Imperial College London

Targeting Repressive Histone Methylation in Cancer

Isabel C A Dye

Department of Surgery and Cancer

Imperial College London

A thesis submitted in part fulfillment for the requirements of the degree of Doctor of Philosophy

April 2024

Abstract

EZH2 and G9a are repressive histone lysine methyltransferases that are heavily implicated in tumourigenesis. Both enzymes silence transposable elements, the expression of which stimulates viral mimicry immune signalling. Additionally, EZH2 and G9a promote oncogenic cMYC activity. Analyses of pan-cancer cell lines demonstrated cell-intrinsic anticorrelation between immune signalling and cMYC-G9a/EZH2 repression, supporting contribution of EZH2 and G9a to cMYC-mediated immune repression.

These features recommend EZH2 or G9a inhibition (EZH2i/G9ai) as an anti-tumour approach. However, physical interaction and redundancy between the two enzymes may limit the efficacy of single EZH2i/G9ai. HKMTi-1-005 is a novel compound based on a peptide-competitive G9ai, modified with the aim of introducing EZH2i activity. ATACseq and RNAseq performed in ovarian high-grade serous carcinoma (HGSC) models supported EZH2i and G9ai effects; dual EZH2/G9a repression targets gained chromatin accessibility and were transcriptionally upregulated by treatment. Additionally, HKMTi-1-005 sensitivity was highly correlated with that of single EZH2i or G9ai in analyses of over 600 cancer cell lines. Biochemical assays revealed minimal direct EZH2i by HKMTi-1-005, suggesting an indirect EZH2 targeting effect.

Further investigations demonstrated that HKMTi-1-005 upregulated transposable elements and induced viral mimicry in HGSC models. Immune signalling was characterised by upregulation of a 7-gene signature (7ISG). In parallel, HKMTi-1-005 de-repressed targets of cMYC-G9a/EZH2 repression and downregulated canonical cMYC targets. Reflecting cMYC and immune effects, *in vitro* HKMTi-1-005 sensitivity was greater in *MYC*-deregulated lines and those with low basal 7ISG expression.

HKMTi-1-005 has been shown to prolong survival of HGSC mouse models, a tumour type with extensive *MYC* deregulation. The KM^2 murine model of lung adenocarcinoma was used to verify treatment efficacy in another *MYC*-deregulated tumour; HKMTi-1-005 induced immune stimulation and prolonged survival.

The data support HKMTi-1-005 as an EZH2 and G9a targeting approach with anti-tumour efficacy. Further, they propose *MYC* deregulation and the 7ISG as stratification biomarkers for HKMTi-1-005 sensitivity.

Contents

1 Introduction **26**

1.1 Chromatin and epigenetics: an introduction 26

1.1.1 Chromatin: the solution to a complex packaging problem 26

1.1.2 Chromatin states and nuclear organisation 26

1.1.3 TADs and chromatin loops 27

1.1.4 Chromatin modifications 29

1.2 Histone 3 lysine 9 methylation (H3K9me): G9a/GLP 32

1.2.1 Mammalian H3K9me enzymes: introduction to G9a/GLP activity and chromatin targeting 32

1.2.2 H3K9me: function in gene expression and chromatin organisation 33

1.2.3 Non-canonical G9a effects 34

1.2.4 H3K9 demethylation 34

1.3 Histone 3 lysine 27 methylation (H3K27me): EZH2 35

1.3.1 Mammalian H3K27me enzymes: introduction to EZH1/2 activity and chromatin targeting 35

1.3.2 H3K27me: function in gene expression and chromatin organisation 37

1.3.3 Non-canonical EZH2 effects 38

1.3.4 H3K27 demethylation 39

1.4 Epigenetic crosstalk and interdependency 39

1.4.1 Crosstalk between histone methylation and DNA methylation 39

1.4.2 Bivalency: co-existence of activating chromatin marks with H3K27me/H3K9me 40

1.4.3 Interactions between EZH2 and G9a 41

1.5 EZH2 and G9a in cancer 43

1.5.1 Dysregulation and prognostic significance 43

1.5.2 Oncogenic mechanisms of action 44

1.6 Targeting EZH2 and G9a as an anti-tumour approach 46

1.6.1 Single inhibition of EZH2 or G9a 46

1.6.2 EZH2i or G9ai in combination with other epigenetic therapies 47
1.6.3 Targeting EZH2-G9a co-operation: dual inhibition 47
1.7 Transposable elements (TEs) 48
1.7.1 Introduction to transposable elements: nomenclature, abundance and evolutionary ages 48
1.7.2 TEs in the genome: effects on genome structure and function 51
1.7.3 A crucial role for epigenetics in TE silencing 53
1.7.4 A brief overview of TEs in cancer 54
1.7.5 Exploitation of TEs in cancer via viral mimicry 54
1.7.6 Targeting EZH2 and/or G9a as a viral mimicry approach 55
1.7.7 Which features prime cells for viral mimicry? 57
1.8 cMYC 58
1.8.1 cMYC: structure, interactions and transcriptional activation effects 58
1.8.2 Transcriptional repression and immune evasion mediated by cMYC 60
1.8.3 Interactions between cMYC and G9a 61
1.8.4 Interactions between cMYC and EZH2 61
1.8.5 cMYC interactions with other H3K27 and H3K9 epigenetic modifiers 63
1.8.6 Therapeutic targeting of cMYC: current approaches 63
1.8.7 Epigenetic compounds: a novel therapeutic approach for cMYC-deregulated cancer? 66
1.9 cMYC-deregulated cancer: ovarian high-grade serous carcinoma (HGSC) 68
1.9.1 Overview 68
1.9.2 EZH2 and G9a: contributions to HGSC carcinogenesis and chemoresistance . . 68
1.9.3 HGSC treatment 69
1.9.4 The immune system in HGSC 69
1.9.5 Epigenetic targeting of EZH2 and G9a in HGSC: an anti-tumour immune boost? 71
1.10 cMYC-deregulated cancer: lung adenocarcinoma (LuAd) 71
1.10.1 Overview 71
1.10.2 cMYC deregulation and the immune system in LuAd 72
1.11 Introduction to HKMTi-1-005 73
1.11.1 Identification of a compound with dual activity against EZH2 and G9a 73
1.12 Aims and objectives 75

2 Methods **76**

2.1 Overview of bioinformatic tools 76

2.2 HGSC cell lines . . . 76
2.2.1 HGSC cell line details . . . 76
2.2.2 Cell culture and HKMTi-1-005 treatment . . . 79
2.2.3 RNAseq . . . 80
2.2.4 ATACseq . . . 81
2.3 ID8 *in vivo* samples . . . 83
2.3.1 ID8 *in vivo* model details . . . 83
2.3.2 HKMTi-1-005 treatment and sample collection . . . 83
2.3.3 RNAseq . . . 83
2.3.4 ATACseq . . . 84
2.4 ID8 *in vitro* samples . . . 86
2.4.1 ID8 cell line details . . . 86
2.4.2 Cell culture and HKMTi-1-005 treatment . . . 86
2.4.3 ATACseq . . . 86
2.5 RNAseq: downstream analyses . . . 87
2.6 ATACseq: downstream analyses . . . 88
2.7 The GDSC2 cell line panel . . . 88
2.7.1 GDSC2 cell line details and data access . . . 88
2.8 Patient cohorts . . . 89
2.8.1 Patient cohort details . . . 89
2.8.2 Patient cohort analyses . . . 90
2.9 Genesets and scoring . . . 91
2.9.1 Genesets . . . 91
2.9.2 Singscore . . . 91
2.10 Statistical analyses . . . 93
2.11 Collaborative projects . . . 93
2.11.1 CRUK Scotland Institute: lung adenocarcinoma experiments . . . 93
2.11.2 AstraZeneca: biochemical investigations into HKMTi-1-005 . . . 95
2.12 Graphics . . . 95

3 Results Section 1: HKMTi-1-005 has EZH2i- and G9ai-like properties 97

3.1 Introduction . . . 97
3.2 *In vitro* and *in vivo* transcriptome analysis: HKMTi-1-005 treatment effect on HGSC 100
3.2.1 RNAseq data processing: HGSC cell lines . . . 100
3.2.2 RNAseq data processing: ID8 *in vivo* murine model of HGSC . . . 102

3.2.3 Overview of HKMTi-1-005 transcriptional effects . . . 105
3.2.4 HKMTi-1-005 transcriptional effects at EZH2 and G9a targets . . . 110
3.2.5 Preliminary investigation into potential non-enzymatic EZH2i effects of HKMTi-1-005 . . . 111
3.3 *In vitro* and *in vivo* chromatin accessibility analysis: HKMTi-1-005 treatment effect on HGSC . . . 115
3.3.1 ATACseq data processing: HGSC cell lines . . . 115
3.3.2 ATACseq data processing: ID8 *in vitro* and *in vivo* murine models of HGSC . 119
3.3.3 Overview of chromatin accessibility effects . . . 124
3.3.4 HKMTi-1-005 chromatin effects at EZH2 and G9a targets . . . 126
3.3.5 Characterisation of differentially accessible chromatin . . . 132
3.3.6 Enhancers and bivalent chromatin . . . 133
3.3.7 Relating changes in gene expression to altered chromatin accessibility following HKMTi-1-005 treatment . . . 138
3.4 Investigating HKMTi-1-005 in comparison to single EZH2 and G9a inhibitors . . . 141
3.4.1 HKMTi-1-005 displays a similar sensitivity profile to single EZH2 or G9a inhibitors . . . 141
3.4.2 Haematological, sarcoma, melanoma and SCLC cell lines display heightened sensitivity to HKMTi-1-005 . . . 144
3.4.3 Aberrant *EZH2* or *G9a* is associated with increased sensitivity to HKMTi-1-005 144
3.4.4 AstraZeneca: studying the biochemical activity of HKMTi-1-005 . . . 151
3.5 Discussion: conclusions, limitations and future work . . . 153

4 Results Section 2: Characterisation of the viral mimicry response induced by HKMTi-1-005 158

4.1 Introduction . . . 158
4.2 HKMTi-1-005 induces immune stimulation in HGSC models . . . 161
4.2.1 Transcriptional changes in immune pathways induced by HKMTi-1-005 . . . 161
4.3 HKMTi-1-005 induces transposable elements in HGSC models . . . 170
4.3.1 De-repression of transposable element transcription: sub-family level analysis . 170
4.3.2 De-repression of transposable element transcription: locus-specific analysis . . 177
4.3.3 Increased accessibility of transposable elements . . . 183
4.4 Investigating the aberrant PEO4 response . . . 191
4.4.1 A high basal immune signalling state . . . 191
4.4.2 Heightened basal expression of HKMTi-1-005-induced TEs . . . 193

4.5 Discussion: conclusions, limitations and future work 200

5 Results Section 3: Investigating the relationship between *MYC*-mediated immune repression and HKMTi-1-005 209

5.1 Introduction 209

5.2 cMYC-G9a and EZH2 repression anticorrelates with immune signalling scores: *in vitro* studies 213

5.2.1 cMYC-G9a/EZH2 repression inversely correlates with immune signalling in the GDSC2 cell line panel 213

5.2.2 *MYC*-amplified cell lines have greater cMYC-G9a/EZH2 repression with lower immune scores 214

5.2.3 Pan-cancer cell line observations are corroborated in individual analyses of ovarian, breast and LuAd cell lines 218

5.3 cMYC-G9a and EZH2 repression anticorrelates with immune signalling scores: *in vivo* studies 221

5.3.1 cMYC-G9a/EZH2 repression inversely correlates with immune signalling in ovarian, breast and LuAd patient cohorts 221

5.3.2 *MYC*-amplified patient samples do not consistently show greater cMYC-G9a/EZH2 repression or lower immune scores 227

5.3.3 *MYC* amplification status impacts immune infiltration in patient samples . . . 229

5.4 HKMTi-1-005 upregulates cMYC-G9a/EZH2 targets 238

5.4.1 HKMTi-1-005 treatment induces de-repression of cMYC-G9a and EZH2 targets, with downregulation of canonical cMYC signalling 238

5.5 Sensitivity to HKMTi-1-005 relates to *MYC* status and the 7ISG 240

5.5.1 *MYC* deregulation is associated with increased HKMTi-1-005 sensitivity . . . 240

5.5.2 Low basal 7ISG scores correlate with greater HKMTi-1-005 sensitivity, independent of canonical cMYC activity 242

5.6 Assessing prognostic significance of the 7ISG in *MYC*-deregulated tumours 247

5.6.1 A high 7ISG score is associated with longer survival in *MYC*-amplified HGSC patients 247

5.7 Examining effects of HKMTi-1-005 treatment in another *MYC*-deregulated tumour model 251

5.7.1 HKMTi-1-005 treatment induces immune signalling with growth inhibition in lung cancer cell lines 251

5.7.2 HKMTi-1-005 treatment increases survival of a murine *MYC*-driven LuAd model252

5.8 Discussion: conclusions, limitations and future work 254

6 Overall discussion **261**

6.1 Dual targeting of EZH2/G9a: the clinical and biological context 261

6.2 Overview of findings 265

6.3 Limitations and future directions 270

6.3.1 Overview 270

6.3.2 cMYC-G9a/EZH2 complexes: formation, genomic targets and treatment effect 271

6.3.3 Understanding the priming effect of *MYC* amplification and basal 7ISG scores 271

6.3.4 *In vivo* experiments 272

6.4 Final remarks 273

7 Bibliography **274**

A Appendix **313**

List of Figures

1.1 The structure of a nucleosome. . . . 27
1.2 DNA compaction: chromatin structures. . . . 28
1.3 Organisation of chromatin and DNA in the nucleus. . . . 29
1.4 Histone modification schematic. . . . 30
1.5 Histone modifications associated with active or inactive genes. . . . 31
1.6 Structures of G9a and GLP domains. . . . 32
1.7 PRC2 and EZH2 structure and organisation. . . . 36
1.8 *EZH2* and *G9a* genomic alterations in cancer. . . . 43
1.9 *EZH2* and *G9a* aberrant expression in cancer. . . . 44
1.10 Classification and structure of transposable elements (TEs). . . . 50
1.11 Mechanisms by which transposable elements (TEs) regulate transcription. . . . 52
1.12 Overview of intracellular innate immune signalling pathways downstream of nucleic acid (NA) recognition. . . . 56
1.13 cMYC structure and interactions. . . . 59
1.14 cMYC inhibition approaches. . . . 64
1.15 cMYC-G9a/EZH2 immune repression schematic. . . . 67
1.16 Histological classifications of lung cancer and the proportion of total lung cancer incidence they constitute. . . . 72
1.17 Chemical structure of HKMT inhibitors. . . . 74

2.1 Schematic overview of RNAseq and ATACseq analyses of HKMTi-1-005 treatment effect. . . . 79
2.2 Schematic describing the cell-free PRC2 inhibition assay performed by AZ. . . . 96

3.1 QC for *in vitro* HGSC RNAseq data. . . . 101
3.2 Alignment metrics for *in vitro* HGSC RNAseq data. . . . 102
3.3 QC following differential expression analyses, *in vitro* HGSC RNAseq data. . . . 103
3.4 Alignment metrics for ID8 *in vivo* RNAseq data. . . . 104
3.5 QC following differential expression analyses, ID8 *in vivo* RNAseq data. . . . 105

3.6 RNAseq volcano plots, HGSC samples. . . . 106
3.7 RNAseq ORA results, HGSC samples. . . . 108
3.8 *Interleukin 24 (IL24)* differential expression following HKMTi-1-005 treatment. . . . 110
3.9 fGSEA on RNAseq data: EZH2 and G9a genesets. . . . 112
3.10 fGSEA on RNAseq data: histone genesets. . . . 113
3.11 PRC2 components: differential expression. . . . 113
3.12 PRC2 recruitment factors: differential expression. . . . 114
3.13 QC for *in vitro* HGSC ATACseq data. . . . 116
3.14 Insert size metrics for *in vitro* HGSC ATACseq data. . . . 117
3.15 Schematic illustrating the two methods of ATACseq analysis. . . . 117
3.16 PCA plots for *in vitro* HGSC ATACseq data. . . . 118
3.17 Heatmap plots for *in vitro* HGSC ATACseq data. . . . 118
3.18 QC metrics for ID8 *in vivo* ATACseq data. . . . 120
3.19 QC metrics for ID8 *in vitro* ATACseq data. . . . 121
3.20 Insert size metrics for ID8 *in vivo* ATACseq data. . . . 122
3.21 Insert size metrics for BRCA1 ID8 *in vitro* ATACseq data. . . . 122
3.22 PCA plots for ID8 *in vivo* and *in vitro* ATACseq data. . . . 123
3.23 ATACseq peaks of accessibility, in control versus HKMTi-1-005-treated conditions. . . . 124
3.24 ATACseq volcano plots. . . . 125
3.25 fGSEA on ATACseq data. . . . 127
3.26 H3K27me3 accessibility heatmaps. . . . 128
3.27 H3K9me3 accessibility heatmaps. . . . 129
3.28 H3K9Ac, H3K27Ac, H3K4me1 and H3K36me3 accessibility heatmaps. . . . 131
3.29 Accessibility changes at genomic features, HGSC cell lines. . . . 132
3.30 Accessibility changes at genomic features, ID8 *in vivo* and *in vitro* samples. . . . 133
3.31 Motif enrichment analysis. . . . 134
3.32 Enhancer accessibility heatmaps. . . . 135
3.33 Bivalent region accessibility heatmaps. . . . 136
3.34 ATACseq fGSEA. . . . 139
3.35 Violin plots illustrating compound sensitivity across the GDSC2 cohort. . . . 142
3.36 Spearman's correlation between sensitivity to HKMTi-1-005, GSK343 (EZH2i), UNC0638 (G9ai) and cisplatin. . . . 142
3.37 Correlation between sensivity to HKMTi-1-005 and to epigenetic drugs. . . . 143
3.38 Correlation between sensivity to HKMTi-1-005 and to alternative drugs. . . . 143

3.39 Sensitivity to HKMTi-1-005, GSK343, UNC0638 or cisplatin: visualisation by tumour type. . . . 145
3.40 Sensitivity to HKMTi-1-005, GSK343, UNC0638 or cisplatin: lung SCLC, melanoma and Ewing's sarcoma cell lines. . . . 145
3.41 *EZH2* or *G9a* mutations versus sensitivity. . . . 146
3.42 *EZH2* or *G9a* OE versus sensitivity. . . . 147
3.43 Spearman's correlation between sensitivity to HKMTi-1-005, GSK343 (EZH2i), UNC0638 (G9ai) and cisplatin: solid cell lines. . . . 148
3.44 *EZH2* or *G9a* mutations versus sensitivity: solid cell lines. . . . 149
3.45 *EZH2* or *G9a* OE versus sensitivity: solid cell lines. . . . 150
3.46 PRC2 inhibition assay results. . . . 152

4.1 fGSEA of hallmark genesets, HGSC samples. . . . 161
4.2 fGSEA of immune genesets, HGSC samples. . . . 162
4.3 IFNG and innate immune gene expression changes. . . . 163
4.4 Consistently upregulated immune genes, HGSC samples. . . . 165
4.5 ISG expression changes. . . . 168
4.6 7ISG expression changes, HGSC cell lines. . . . 169
4.7 7ISG expression changes, ID8 *in vivo* model. . . . 169
4.8 TE and ERV nomenclature used in analyses. . . . 170
4.9 TE expression changes by superfamily, at the sub-family level. . . . 171
4.10 Volcano plot of TE changes in HGSC samples, sub-family level. . . . 172
4.11 Total up- versus down-regulated TEs in HGSC samples, sub-family level. . . . 173
4.12 TE changes at sub-family level, HGSC samples. . . . 174
4.13 Comparison of TEs induced in common across samples, sub-family level. . . . 175
4.14 LTR/ERV changes at sub-family level, HGSC samples. . . . 176
4.15 TE expression changes by superfamily, at the locus-specific level. . . . 178
4.16 Volcano plot of TE changes in HGSC samples, locus-specific level. . . . 179
4.17 Total up- versus down-regulated TEs and LTRs/ERVs in HGSC samples, locus-specific level. . . . 180
4.18 Comparison of TEs induced in common across samples, locus-specific level. . . . 180
4.19 TE changes at locus-specific level, HGSC samples. . . . 181
4.20 LTR/ERV changes at locus-specific level, HGSC samples. . . . 182
4.21 Volcano plot of TE accessibility in HGSC samples. . . . 184

4.22 Total up- versus downregulated DACs at TEs and LTRs/ERVs in PEO1, PEO4 and ID8 *in vivo* samples. 185
4.23 Total up- versus downregulated DACs at TEs and LTRs/ERVs in ID8 *in vitro* samples. 185
4.24 Accessibility changes across TE superfamilies, in PEO1, PEO4 and ID8 *in vivo* samples: DAC method. 186
4.25 Accessibility changes across TE superfamilies, in ID8 *in vitro* samples: DAC method. 187
4.26 Accessibility changes across LTR/ERV families, in PEO1, PEO4 and ID8 *in vivo* samples: DAC method. 188
4.27 Accessibility changes across LTR/ERV families, in ID8 *in vitro* samples: DAC method. 189
4.28 Accessibility changes across TE superfamilies and LTR/ERV families, in PEO1, PEO4 and ID8 *in vivo* samples: subtraction method. 190
4.29 Accessibility changes across TE superfamilies and LTR/ERV families, in ID8 *in vitro* samples: subtraction method. 190
4.30 Comparison of PEO1 and PEO4 at baseline: gene expression. 192
4.31 Basal expression of VM ISGs across HGSC cell lines. 193
4.32 Comparison of PEO1 and PEO4 at baseline: TE expression. 194
4.33 PEO1 versus PEO4: barplot of TE and LTR/ERV differences, sub-family level. 194
4.34 PEO1 versus PEO4: barplot of TE and LTR/ERV differences, locus-specific level. . . 195
4.35 PEO1 versus PEO4: proportion of TEs by superfamily, measured at sub-family level. 195
4.36 PEO1 versus PEO4: proportion of LTRs/ERVs by family, measured at sub-family level. 196
4.37 PEO1 versus PEO4: proportion of TEs by superfamily, measured at locus-specific level. 196
4.38 PEO1 versus PEO4: proportion of LTRs/ERVs by family, measured at locus-specific level. 196
4.39 PEO1 versus PEO4: differences in HERVH expression at locus-specific level. 197
4.40 PEO1 versus PEO4: differences in TE and HERVH accessibility. 198
4.41 PEO1 versus PEO4: up- versus down-regulated DACs at TEs and LTR/ERVs. 198
4.42 PEO1 versus PEO4: differences in TE accessibility, proportion by superfamily. 199
4.43 PEO1 versus PEO4: differences in LTR/ERV accessibility, proportion by family. . . . 199
4.44 Comparison of 7ISG changes in response to DNMTi, HDACi or combination in 4 ovarian cancer cell lines. 201
4.45 Comparison of 7ISG changes in response to DNMTi, HDACi or combination in the TykNu ovarian cancer cell line. 201
4.46 ERV classes versus CpG density, from [153]. 205

5.1 Correlation between cMYC-G9a/EZH2 repression scores and canonical cMYC signalling in the GDSC2 cell line panel. . . . 214
5.2 Correlation between immune genesets and genesets relating to the extent of cMYC-G9a/EZH2 repression or the extent of hallmark cMYC activity in the GDSC2 cell line panel. . . . 215
5.3 Correlation between cMYC-G9a/EZH2 repression scores and canonical MYC signalling or immune signalling scores in the GDSC2 cell line panel: stratified by *MYC* status. . . . 216
5.4 *MYC* amplification versus cMYC-G9a/EZH2 scores in GDSC2 cell lines. . . . 217
5.5 *MYC* amplification versus immune scores in GDSC2 cell lines. . . . 217
5.6 Correlation between cMYC-G9a/EZH2 repression scores and canonical cMYC signalling in ovarian, breast and LuAd cell lines. . . . 219
5.7 Correlation between immune genesets and genesets relating to the extent of cMYC-G9a/EZH2 repression or the extent of hallmark cMYC activity in ovarian, breast and LuAd cell lines. . . . 220
5.8 Correlation between cMYC-G9a/EZH2 repression scores and canonical cMYC signalling in *in vivo* patient cohorts. . . . 221
5.9 Correlation between immune genesets and genesets relating to the extent of cMYC-G9a/EZH2 repression or the extent of hallmark cMYC activity in *in vivo* patient cohorts. . . . 222
5.10 Correlation between cMYC-G9a/EZH2 repression scores and canonical MYC signalling or immune signalling scores in *in vivo* patient cohorts: stratified by *MYC* status. . . . 225
5.11 *MYC* amplification versus cMYC-G9a/EZH2 scores or immune scores in *in vivo* patient cohorts. . . . 227
5.12 *MYC* amplification versus 7ISG scores in *in vivo* patient cohorts. . . . 228
5.13 CIBERSORTx results in the ICGC-OV cohort. . . . 229
5.14 cTME results in the ICGC-OV cohort. . . . 230
5.15 cTME results in the TCGA-OV cohort, stratified by *MYC* status. . . . 231
5.16 CIBERSORTx results in the TCGA-OV cohort, stratified by *MYC* status. . . . 232
5.17 CIBERSORTx results in the TCGA-LUAD cohort, stratified by *MYC* status. . . . 232
5.18 cTME results in the TCGA-BRCA cohort, stratified by *MYC* status. . . . 233
5.19 CIBERSORTx results in the TCGA-BRCA cohort, stratified by *MYC* status. . . . 234
5.20 Inferred immune infiltration in the TCGA-OV cohort: correlation with geneset scores. 235

5.21 Inferred immune infiltration in the TCGA-BRCA cohort: correlation with geneset scores. . . . 236
5.22 CIBERSORTx results in the TCGA-OV and TCGA-BRCA cohorts: correlation with 7ISG and canonical cMYC scores. . . . 237
5.23 fGSEA results for cMYC-G9a/EZH2 and canonical cMYC genesets, HGSC samples. . 239
5.24 *MYC* deregulation versus sensitivity in the GDSC2 cell line panel. . . . 241
5.25 Immune and cMYC scores versus sensitivity in the GDSC2 cell line panel. . . . 243
5.26 Immune and cMYC scores versus sensitivity in the GDSC2 cell line panel: stratified by *MYC* status. . . . 244
5.27 Correlation between canonical cMYC scores and 7ISG scores in the GDSC2 panel, stratified by *MYC* status. . . . 244
5.28 Multiple linear regression: HKMTi-1-005. . . . 245
5.29 Multiple linear regression: UNC0638. . . . 246
5.30 7ISG scores versus overall survival in the ICGC-OV cohort. . . . 248
5.31 7ISG scores versus disease-specific survival in the *MYC*-amplified TCGA-OV cohort. 249
5.32 7ISG scores versus disease-specific survival in the non-amplified TCGA-OV cohort. . . 250
5.33 Analysis of HKMTi-1-005 treatment effect on lung adenocarcinoma cell lines. . . . 252
5.34 Analysis of HKMTi-1-005 treatment effect on a lung adenocarcinoma (LuAd) model: survival effects. . . . 253
5.35 Analysis of HKMTi-1-005 treatment effect on a LuAd model: tumour burden and proliferation. . . . 253

6.1 Schematic illustrating the paradox surrounding inflammatory signalling in *MYC*-amplified cells. . . . 264
6.2 Schematic overview of hypothesised cMYC-G9a/EZH2 and EZH2/G9a repression of immune genes and ERVs, the effect of *MYC* amplification on these axes and the disruption of repression by HKMTi-1-005. . . . 269
6.3 Outstanding questions: cMYC-G9a/EZH2 repression and HKMTi-1-005 treatment mechanisms. . . . 270

A.1 7ISG selection criteria. . . . 314
A.2 Tumour growth inhibitory effect of HKMTi-1-005 in IGROV1 xenograft model. . . . 315
A.3 RNAseq GSEA, GOBP pathways: HGSC samples. . . . 316
A.4 RNAseq GSEA, Hallmark pathways: HGSC samples. . . . 317
A.5 ATACseq FastQC sequence counts, before trimming: ID8 *in vivo* samples. . . . 318
A.6 H3K9Ac accessibility heatmaps. . . . 319

A.7 H3K27Ac accessibility heatmaps. 320
A.8 H3K36me3 accessibility heatmaps. 321
A.9 H3K4me1 accessibility heatmaps. 322
A.10 Accessibility changes versus TSS location, *in vitro* HGSC or ID8 *in vivo* samples. . . 323
A.11 Accessibility changes versus TSS location, ID8 *in vitro* samples. 323
A.12 Motif enrichment analysis, ATACseq subtraction approach. 324
A.13 ATACseq fGSEA: GOBP database. 326
A.14 Bivalent region accessibility heatmaps. 328
A.15 Bivalent H3K9me3-H3K4me3 accessibility heatmaps. 330
A.16 EZH2 or G9a mutation or OE versus sensitivity: haematological cell lines. 331
A.17 Comparison of ISG changes in response to DNMTi, HDACi or combination in 4 ovarian cancer cell lines. 332
A.18 Comparison of ISG changes in response to DNMTi, HDACi or combination in the TykNu ovarian cancer cell line. 332
A.19 *MYC* amplification versus doubling time in the GDSC2 cell line panel. 333
A.20 *MYC* amplification versus *MYC* OE in the GDSC2 cell line panel: Fisher's test. . . . 333
A.21 Multiple linear regression: olaparib. 335
A.22 Multiple linear regression: vorinostat. 335
A.23 Schoenfeld residuals. 336
A.24 7ISG scores versus overall survival in the *MYC*-amplified TCGA-OV cohort. 337
A.25 7ISG scores versus overall survival in the non-amplified TCGA-OV cohort. 337
A.26 *MYC* amplification versus sensitivity in SCLC lines. 337

List of Tables

1.1 Table summarising current EZH2 and G9a inhibition approaches. 46
1.2 MYC inhibition approaches. 65

2.1 Core bioinformatic tools: part 1. 77
2.2 Core bioinformatic tools: part 2. 78
2.3 FRiPs and total reads in PEO1 and PEO4 ATACseq samples. 82
2.4 FRiPs and total reads in ID8 *in vivo* ATACseq samples. 85
2.5 FRiPs and total reads in F3, BRCA1 and BRCA2 ID8 *in vitro* ATACseq samples. . . 87
2.6 List of genesets. 92
2.7 Primers used in qRT-PCR of lung cancer cell lines. 95

3.1 ATACseq versus RNAseq comparison: upregulation. 138
3.2 ATACseq versus RNAseq comparison: downregulation. 138
3.3 IC_{50} values for H3K27me3 inhibition. 151

4.1 Genes in the ISG geneset defined in [439]. 167
4.2 PEO1 versus PEO4: top GSEA results. 191
4.3 LTR induction by epigenetic therapy, from [153]. 205

5.1 CoxPH model in the ICGC-OV cohort. 248
5.2 CoxPH model in the *MYC*-amplified TCGA-OV cohort. 249
5.3 CoxPH model in the non-amplified TCGA-OV cohort. 250

A.1 Co-occurrence of *EZH2* and *G9a* alterations in TCGA patient cohorts. 314
A.2 Co-occurrence of *EZH2* or *G9a* alterations with *MYC* alterations in TCGA patient cohorts. 314
A.3 Statistical comparison of RNAseq results using mm39 versus mm10 reference genome annotation. 314
A.4 Statistical comparison of ATACseq approach results. 315
A.5 Overlap in transcription factor motif enrichment across HGSC samples. 325

A.6 Statistical testing of the overlap between more accessible TEs and transcriptionally upregulated TEs at the locus-specific level. 326
A.7 95% confidence intervals for correlation between EZH2 repression geneset scores and immune geneset scores, stratified by *MYC* status. 327
A.8 95% confidence intervals for correlation between cMYC-G9a repression geneset scores and immune geneset scores, stratified by *MYC* status. 329
A.9 TCGA-BRCA cohort: 95% confidence intervals for correlation between cMYC-G9a/EZH2 repression geneset scores or canonical cMYC scores and immune geneset scores. . . . 333
A.10 Changes in expression of *MYC* following HKMTi-1-005 treatment in HGSC samples. 333
A.11 95% confidence intervals for correlation between immune geneset scores and sensitivity to HKMTi-1-005. 334
A.12 95% confidence intervals for correlation between immune geneset scores and sensitivity to HKMTi-1-005, stratified by *MYC* status. 334

Acknowledgements

Enormous thanks must go to my supervisors, Professors Iain McNeish and Bob Brown. You have both provided excellent mentorship and supervision from the very start of my PhD, with an invaluable combination of encouragement and advice. It has been a privilege to learn from such extensive expertise and experience. Thank you also for being hugely accommodating in the face of various trials and tribulations; I have thoroughly enjoyed undertaking this PhD under your guidance.

I am incredibly grateful to my whole lab group for being particularly welcoming and for constantly impressing me with exciting data, rigorous science and an unrelenting sense of humour. Thanks also to those of you who guided me during my short stint in the lab, including Dr Ian Garner and Nahal Masrour who were essential in steering the first months of my PhD. To Dr Sarah Spear, for being an inspiration as a scientist and for encouraging me with your belief in my abilities. To Hasan Mirza, for providing bioinformatic advice and bioinformatic commiseration, for encouraging my development as a cricket fan and for instilling in me that no number of backups is too many backups. Also, to my fellow PhDs, for providing two necessities: camaraderie and PhD-based humour.

Many thanks must also go to my collaborators Professor Daniel Murphy and Dr Sarah Laing at the CRUK Scotland Institute. Your work investigating HKMTi-1-005 was a crucial component of this project, and your infectious enthusiasm persuaded me to persevere with *MYC* despite its many attempts to put me off.

Thanks to Professor Matthew Fuchter for supplying HKMTi-1-005, and to Dr Francesca Fiorentino for help with all things statistical. To the AstraZeneca team, including Dr James Robinson, thank you for sharing your insights into the mechanism of action of HKMTi-1-005. For funding this PhD, thanks go to the MRC DTP and the STRATiGRAD programme.

To the Gandhi family, thank you for making Bristol a home away from home, and for ensuring I am always (very) well-fed. To my friends, thank you for all of the many adventures and for generally being an exceptional bunch of people. Thanks also for being interesting and interested, but not so interested that you read beyond your mention here. To my family and to Sanchit – it's unimaginative but true – I could not have done it without you. Thank you for your unwavering support throughout these four years and beyond, and for being such an immense source of happiness and inspiration for me.

Statement of Originality

This thesis consists of work performed by the author, except where stated. Collaborators are appropriately credited for their contributions, and published work is referenced and listed in the bibliography. A publication license has been obtained for all graphics used, with inclusion of the relevant citation.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abbreviations

List of Abbreviations

AP1 Activator Protein 1

AR androgen receptor

ARKS Alanine-Arginine-Lysine-Serine

ATACseq assay for transposase-accessible chromatin with sequencing

AZ AstraZeneca

bHLHZip basic helix-loop-helix-leucine zipper

BRCA1 BRCA1 DNA repair associated

BRCA2 BRCA2 DNA repair associated

BRD4 Bromodomain-containing Protein 4

CCL5 CC Motif Chemokine Ligand 5

CGI CpG island

cHC constitutive heterochromatin

ChIP Chromatin immunoprecipitation

CI confidence interval

co-IP co-immunoprecipitation

CoxPH Cox proportional hazards

CTCF CCCTC-binding factor

cTME Consensus^{TME}

CXCL10 CXC Motif Chemokine Ligand 10

CXCL9 CXC Motif Chemokine Ligand 9

DAC differentially accessible chromatin region

DEG differentially expressed gene

DNMT DNA methyltransferase

dsRNA double stranded RNA

DSS disease-specific survival

E-box Enhancer-box

EC euchromatin

EED Embryonic Ectoderm Development

EOC Epithelial ovarian cancer

ERV endogenous retrovirus

ESC embryonic stem cell

EZH1 Enhancer of Zeste Homologue 1

EZH2 Enhancer of Zeste Homologue 2

fGSEA Fast gene set enrichment analysis

fHC facultative heterochromatin

FIGO Fédération Internationale de Gynécologie et d'Obstétrique

FPKM Fragments Per Kilobase of transcript per Million mapped reads

FRET fluorescence resonance energy transfer

FRiP fraction of reads in peaks

GDSC2 Genomics of Drug Sensitivity in Cancer 2

GLP G9a-Like Protein

GOBP Gene Ontology Biological Process

GSEA gene set enrichment analysis

HC heterochromatin

HDAC histone deacetylase

HERVH human endogenous retrovirus sub-family H

HGSC Ovarian high-grade serous carcinoma

HIF1A Hypoxia Inducible Factor 1 Alpha

HKMT histone lysine methyltransferase

HMT histone methyltransferase

HP1 Heterochromatin Protein 1

HR homologous recombination

IC_{50} half-maximal inhibitory concentration

ICGC International Cancer Genome Consortium

ICR Institute of Cancer Research

IFN interferon

IFNA interferon alpha

IFNB interferon beta

IFNG interferon gamma

IL24 Interleukin 24

ISG interferon-stimulated gene

JARID2 Jumonji And AT-Rich Interaction Domain Containing 2

JmjC Jumonji C

KDM histone lysine demethylase

LAD lamina-associated domain

LINE long-interspersed nuclear element

lncRNA long non-coding RNA

log2FC log2 fold-change

LTR long terminal repeat

LuAd lung adenocarcinoma

MAX MYC Associated Factor X

MBII Myc Box II

MoA mechanism of action

MoR median of ratios

mRNA messenger RNA

MSigDB Molecular Signatures Database

mya million years ago

NK natural killer

OE overexpression

ORA over-representation analysis

OS overall survival

p-adj adjusted p-value

PARPi Poly-ADP Ribose Polymerase inhibitor

PCa prostate cancer

PCA principal component analysis

PFS progression-free survival

PRC1 Polycomb Repressor Complex 1

PRC2 Polycomb Repressor Complex 2

QC quality control

qRT-PCR Quantitative reverse transcription polymerase chain reaction

RBBP Retinoblastoma-Binding Protein

RiP reads in peaks

RNAi RNA interference

RNAPII RNA polymerase II

SAM S-Adenosyl-l-methionine

SCLC small cell lung cancer

SET Su(var)3–9 Enhancer-of-zeste Trithorax

SETDB1 SET Domain Bifurcated 1

SETDB2 SET Domain Bifurcated 2

SINE short-interspersed nuclear element

siRNA short interfering RNA

STAT1 Signal Transducer And Activator Of Transcription 1

STING Stimulator Of Interferon Response CGAMP Interactor

SUV39H1 Suppressor of Variegation 3-9 Homologue 1

SUV39H2 Suppressor of Variegation 3-9 Homologue 2

SUZ12 SUZ12 polycomb repressive complex 2 subunit

TAD topologically associating domain

TBK1 TANK-Binding Kinase 1

TCGA The Cancer Genome Atlas

TE transposable element

TF transcription factor

Tfh T follicular helper

TME tumour microenvironment

TPM transcripts per million

TSS transcription start site

VM viral mimicry

Chapter 1

Introduction

1.1 Chromatin and epigenetics: an introduction

1.1.1 Chromatin: the solution to a complex packaging problem

Within a eukaryotic cell, roughly 2 metres of DNA must be condensed to fit within a nucleus of 6µm in diameter [1]. A vast compaction of DNA is required to achieve this. However, processes such as DNA transcription, replication or repair demand that much of the DNA must retain the ability to become rapidly accessible. The solution to this problem in eukaryotes is an intricate packaging system called chromatin.

The basic unit of chromatin is the nucleosome. Nucleosomes are made up of 145-147 base pairs of DNA wrapped around an octamer of histone proteins, comprising two molecules of each of H2A, H2B, H3 and H4 [2] (figure 1.1). An additional H1 histone protein stabilises this structure by binding nucleosomes and the linker DNA that joins them [3]. Interaction between nucleosomes allows formation of chromatin fibres, with further compaction possible via different levels of interaction between these chromatin fibres [3] (figure 1.2).

1.1.2 Chromatin states and nuclear organisation

Chromatin is often broadly categorised into two states describing the level of compaction: euchromatin (EC) and heterochromatin (HC) [4]. EC is decondensed and transcriptionally active, whilst HC is condensed and transcriptionally silent, found either as facultative heterochromatin (fHC) or constitutive constitutive heterochromatin (cHC). cHC remains permanently inactive, mostly playing a structural role such as maintaining compaction of telomeres and centromeres, alongside silencing of repetitive and non-coding regions [4]. In contrast, fHC is a more dynamic heterochromatin state that retains the ability to establish an open conformation in response to developmental or environmental cues [4].

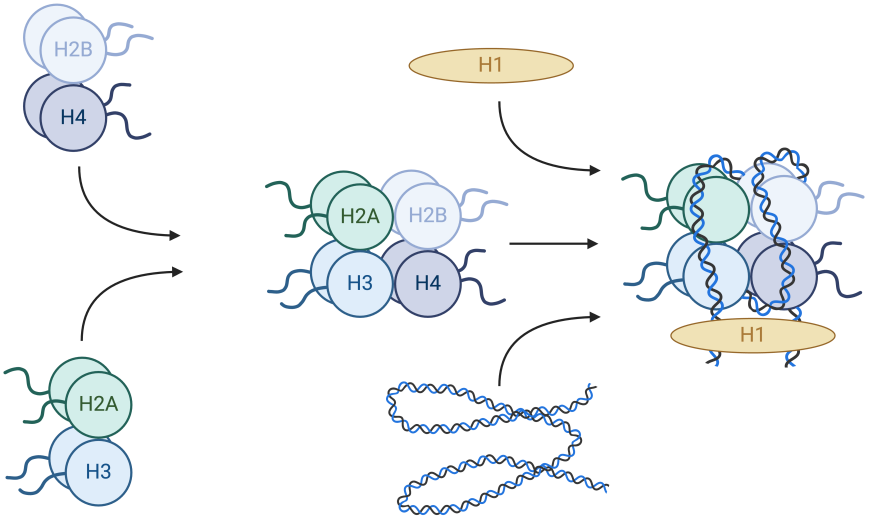


Figure 1.1 **The structure of a nucleosome.** Each nucleosome comprises an octamer of histone proteins: two molecules of each of H2A, H2B, H3 and H4 histones. ∼147bp of DNA is wrapped around the histone octamer, with an additional H1 histone protein stabilising the structure at the point of DNA entry and exit. Image created with BioRender.com.

In the nucleus, there is a strong segregation of these chromatin states. Indeed, the whole genome can be divided into two spatially distinct domains that are cell type-specific and defined at the megabase scale [5]. These domains are known as the A (active) and B (inactive) compartments (figure 1.3), associated with EC and HC respectively [5, 6]. A/B compartments are further subdivided into topologically associating domains (TADs) [7] and chromatin loops [8–10], building a hierarchical 3D genome structure to allow fine-tuned control of gene expression within each compartment.

1.1.3 TADs and chromatin loops

Topologically associating domains comprise groups of loci that demonstrate significantly greater intra- versus inter-domain interactions [8, 12]. This effect promotes interaction between regulatory elements and their gene targets within TADs, such as enhancer or promoter elements [13], whilst insulating genes from external regulatory elements [14]. Loci within a TAD also tend to have similar chromatin states [10], demonstrating the importance of TAD structures not only for co-regulation of genes within the TAD, but also for blocking the spread of chromatin states to neighbouring domains [15]. For example, TADs found in the A compartment are broadly transcriptionally active, with extensive looping between regulatory elements and their gene targets [8]. In contrast, many TADs within the heterochromatic B compartment correspond to lamina-associated domains (LADs), which are found near the nuclear periphery and retained in a transcriptionally repressed state [16].

TAD boundaries are most often demarcated by binding of the insulator-binding protein CCCTC-binding factor (CTCF) [7]. Loop extrusion allows formation of TADs, whereby the ring-like cohesin complex encircles two strands of DNA and DNA is extruded through its centre, until a TAD bound-

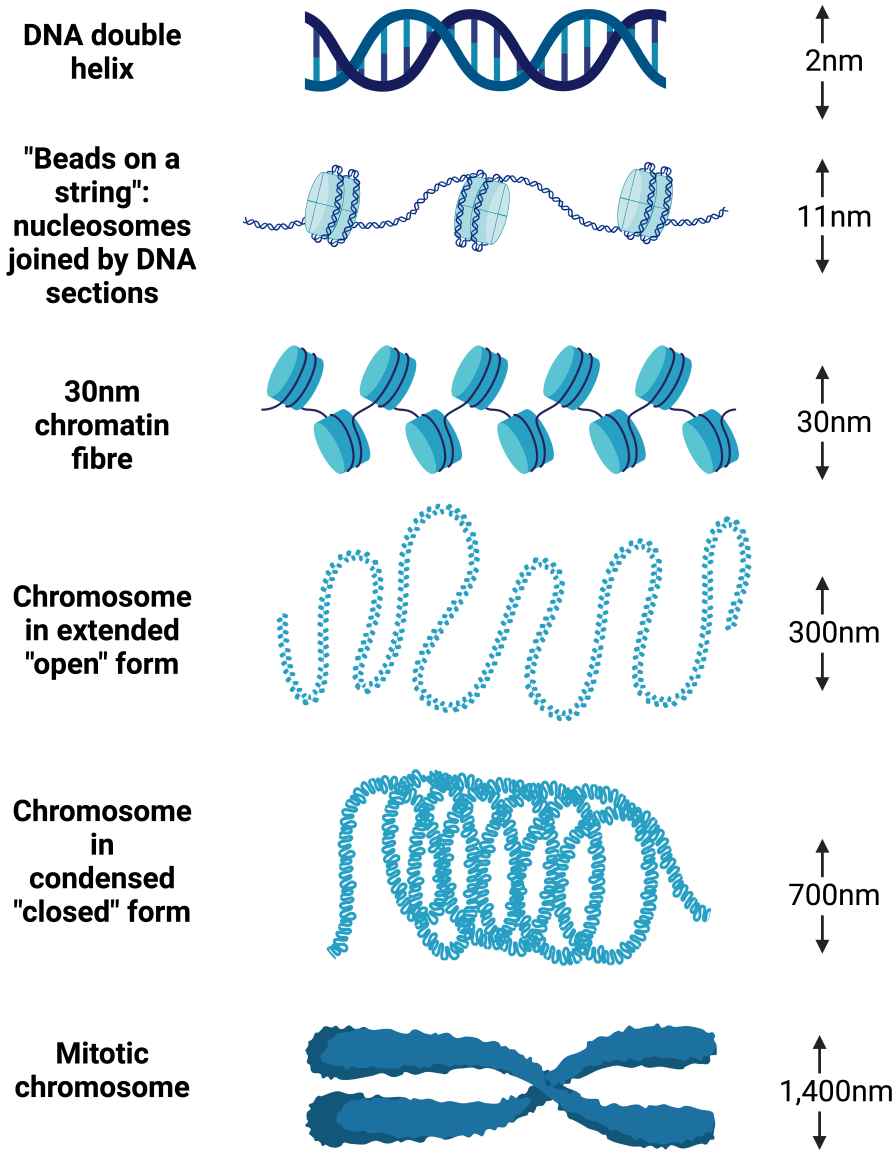


Figure 1.2 **DNA compaction: chromatin structures.** Formation of the nucleosome structure - consisting of DNA wrapped around a histone octamer - achieves the first level of DNA compaction. Linker DNA joining adjacent nucleosomes leads to a "beads on a string" appearance. Nucleosomes can be packed into a 30nm fibre, which is the basis of further folding. Higher-order structures based on interactions between 30nm fibres allow varying levels of chromatin compaction. Image created with BioRender.com, a reproduction of [3].

ary is reached [17]. The boundaries of TADs are largely cell type-invariant, and are even conserved across species [7]. Indeed, TADs are now considered to be a universal unit of eukaryotic chromosome organisation [12, 15]. Therefore, an additional level of control is needed for more precise gene regulation, achieved via the formation of sub-TAD structures known as chromatin loops or insulated neighbourhoods [8–10]. Like TADs, loci within chromatin loops display similar chromatin environments, frequently interact with one another and demonstrate similar long-range interactions [10]. Yet in contrast to TADs, sub-TAD looping is largely cell type-specific [12, 15].

Importantly, recent findings have challenged the strict hierarchical model of chromatin, and demonstrated that the A/B compartmentalisation may be more precise than the megabase-sized

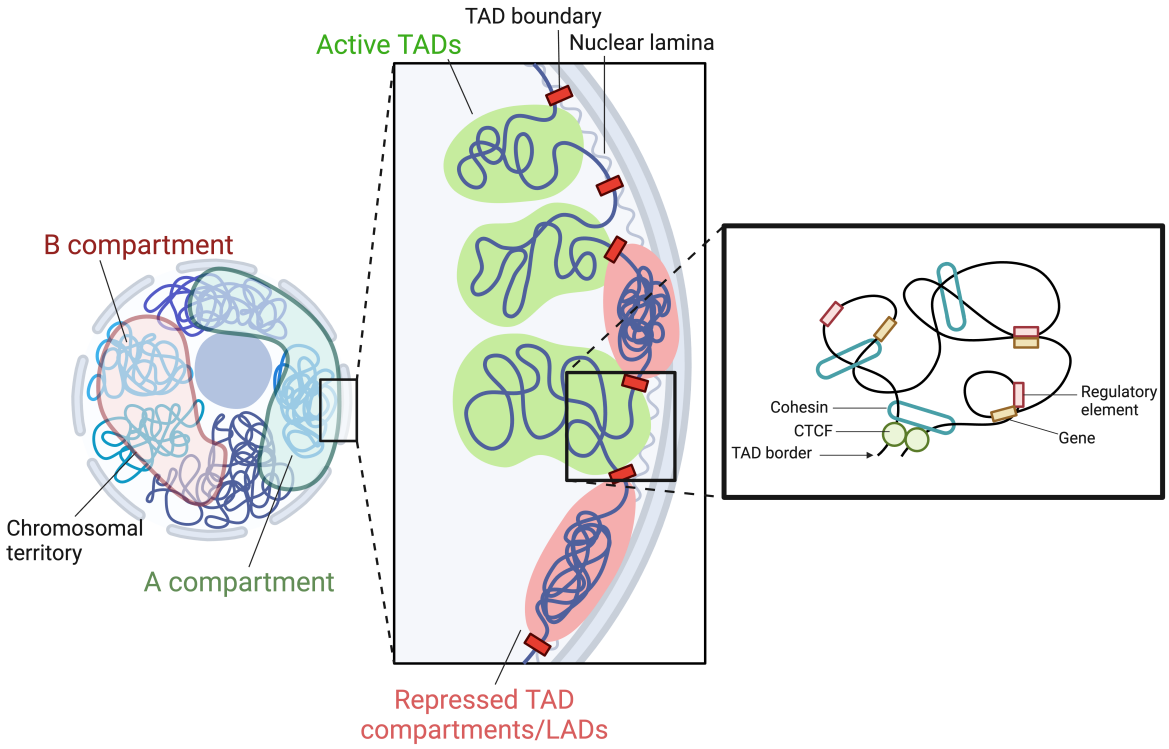


Figure 1.3 **Organisation of chromatin and DNA in the nucleus.** Left: the nucleus is organised into two broad domains, the A (active, shown in green) and B (inactive, shown in red) compartments. Additionally, chromosomes are found in distinct territories. Centre: at a smaller scale, topologically associating domains (TADs) may be repressed or active. Repressed TADs and lamina-associated domains (LADs) are often found at the nuclear lamina. Active TADs are usually distal from the nuclear lamina. Right: TADs are demarcated by binding sites for CCCTC-binding factor (CTCF), with the cohesin complex encircling DNA. Within TADs, cohesin-mediated loops regulate further chromatin folding and interactions, modulating interactions between genes and regulatory elements. Adapted from [11], based on modification of the "Chromosome Organization in Nucleus: TADs" template by BioRender.com (2024), retrieved from `https://app.biorender.com/biorender-templates`.

domains described [18]. Even within a single gene locus, some lengths of DNA may be in the A compartment whilst others are found in the B [18]. Analysis at this resolution suggested that active promoters and their distal enhancers are almost ubiquitously localised in the A compartment, regardless of TAD/chromatin loop localisation, and that CTCF-mediated loops are much more diffuse than previously observed [18]. As sequencing capacity increases, it is likely a more granular understanding of genome organisation will emerge, but current research largely supports the hierarchical organisation described.

1.1.4 Chromatin modifications

Underlying the broad EC versus HC definition of chromatin is an incredibly complex chromatin landscape. Indeed, as many as 51 distinct chromatin states have been identified in human cells [19]. These many chromatin states are regulated by epigenetic modifications.

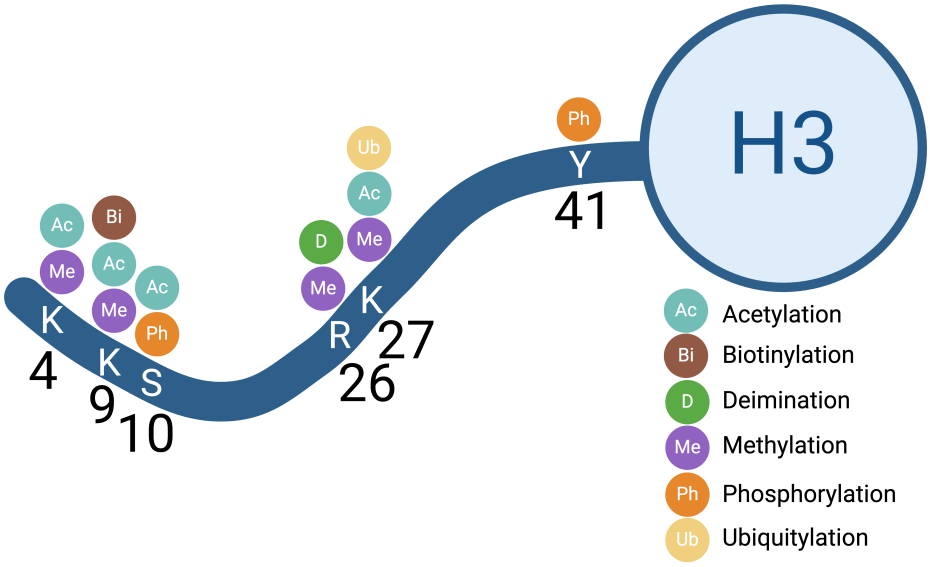


Figure 1.4 **Histone modification schematic.** Examples of some of the many potential modifications of the H3 histone tail. K = lysine, S = serine, R = arginine, Y = tyrosine residues. Numbers indicate the amino acid position number on the histone tail. Created using BioRender.com, adapted from [24].

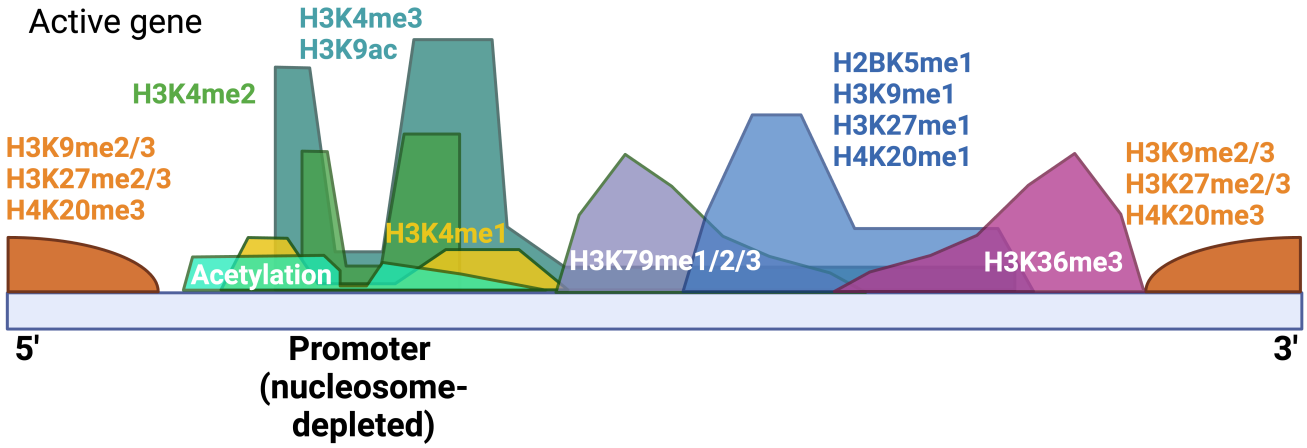
Epigenetic modifications may be added either to the underlying DNA sequences or to the core histone proteins around which the DNA is wrapped [20]. Functionally, the addition of modifications affects histone-DNA and histone-histone interactions, as well as the recruitment of chromatin-binding proteins, thereby altering local chromatin environments and the higher-order structure of chromatin [21]. The histone octamer may be modified in an enormous number of ways: in addition to variants of the histone proteins themselves, each of the histone proteins has an N-terminal "tail" that can be modified in more than a dozen positions [19] and with at least 8 distinct types of modification [2] (figure 1.4).

The relationship between specific modifications and associated transcriptional states was hypothesised to represent a histone or nucleosome 'code' [22]. Attempts to decipher these codes were accelerated by the improvement of genomic technologies, yet revealed more of a combinatorial and context-dependent language than a precise code [23]. However, general associations between transcriptional states and certain epigenetic marks have been identified (figure 1.5).

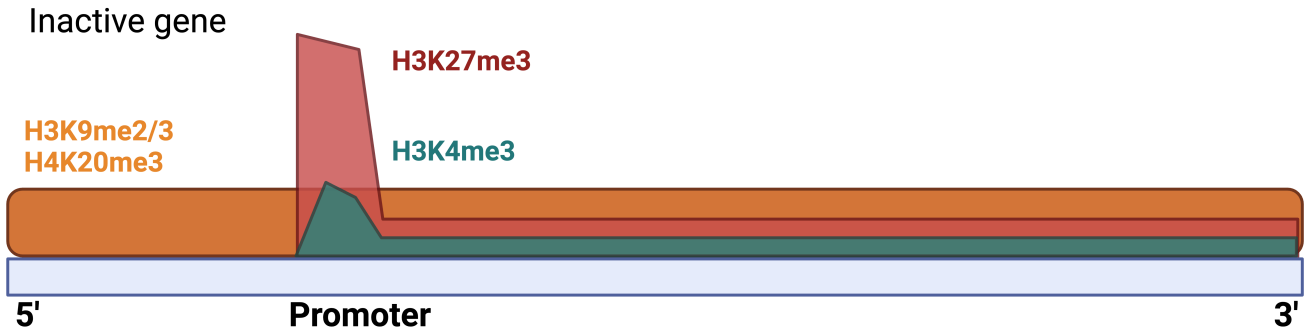
A prototypical example of this is the acetylation of histone tails as a marker of active chromatin. Acetylation can neutralise the basic charge of lysine residues and therefore promote de-condensing of the chromatin, via reduced histone-DNA contacts [26]. In addition to direct effect on histone-DNA interactions, acetylation provides a binding site for transcription factors (TFs) containing bromodomains, which can promote remodelling of nucleosomes contributing to chromatin accessibility [26].

In contrast, the transcriptional silencing associated with heterochromatin is centred around three essential systems: DNA methylation at cytosine residues, methylation at lysine 9 of the H3 histone (H3K9me) and methylation at lysine 27 of the H3 histone (H3K27me) [27]. These modifications do

not alter histone charges, but can instead mediate repressive effects via recruitment of chromatin-binding proteins that promote chromatin compaction, as well as via recruitment of further repressive epigenetic modifiers [26]. This feedback system results in a condensed chromatin state with reduced accessibility for transcriptional machinery, thereby reducing gene expression. The genomic effects, oncogenic potential, and therapeutic targeting of the H3K9me and H3K27me histone marks will be the focus of this thesis, with a particular concentration on the EZH2 and G9a enzymes responsible for their deposition.



(a)



(b)

Figure 1.5 **Histone modifications associated with active or inactive genes.** Localisation and type of histone modifications associated with (a) actively transcribed genes and (b) inactive, silent genes. 5' = 5 prime end, 3' = 3 prime end. H = histone, K = lysine, me1/2/3 = mono-/di-/tri-methylation, ac = acetylation. Figure adapted from [25].

1.2 Histone 3 lysine 9 methylation (H3K9me): G9a/GLP

1.2.1 Mammalian H3K9me enzymes: introduction to G9a/GLP activity and chromatin targeting

Mono-, di-, or tri-methylation of histone 3 lysine 9 (H3K9me1/2/3) can be placed by several H3K9 histone lysine methyltransferases (HKMTs) in mammals [28], via the transfer of a methyl group from S-Adenosyl-l-methionine (SAM) [29]. These HKMTs include Suppressor of Variegation 3-9 Homologue 1 (SUV39H1) and Suppressor of Variegation 3-9 Homologue 2 (SUV39H2), SET Domain Bifurcated 1 (SETDB1) and SET Domain Bifurcated 2 (SETDB2), G9a and G9a-Like Protein (GLP) [28]. Each of the H3K9 HKMTs contains the catalytic Su(var)3–9 Enhancer-of-zeste Trithorax (SET) domain responsible for methylation of H3K9 (figure 1.6), and there is considerable interplay and redundancy between the enzymes [28]. Moreover, SUV39H1, SETDB1, G9a and GLP have been observed to come together to form a 'megacomplex' and to have interdependency in their protein stability [30]. These features complicate the demarcation of specific roles for each of the individual HKMTs in the deposition of H3K9me1/2/3. Of note, aside from the HKMTs discussed, several PRDI-BF1 and RIZ homology domain (PRDM) proteins exist that can also contribute to H3K9me [31]. However, whether this H3K9me effect is direct or indirect remains unclear [31].

The G9a (or Euchromatic histone-lysine N-methyltransferase 2, EHMT2/KMT1C) and GLP (or KMT1D) HKMTs are ubiquitously expressed, with a knockout of either enzyme leading to embryonic

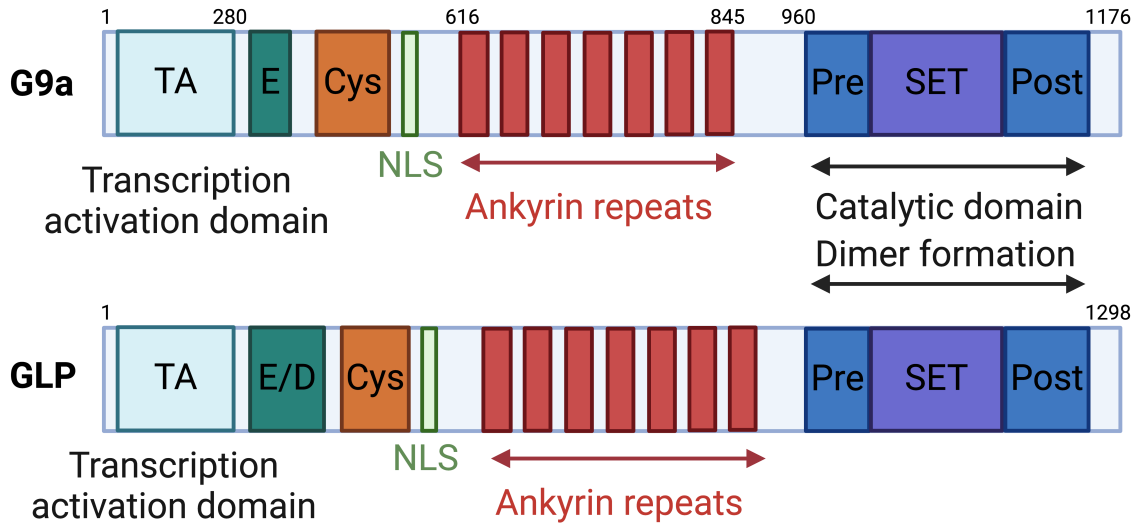


Figure 1.6 **Structures of G9a and GLP domains.** TA = transcription activation domain, E = E (glutamine)-rich domain, E/D = glutamine/aspartic acid-rich region, Cys = cysteine-rich region, NLS = nuclear localisation signal, Pre= pre-SET domain, SET = Su(var)3–9 Enhancer-of-zeste Trithorax domain, Post = post-SET domain. Number of amino acids indicated above domains. Note that there is 80% homology in the SET domains between G9a and GLP [32]. Figure adapted from [33] and [34].

lethality [35, 36]. Despite distinct activities from one another in certain cell types, the two enzymes are predominantly found as a G9a/GLP heterodimer [28], with G9a being the dominant protein responsible for *in vivo* H3K9 methylation [37].

The G9a/GLP heterodimer can place H3K9me1/me2 and is responsible for the majority of H3K9me2 deposition in mammalian genomes [38]. H3K9me1 provides a crucial template for di- and tri-methylation, as the substrate for further methyltransferase activity by H3K9 HKMTs [28, 39]. Large domains of H3K9me2 exist [28], but the processive nature of H3K9 methylation can result in conversion of H3K9me2 to H3K9me3 [40]. Additionally, several reports suggest direct H3K9me3 activity of G9a/GLP [41]. Hence, G9a/GLP can contribute to accumulation of each of the mono-, di- and tri-methylation marks at H3K9.

Importantly, G9a/GLP does not have direct DNA-binding capacity; instead, recruitment to chromatin is achieved by recognition of H3K9me0/1 (via their SET domains) or H3K9me1/2 (via ankyrin repeats) [42]. Additionally, G9a can be recruited by TFs such as zinc finger proteins (ZFPs), via chromatin-binding proteins such as chromodomain Y-like protein (CDYL), or via non-coding RNAs such as KCNQ1 overlapping transcript 1 (*KCNQ1OT1*) [41].

1.2.2 H3K9me: function in gene expression and chromatin organisation

H3K9me2/3 is considered a hallmark of heterochromatin; its repressive effects are essential for the silencing of genes, transposable elements and satellite repeats [28]. Gene targets of H3K9me are diverse, including those involved in regulation of processes such as development, pluripotency, cell cycle regulation and DNA damage [29, 35, 36, 43].

In the heterochromatic B compartment, H3K9me2/3 modifications are recognised by chromatin-binding proteins such as Heterochromatin Protein 1 (HP1) family members [44, 45]. HP1 proteins can subsequently act in a feedback loop to recruit further H3K9me2/3 deposition and heterochromatin via self-oligomerisation and recruitment of additional H3K9 HKMTs or DNA methyltransferases (DNMTs), alongside enzymes responsible for removal of active histone marks such as histone deacetylases (HDACs) [26, 46–49]. The feedback nature of this interaction allows spreading of broad heterochromatic domains from primary H3K9me sites [46]. H3K9me is also implicated in the spatial segregation of inactive chromatin domains from active regions; LADs are enriched for H3K9me2/3, which has now been proven to be essential for tethering of these repressive regions to the nuclear periphery [38, 50, 51].

Whilst these effects regulate the persistent heterochromatin associated with the B compartment, H3K9me is also essential in the more dynamic A compartment. Indeed, loss of the G9a/GLP complex results in loss of H3K9me throughout euchromatin regions of the genome [36]. At the level of individual gene repression, H3K9me effects are similar to those described in the B compartment, with

recruitment of many repressive chromatin modifiers including HP1 and DNA methylation marks [33], whilst also blocking deposition of activating acetylation marks at the H3K9 residue [52]. However, in contrast to cHC regions, the spreading of H3K9me is limited in the A compartment using various mechanisms [46]. These include inhibition of H3K9me deposition by active chromatin marks, erasure of H3K9me marks, and depletion or turnover of nucleosomes [46]. In addition to repression at individual gene loci, higher-order chromatin structures are also regulated by H3K9me in the A compartment. Specifically, H3K9me at TAD or loop anchor sites may prevent CTCF and cohesin binding, disrupting TAD formation as well as controlling the strength of chromatin looping [28, 53] (figure 1.3), thereby influencing the expression of multiple genes simultaneously.

1.2.3 Non-canonical G9a effects

G9a is known to methylate non-histone targets alongside the H3K9 residue. The effects of methylation in this instance may be either repressive or activating [54]. Intriguingly, one non-canonical activity of G9a is to methylate itself: automethylation of G9a allows binding of the HP1 protein [55], allowing G9a to act as a coactivator for the glucocorticoid receptor via recruitment of RNA polymerase II (RNAPII) and other activator proteins [56]. Phosphorylation of the threonine 186 residue on G9a inhibits its co-activator function, regulating repressive versus activating G9a effects [56].

Additional non-histone methylation substrates for the G9a/GLP complex are the chromatin remodelling factors Pontin and Reptin, methylated by G9a under hypoxic conditions [57, 58]. Methylated Pontin increases recruitment of the epigenetic coactivator p300 to the Hypoxia Inducible Factor 1 Alpha (HIF1A) protein, leading to strong transcriptional activation of a subset of HIF1A target genes [58]. Yet, methylation of Reptin has the converse effect: Reptin is recruited to HIF1A target genes and results in transcriptional repression, via recruitment of co-repressors such as HDACs [57]. Several other non-canonical G9a effects have been identified in different cell types [54], including those independent of the catalytic activity of G9a [53, 59]. Hence, G9a may affect histone and non-histone targets, dependent or independent of methylation, with potential for activation or repression effects.

1.2.4 H3K9 demethylation

Methylation was initially thought to be a permanent epigenetic mark [31]. However, this has since been proven untrue. Two different families of histone lysine demethylases (KDMs) exist, distinguished by their catalytic mechanism: amine oxidases and Jumonji C (JmjC)-domain enzymes [60]. The KDM1 family, also known as lysine-specific demethylase (LSD), utilises amine oxidation with a flavin adenine dinucleotide (FAD) cofactor for demethylation [60]. The KDM1 family includes

KDM1A (LSD1) and KDM1B (LSD2), enzymes that were initially recognised as H3K4 demethylases [60]. However, a range of cofactors such as the androgen receptor (AR) allow a broader target range for demethylation activity, including H3K9me1/2 [60].

The second family is much larger, consisting of KDMs with demethylase activity towards several H3 and H4 lysine residues: substrates include H3K4, H3K9, H3K27, H3K36 and H4K20 [60]. Six subfamilies exist within this family (KDM2-7), all of which have the catalytic JmjC-domain, with dioxygenase activity allowing demethylation of substrates. H3K9me1/2 is demethylated by all KDM3 subfamily members (also known as JMJD1, or JHDM2), including KDM3A/3B/3C, and by the KDM7 subfamily members KDM7B/7C (PHF8/PHF2) [60]. Importantly, the KDM7B (PHF8) enzyme has a PHD domain which recognises H3K4, and this enzyme therefore contributes to the mutual exclusivity of H3K9 and H3K4 methylation [31]. H3K9me2/3 can be demethylated by all members of the KDM4 (JMJD2/JHDM3) subfamily (KDM4A/4B/4C/4D/4E) [60]. Non-histone substrates can also be demethylated by KDM enzymes [60]. Hence, H3K9 methylation status and other G9a targets are regulated by the opposing action of numerous H3K9 methylase and demethylase enzymes.

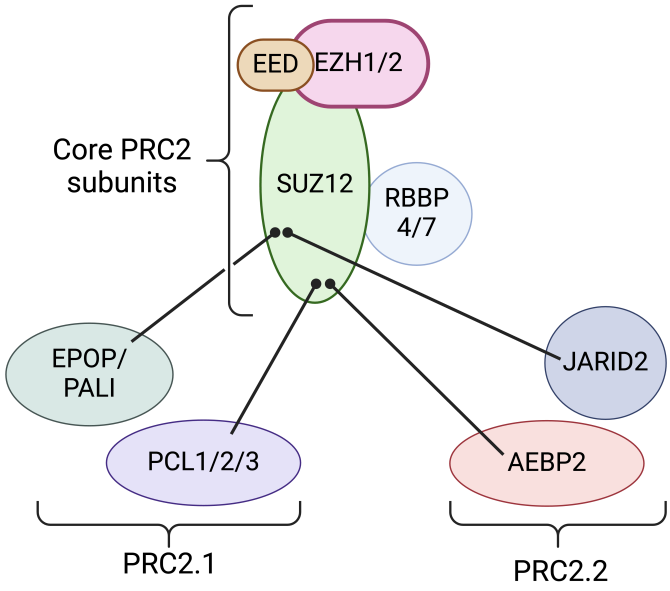
1.3 Histone 3 lysine 27 methylation (H3K27me): EZH2

1.3.1 Mammalian H3K27me enzymes: introduction to EZH1/2 activity and chromatin targeting

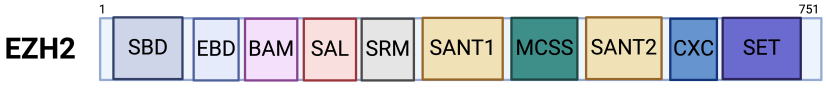
The Polycomb Repressor Complex 2 (PRC2) is responsible for deposition of H3K27me1/2/3 [61]. PRC2 consists of a minimal core of three stoichiometric proteins: the SET-domain-containing Enhancer of Zeste Homologue 1 (EZH1)/Enhancer of Zeste Homologue 2 (EZH2), the SUZ12 polycomb repressive complex 2 subunit (SUZ12) and the WD40 protein Embryonic Ectoderm Development (EED) [62, 63]. EZH1/EZH2 provides the catalytic activity in the context of PRC2, via a C-terminal SET domain [62, 63] that transfers a methyl group from the SAM donor utilised by G9a/GLP [64]. EZH2 is highly dependent on SUZ12 and EED for enzymatic activity, and depletion of any one of these three proteins is sufficient for embryonic lethality [63]. In contrast, EZH1 is dispensable for normal development [65]. The specifics of EZH1 versus EZH2 activity are debated: PRC2-EZH1 complexes have lower methyltransferase activity compared to their PRC2-EZH2 counterparts which are responsible for the majority of H3K27me deposition [66], hence *EZH1* knockout has a limited effect on H3K27me levels [65]. However, EZH1 appears to have chromatin compaction effects even in the absence of SAM and is able to bind nucleosomes and DNA more strongly than EZH2 [66]. Additionally, whilst EZH2 expression is highest in proliferative cells and tends to decrease with differentiation, EZH1

expression is higher in differentiated non-proliferative tissue [65, 66]. These findings support distinct roles for the EZH1 and EZH2 enzymes within the PRC2 complex, yet the specifics of such roles remain poorly defined.

The Retinoblastoma-Binding Protein (RBBP)4/7 comprises a fourth ubiquitous PRC2 component that exists at slightly lower stoichiometry [62]. Beyond these core components, the diverse activity of the PRC2 complex is governed by a large number of sub-stoichiometric factors and non-coding RNAs that have been shown to bind PRC2 in a context-dependent manner to modulate its activity and specificity [62]. Experimental evidence has revealed the existence of at least two PRC2 complex versions (named PRC2.1 and PRC2.2) dependent on the associated factors, that work both independently and synergistically in their deposition of H3K27me3 (figure 1.7) [67].



(a)



(b)

Figure 1.7 **PRC2 and EZH2 structure and organisation.** (a) Schematic of PRC2 structure, including the alternative cofactors associated with PRC2.1 and PRC2.2 complex versions. EED = Embryonic Ectoderm Development, EZH1/EZH2 = Enhancer Of Zeste Homologue 1/2, SUZ12 = SUZ12 polycomb repressive complex 2 subunit, RBBP4/7 = Retinoblastoma-Binding Protein 4/7, EPOP = Elongin BC and PRC2 associated protein, PALI = Phylogeny and ALIgnment of homologous protein structures, PCL1/2/3 = Polycomblike 1/2/3, JARID2 = Jumonji And AT-Rich Interaction Domain Containing 2, AEBP2 = AE Binding Protein 2, PRC2 = Polycomb Repressive Complex 2. (b) EZH2 domain organisation. SBD = SANT1-binding domain, EBD = the EED-binding domain, BAM = β -addition motif, SAL = SET activation loop, SRM = stimulation-responsive motif, MCSS = motif connecting SANT1 and SANT2, SANT = SWI3, ADA2, N-CoR and TFIIIB DNA-binding domain, CXC = cysteine-rich domain, SET = Su(var)3–9 Enhancer-of-zeste Trithorax domain. Number of amino acids indicated above domains. (a) Adapted from [61]. (b) Adapted from [68].

Similarly to G9a/GLP, the PRC2 complex does not have direct DNA-binding capacity [42]. Yet, PRC2 binds DNA preferentially at gene promoters and regulatory elements [69], evidencing the existence of a specific targeting mechanism. A small subset of this targeting is achieved via DNA-binding transcription factors such as REST and SNAIL, but this cannot account for the majority of PRC2 targeting [69]. Other mechanisms include binding to non-coding RNAs and CpG island (CGI) regions, alongside recognition of H3K27me3 marks by the EED subunit of PRC2 [69]. This latter mechanism also activates PRC2 enzymatic activity, contributing to the spread of H3K27me3 domains [69]. Intriguingly, both PRC2 (including the EZH2 subunit) and G9a/GLP proteins have been proposed to be inhibited by active transcription and enriched at chromatin upon transcriptional inhibition, suggesting that nascent RNA may also regulate chromatin recruitment of histone modifiers [70].

1.3.2 H3K27me: function in gene expression and chromatin organisation

H3K27me3 is by far the best described H3K27me mark [71]. However, it is not the most abundant: whilst H3K27me3 and H3K27me1 exist at only 5-15% H3 histones [63, 72], H3K27me2 is found at up to 70% [71]. Given the processive nature of H3K27 methylation [72], H3K27me2 was initially thought to be solely an intermediate product prior to H3K27me3 [71]. However, it has recently been shown to act as a repressive mark, preventing non-cell-type-specific enhancer firing via protection from the activating H3K27 acetylation mark [71]. Similarly, literature describing the role of H3K27me1 is emerging: H3K27me1 accumulation across gene body regions positively correlates with active transcription [71], whilst its deposition elsewhere is associated with the stably repressive cHC [72].

The predominant role of H3K27me3 is in transcriptional repression, and it is considered a hallmark feature of fHC [73]. Indeed, its distribution is mostly distinct from histone modifications associated with active transcription [72]. It is found enriched at CGI promoters of silent genes [61] as well as in repressive LADs [16], with targets including many developmental genes as well as tumour suppressor genes [62, 74]. In addition to focal peaks of H3K27me3 at gene promoters, H3K27me3 has been described to form large domains over silent genes and intergenic regions [75], as well as over repetitive elements such as short-interspersed nuclear elements (SINEs) [75].

The silencing effects of H3K27me may be achieved via recruitment of additional repressive chromatin factors, or via indirect or direct hindrance of chromatin-binding proteins with activating properties [72]. For example, PRC2 can interact with RNA polymerase enzymes and elongation factors to regulate transcription, and can antagonise histone acetyltransferase enzymes such as p300/CBP to prevent deposition of the activating H3K27Ac mark [76]. Of note, repressive activity by PRC2

is enhanced by the closely related Polycomb Repressor Complex 1 (PRC1). H3K27me deposited by the PRC2 complex may provide a docking site for PRC1, which contributes to additional chromatin repression via mono-ubiquitination of H2AK119 (H2AK119ub1) [72]. Complicating this picture, H2AK119ub1 has also been described to recruit PRC2 to target sites and to stimulate its enzymatic activity [63], illustrating the interdependency of PRC1 and PRC2 activities and contributing to the spreading capacity of repressive H3K27me3 chromatin domains [69].

The role of H3K27me in chromatin architecture is less evident than described for H3K9me, but recent reports have begun to delineate these effects. For example, PRC2 has been described to mediate long-range chromatin loops spanning 10-100 megabases and across multiple TADs within the same A/B compartment [77]. These are thought to be anchored by key PRC2 developmental gene targets and to function in coordinated gene silencing over large genomic distances, promoting spreading of H3K27me3 [77]. PRC2 has also been evidenced to mediate looping between enhancer elements and target genes in embryonic stem cells (ESCs) [78], or alternatively between silencer elements and target genes [79].

1.3.3 Non-canonical EZH2 effects

Non-canonical EZH2 activity includes methylation of non-histone proteins, PRC2-independent EZH2 activity and transactivation effects [64]. For example, EZH2 is responsible for methylation of the TF GATA4 during development [80]. This non-histone methylation reduces the interaction between GATA4 and the coactivator p300, contributing to transcriptional repression in an H3K27me3-independent manner [80]. EZH2 has also been demonstrated to activate gene expression, including in a manner independent of PRC2. This is exemplified by the effects of EZH2 on the androgen receptor (AR) [81]. EZH2 can directly occupy the AR promoter and promote its expression, without requirement for other components of the PRC2 complex and not dependent on methylation [81]. The activity of EZH2 as a transcriptional activator or repressor may depend on numerous variables, such as the local chromatin environment including presence of H3K27Ac versus H3K27me3 marks, or the existence of TF motifs for co-activators [81]. Another non-canonical role for EZH2 is found in a different developmental process: EZH2 has been shown to regulate the development of natural killer T (NKT) cells, via methylation of a lineage-defining TF that promotes its ubiquitination and degradation [82]. Hence, non-histone targets of EZH2 methylation and methylation-independent effects play important physiological roles alongside H3K27me3-mediated transcriptional regulation.

1.3.4 H3K27 demethylation

H3K27 di- and tri-methylation is mostly removed by two JmjC demethylase enzymes: UTX (KDM6A) and JMJD3 (KDM6B) [83]. UTX and JMJD3 catalyse the conversion of H3K27me3 to H3K27me2 and H3K27me2 to H3K27me1 [83]. UTX may also couple the demethylation of H3K27me2/3 with the methylation of H3K4 to promote transcriptional activation, via association with the Mixed-Lineage Leukemia (MLL) 2/3 histone methyltransferase complexes [84]. A third member of the same JmjC subfamily is the UTY (KDM6C member). However, this has not been demonstrated to demethylate H3K27me3 *in vitro*; it is unclear whether it requires additional cofactors for functionality, or plays a distinct role from other JmjC members [83].

H3K27me1 demethylation mechanisms remain obscure; UTX and JMJD3 were able to demethylate H3K27me1 *in vitro* but failed to do so *in vivo* [84]. The KDM7A (KIAA1718) enzyme has been reported as a dual demethylase of both H3K9me2 and H3K27me2 simultaneously, with potential additional activity at H3K27me1 [84]. Hence, UTX and JMJD3 appear to be the major erasers of H3K27me2/3 methylation placed by EZH2, whilst erasure of H3K27me1 is still under debate.

1.4 Epigenetic crosstalk and interdependency

1.4.1 Crosstalk between histone methylation and DNA methylation

The H3K9 and H3K27 lysine residues both exist within Alanine-Arginine-Lysine-Serine (ARKS) motifs [85]. This motif provides a clear example of the vast potential for crosstalk between different epigenetic marks: methylation of the preceding arginine or phosphorylation of the succeeding serine residue are both suggested to impact the ability of EZH2 or G9a to methylate their respective lysine targets, potentially via steric hindrance [86]. For example, H3S10 phosphorylation appears to prevent H3K9 methylation by G9a, and similarly H3S28 phosphorylation and H3K27me are thought to be mutually exclusive [86]. Whilst this describes antagonistic crosstalk between epigenetic modifications, H3K9 and H3K27 methylation marks also act in concert with other inhibitory epigenetic marks to achieve their repressive effects. One such mark is DNA methylation, a ubiquitous modification that is essential for transcriptional control of genes and repetitive elements [87].

DNA methylation is achieved by the transfer of a methyl group to the fifth carbon of cytosines (5-methylcytosine, or 5mC) by DNA methyltransferases (DNMTs) [87]. There are three major DNMTs, including DNMT1, DNMT3A and DNMT3B [88]. These enzymes utilise SAM to methylate target cytosines, either *de novo* (DNMT3A/B) or in the context of hemi-methylation (DNMT1) for methylation maintenance [88].

Whilst DNA methylation can occur at any cytosine base, over 98% of DNA methylation oc-

curs at symmetrical CpG dinucleotides (cytosine nucleotide followed by a guanine nucleotide) in somatic mammalian cells [89]. DNA methylation is broadly associated with transcriptional repression; however, this effect is dependent on the localisation of DNA methylation, with potential for transcriptional activation effects in certain contexts [90]. Importantly, DNA methylation is excluded from particular CpG-dense regions at gene promoters called CGIs [88]. Roughly 60% of promoters and first exons are associated with CGIs [91] and their hypomethylation allows destabilisation of nucleosomes, increased chromatin accessibility and binding of transcription initiation factors to facilitate gene expression [88]. The repressive effect of DNA methylation is achieved via both exclusion of transcriptional (co-)activators and the recruitment of methyl-binding proteins that can recruit further repressive effectors [91], including many chromatin remodellers [87]. This includes histone lysine methyltransferases; both H3K9me and H3K27me are strongly associated with DNA methylation [89, 92]. EZH2 and G9a have been shown to bind all three DNMTs directly [89, 93], with additional indirect interactions, for example via the replication complex or chromodomain proteins [93]. Additionally, knockout of either *EZH2* or *G9a* results in DNA hypomethylation of certain target loci of these enzymes [89]. In some instances, histone methylation by EZH2 or G9a is thought to be an intermediate silencing mechanism ahead of permanent repression by DNA methylation. For example, H3K27me3 at promoters in stem cells is often replaced by DNA methylation during differentiation, which may achieve stable repression of target genes [89].

The close co-regulation between histone and DNA methylation is in part achieved by overlap in recruitment mechanisms. This is particularly evident for H3K9me placed by G9a. For example, HP1, which both recruits G9a and is recruited by H3K9me, is able to interact with and recruit DNMTs [89]. Another example is that of Ubiquitin Like With PHD And Ring Finger Domains 1 (UHRF1), a coordinating protein involved in DNMT1 recruitment to the DNA replication fork. UHRF1 binds hemi-methylated DNA, yet also exists in a complex with G9a, thereby allowing coordinated deposition of DNA and histone methylation during DNA replication [89]. Hence, DNMTs and HKMTs utilise crosstalk and feedback mechanisms to cooperate for gene silencing.

1.4.2 Bivalency: co-existence of activating chromatin marks with H3K27me/H3K9me

Whilst H3K9me and H3K27me are generally associated with repressive chromatin, this is not universally true. Both marks have been found to co-exist with activating epigenetic marks, creating a chromatin state described as bivalent, or poised [94, 95].

The best described bivalent chromatin exhibits H3K27me3 with the activating H3K4me3 or H3K4me1 marks, associated with bivalent promoters or enhancers [96, 97]. In fact, recent obser-

vations have suggested that many bivalently marked promoters may instead be trivalent, with each of these three histone modifications present simultaneously [97]. The purpose of these domains has been debated. Initial hypotheses centred around the ability to allow rapid transcriptional activation or repression of genes following removal of repressive or activating marks [94, 97–99], whilst several recent observations have suggested that instead its role is to prevent permanent transcriptional silencing via DNA hypermethylation [100].

Physiologically, bivalent chromatin at H3K27me3 is an important feature of ESCs, allowing a pluripotent cell to follow different routes of differentiation depending on cell fate [94, 96, 98, 99, 101]. Bivalent enhancers can physically interact with bivalent promoter regions to regulate developmental gene expression, preferentially within the same TADs [96]. Although initially thought to be ESC-specific, bivalency has now also been identified in terminally differentiated cells [100], with roughly 40% of ESC bivalent regions being retained after differentiation [89].

Bivalently marked H3K9me chromatin is less well established than that of H3K27me. It is mostly found in euchromatic DNA, as is observed with the combination of H3K9me3 and the activating H3K36me3 mark [102]. These domains have been found at genes with low levels of transcription associated with processes such as development and cell cycle transition [103], and have also been evidenced to play a role in poising enhancer elements [102, 103]. Aside from H3K36me3 bivalency, H3K9me3 can also be found in combination with H3K4me3. H3K9me3-H3K4me3 bivalent regions have been identified in lineage-committed mesenchymal stem cells and pre-adipocytes, poising transcription of two master regulator TFs involved in the subsequent commitment to an adipocyte cell fate [95]. Maintenance of H3K9me3 at each of these dual-marked regions is thought to depend on SETDB1 activity [95, 102, 103], although the interplay between H3K9 HKMT activities and the physical interactions between the enzymes suggest the possibility of a role for G9a.

Generally, bivalent modifications imply multiple modifications at different histone proteins within the same nucleosome or spanning a short stretch of DNA, as the existence of alternative modifications on the same histone residue were thought to be mutually exclusive [104]. However, recent findings demonstrated both acetylation and methylation at the same histone H4 lysine residue (named Kacme), suggesting the possibility of an added layer of complexity in bivalently marked chromatin [104].

1.4.3 Interactions between EZH2 and G9a

Whilst bivalent/poised chromatin generally refers to active histone modification in combination with either H3K27 or H3K9 methylation, recent literature has reported enhancers co-occupied by all three modifications: H3K9me3, H3K27me3 and the active H3K4me1 mark [105]. Both H3K27me3 and H3K9me3 must be removed, alongside gain of H3K27Ac, for activation of the poised enhancers

[105]. The co-regulation of genomic elements by dual H3K9 and H3K27 methylation is not limited to this subset of poised enhancers; the two repressive histone marks are known to significantly co-localise genome-wide [106]. Several groups have identified genes repressed by dual H3K9me/H3K27me, including those involved in development, differentiation, apoptosis and chemokine signalling [106–108] amongst other targets [109].

Inhibition of both EZH2 and G9a enzymes simultaneously induces gene expression to a greater extent than single inhibition [110] and additionally allows re-expression of genomic targets that are not affected by single inhibition [107, 109]. Further, knockout of either *G9a* or *GLP* has been shown to reduce global levels of H3K27me1 [111], whilst also decreasing H3K27me3 and EZH2 occupancy at a subset of PRC2 target genes [106]. These effects may be explained by multiple levels of direct and indirect EZH2-G9a interaction: co-localisation of EZH2 and G9a, physical interaction between the enzymes and overlap in their enzymatic activities.

Most basically, PRC2 and G9a/GLP have been shown to co-localise at promoters and CGIs [106] throughout the genome. Indeed, one third of G9a targets have been shown to be bound by PRC2 components, whilst roughly 10% of PRC2 targets are bound by G9a [106]. This may in part be due to overlap in chromatin recruiting mechanisms for PRC2 and G9a: sequence-specific TFs such as SNAIL, REST, CDYL and Msx1 have been identified as recruitment factors for both PRC2 and G9a, and additional factors such as the PRC2 cofactor Jumonji And AT-Rich Interaction Domain Containing 2 (JARID2) and multiple long non-coding RNAs (lncRNAs) are also recognised as common chromatin-targeting factors [42]. Additionally, the PRC2 subunit EED is able to recognise H3K9me3 for chromatin targeting of the complex [85], whilst H3K27me3 has been shown to stabilise HP1, a key recruitment protein for G9a/GLP [112]. Hence, overlapping recruitment mechanisms in combination with positive feedback loops may contribute to co-localisation of EZH2 and G9a.

Moreover, PRC2 and G9a physically interact both *in vitro* and *in vivo*, in a manner independent of DNA or RNA [106]. The contact may occur via cofactor interactions: the PRC2 components SUZ12, EED, JARID2 (PRC2.1 cofactor) [106] and PALI1 (PRC2.2 cofactor) [113] have all been shown to interact with G9a/GLP. For example, a G9a-PALI1-PRC2 super-complex has been described in prostate cancer with both H3K9 and H3K27 methyltransferase activity [113]. An additional level of EZH2-G9a interaction is evidenced *in vitro* but not yet *in vivo*, whereby EZH2 is capable of directly modifying G9a via methylation at the lysine 185 residue, a modification that promotes increased G9a repressive effects [114].

Finally, redundancy in EZH2 and G9a enzymatic activity adds to the complexity of the relationship between H3K9me and H3K27me marks. G9a is known to place the H3K27me mark *in vitro* and *in vivo* [111, 115]. EZH2 has also been shown to have H3K9me capacity *in vitro* [116], although *in vivo* data have not yet confirmed this effect. The promiscuity of these enzymes may be explained by

the fact that the lysine residues of both H3K9 and H3K27 are embedded in ARKS motifs, providing a similar substrate that may allow recognition of both substrates by the same enzyme [85]. Hence, EZH2 and G9a methyltransferase activities and downstream effects are intricately linked.

1.5 EZH2 and G9a in cancer

1.5.1 Dysregulation and prognostic significance

Epigenetic alterations are increasingly understood to play a crucial role in tumourigenesis [117]. Indeed, they have been linked to every single hallmark of cancer [118]. The EZH2 and G9a enzymes typify such roles: they contribute to the initiation, progression, metastasis and chemoresistance of various tumour types [54, 64]. Alterations such as mutation, amplification, or overexpression of either enzyme are found in many tumour types (figures 1.8 and 1.9) and higher expression has been associated with poorer prognosis in several cancers [54, 119]. In addition, *EZH2* and *G9a* alterations significantly co-occur (appendix table A.1) and their expression levels have been shown to significantly correlate with each other in breast, ovarian, melanoma and bladder tumours and cell line models, suggesting dysregulation of the two enzymes may be functionally linked [107, 120].

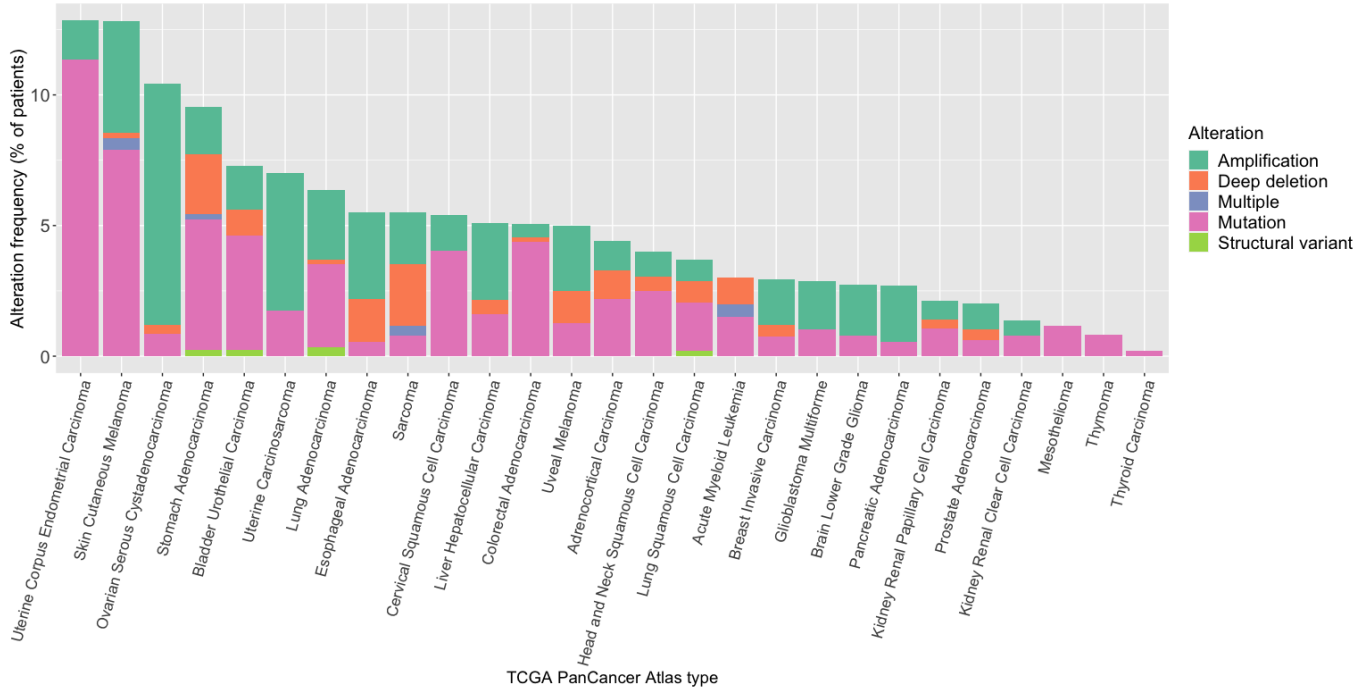


Figure 1.8 ***EZH2* and *G9a* genomic alterations in cancer.** Bar chart illustrating combined alterations in either *EZH2* or *EHMT2* (*G9a*) across The Cancer Genome Atlas (TCGA) cohorts with a minimum of 50 cases. Mutations shown in green, structural variants in purple, amplifications in red, deep deletions in dark blue and multiple alterations in grey. No annotation regarding the significance of *EHMT2* mutations in the cohort is available. 4/168 *EZH2* mutations are annotated as gain-of-function, 32 as loss-of-function and the remaining 132 are mutations of unknown significance or are lacking annotation. Figure produced using cBioPortal data [121–123].

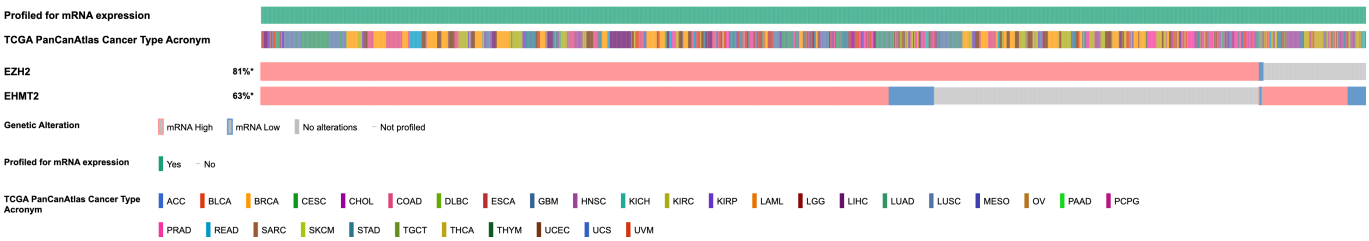


Figure 1.9 ***EZH2* and *G9a* aberrant expression in cancer.** Diagram illustrating aberrant expression of either *EZH2* or *EHMT2* (*G9a*) across The Cancer Genome Atlas (TCGA) cohorts with matched normal samples. High messenger RNA (mRNA) expression is shown pink, low mRNA expression in light blue. High and low are defined by z-score threshold ± 1. Figure produced using cBioPortal [121–123].

1.5.2 Oncogenic mechanisms of action

The various mechanisms by which EZH2 and G9a contribute to carcinogenesis reflect their diverse physiological effects: oncogenic roles may be the result of aberrant histone methylation-mediated transcriptional effects, disrupted genome organisation or non-canonical effects [64, 124].

Most basically, histone methylation by EZH2 and G9a can silence tumour suppressor genes such as those involved in regulation of the cell cycle, immune signalling, apoptosis, chemoresistance and differentiation [64, 124, 125]. For example, G9a can place mono-, di-, or tri-methylation to silence the deubiquitylase USP37 [126]. USP37 is a crucial modulator of cell cycle progression, regulating stability of the cyclin-dependent kinase inhibitor p27 amongst other proteins. Hence, depletion of G9a can induce cell cycle arrest and apoptosis [126]. Immune suppression via transcriptional silencing is demonstrated by EZH2-mediated methylation; EZH2 can silence Major Histocompatibility Complex I (MHC-I) expression [127] as well as the Th1-type chemokines CXC Motif Chemokine Ligand 9 (CXCL9)/CXC Motif Chemokine Ligand 10 (CXCL10) [128], both of which may promote escape from immune recognition by reducing antigen presentation and immune cell infiltration [127].

Additionally, transcriptional silencing is hypothesised to contribute to the existence of cancer stem cell and chemoresistant populations. This has been demonstrated in prostate cancer, where the G9a-PALI1-PRC2 complex represses genes associated with a differentiated state, promoting stem cell-like gene expression patterns [113]. Similarly, drug resistance mediated by a reservoir of drug-tolerant persister cells in multiple cancer models have been shown to depend partly on EZH2- and G9a-mediated H3K27me and H3K9me and the associated transcriptional silencing effects [114, 129]. Hence, transcriptional silencing via H3K9/H3K27 methylation contributes to multiple hallmarks of tumourigenesis.

In addition to repressive effects of H3K27me/H3K9me marks, the plasticity afforded by bivalent H3K27/H3K9 chromatin is exploited by cancer cells. For example, promoter bivalency has been shown to resolve into a hypermethylated state following chemotherapy in ovarian cancer models, resulting in gene silencing and potentially contributing to drug resistance [130]. A similar effect has

been proposed in the development of resistance to immunotherapies, via enabling immune evasion [127]. In this case, aberrant PRC2 activity allows maintenance of bivalent H3K4me3/H3K27me3 domains, where resolution of bivalent chromatin would usually allow antigen presentation and immune recognition [127]. Other tumourigenic processes are also evidenced to rely on bivalent chromatin. In melanoma, bivalently marked H3K4me3/H3K27me3 domains – including promoter and enhancer elements – are implicated in metastasis, via regulation of epithelial-to-mesenchymal TF expression such as ZEB1, TWIST1, SNAI1 and CDH1 TFs [131]. In fact, most hypermethylated genes in adult cancers are those that are bivalently marked in ESCs [100], demonstrating the relevance of bivalent chromatin in tumourigenesis.

Whilst most oncogenic effects ascribed to EZH2 and G9a relate to their canonical effects at H3K27 and H3K9, several non-canonical roles contributing to carcinogenesis have also been described. These may involve transcriptional activation roles, methylation of non-histone targets, or non-enzymatic effects of EZH2 and G9a. Exemplifying this are the many roles of EZH2 elucidated in prostate cancer (PCa). First, EZH2 has been shown to drive transcription of the AR in PCa as described in section 1.3.3, in a manner that is PRC2- and methylation-independent and driven instead by occupancy of EZH2 at the AR promoter [81]. Second, EZH2 has been shown to act as a transcriptional activator in PCa and glioblastoma following phosphorylation at its serine 21 residue, targeting distinct genes from those repressed by H3K27me3 [132, 133]. In PCa, EZH2 performs this role via modulating AR recruitment to target sites; this activity relies on the methyltransferase activity of EZH2 but is PRC2-independent, and has been suggested to result from direct methylation of the AR by EZH2 [133]. Similarly, G9a/GLP has non-canonical oncogenic effects: it is able to methylate p53 at lysine 373 to inactivate this key tumour suppressor [134], and contributes to transcriptional activation or repression of hypoxia genes (via methlyation of Pontin and Reptin respectively, as described in section 1.2.3) promoting cancer cell migration and tumour growth [57, 58].

Finally, deregulation of HKMT enzymes may alter chromatin conformation on a larger scale. For example, crosstalk between G9a and the lysine demethylase KDM3B has been evidenced to regulate CTCF binding in mixed lineage leukaemia [135]. Dysregulation of this process promotes aberrant looping that results in amplification and rearrangement of the Mixed-Lineage Leukemia 1/Lysine Methyltransferase 2A (*MLL/KMT2A*) locus that characterises several leukaemia subtypes [135].

1.6 Targeting EZH2 and G9a as an anti-tumour approach

1.6.1 Single inhibition of EZH2 or G9a

There have been numerous pre-clinical investigations using inhibitors of EZH2 or G9a in cancer models. The majority of these compounds are focused on inhibiting enzymatic effects, although some alternative approaches including protein depletion are under investigation (table 1.1).

Inhibition of EZH2 or G9a individually has demonstrated anti-tumour efficacy in various models [54, 136], with a multitude of possible mechanisms of action reflecting the diverse roles of EZH2 and G9a in cancer. For example, in different models EZH2 or G9a inhibition has been proposed to block cell cycle progression [137, 138], induce apoptosis or autophagic cell death [139, 140], stimulate an immune response [141, 142], or reduce the survival of drug resistant populations [129].

Pre-clinical findings and promising results from clinical trials lead to FDA approval of the EZH2i tazemetostat for the treatment of follicular lymphoma and epithelioid sarcoma [143, 144]. Particularly, lymphomas are rendered sensitive to EZH2i due to frequent activating mutations in the catalytic SET domain of *EZH2* at the tyrosine 641 residue (e.g. Y641H, Y641S, Y641C, Y641F and Y641N), associated with increased EZH2 activity and H3K27me3 deposition [145]. However, there has been limited efficacy of single inhibition approaches in most solid tumour types, and in *EZH2*-wildtype tumours resistance can emerge [136]. For example, the EZH2i GSK126 was trialled in patients with advanced hematologic and solid tumours but terminated with insufficient clinical activity [146]. No G9a inhibitors are currently in clinical trials [147].

Target	**Mechanism**	**Examples of inhibitor(s)**
Global histone methylation	SAH hydrolase inhibitor	3-deazaneplanocin A (DZNep)
EZH2	SAM-competitive	**Tazemetostat (EPZ6438)**, **SHR2554**, **GSK126**, GSK343, **CPI-1205**, **CPI-0209**, **MAK683**, UNC1999, EPZ005687
EZH2	Disrupt EZH2-EED interaction	Astemizole, SAH-EZH2 (stabilised alpha-helix of EZH2 peptide), nifedipine, EED226, AZD9291, Wedelolacton
EZH2	Disrupt EZH2-SUZ12 interaction	A769662
EZH2	EZH2 degradation	GNA022
G9a	Substrate-competitive	UNC0638, UNC0642, BIX-01294, CM-272, A-366, E72, CT13, HKMTI-1-247, HKMTI-1-248
G9a	SAM-competitive	BRD4770, BRD9539, BIX-01338, Chaetocin
G9a	Unknown	DCG066, TM2-115, CPUY074001

Table 1.1 **Table summarising current EZH2 and G9a inhibition approaches.** Current EZH2 and G9a inhibition approaches, including mechanism of action and example compounds. Compounds in bold are those in clinical trials at the time of writing. Information from [64, 147, 148].

1.6.2 EZH2i or G9ai in combination with other epigenetic therapies

The efficacy of single inhibitors may be limited by the vast network of interactions exhibited by EZH2 and G9a (section 1.4.3). Hence, there has been investigation into combination therapy approaches to combat this effect. Clinical and pre-clinical investigations include EZH2i or G9ai in combination with epigenetic agents [136], immune-modulatory agents, DNA damage agents and hormone receptor agonists [147, 149].

For example, Bromodomain-containing Protein 4 (BRD4) is an epigenetic regulator protein capable of binding H3K27Ac and increasing recruitment of the RNAPII transcription machinery [136]. BRD4 is heavily implicated in carcinogenesis and there are several BRD4 inhibitors under development [136]. Combination of EZH2i and BRD4i demonstrate improved anti-tumour effects versus single agent use, potentially due to the increase in H3K27Ac observed in EZH2i resistance [150].

Another combination approach is that of EZH2i or G9ai with DNA methyltransferase inhibitors (DNMTi), based on literature demonstrating that EZH2 and G9a both cooperate with DNA methylation in various tumour models (section 1.4.1). Treatment with EZH2i or G9ai and a DNMTi have suggested a potential immune stimulatory effect. For example, in hepatocellular carcinoma cell lines, EZH2i and DNMTi induced immune response pathways including interferon (IFN)-related genes [151]. In ovarian models, Th1-type chemokines CXCL9-10 were induced by the same combination and contributed to increased effector T cell infiltration alongside reducing tumour burden [128]. Likewise, groups have shown that in bladder tumour models and cell lines, a combination of G9ai and DNMTi induces IFN-related immune signalling [120], whilst in ovarian cancer models a similar IFN-based immune response is induced [152]. In addition to pharmacological intervention, DNMTi in combination with knockdown of *EZH2* or *G9a* can induce IFN signalling in a variety of cancer cell lines [153].

Yet, these effects have not been ubiquitously observed. In cell line models of melanoma, breast and ovarian cancer, DNMTi in combination with either a G9ai or an EZH2i was not more effective than single inhibition alone [107]. Potentially, this lack of efficacy is driven by the extensive EZH2-G9a crosstalk and redundancy described (section 1.4.3), prompting investigations into combination approaches targeting these two HKMT enzymes.

1.6.3 Targeting EZH2-G9a co-operation: dual inhibition

In melanoma, breast, ovarian, rhabdoid, myeloma and prostate cell line models, use of two single EZH2 and G9a inhibitors has a greater growth inhibitory effect than either inhibitor individually [107, 109, 110, 113, 154]. Additionally, xenograft models have demonstrated reduced tumour growth in the context of dual EZH2/G9a inhibition in multiple tumour models [107, 110, 113, 154]. These

have included various combinations of the EPZ-6438, GSK126, UNC0642, UNC0638 and BIX-01294 compounds described in table 1.1.

The reported mechanisms of anti-tumour action are diverse. Casciello et al. showed that in breast, melanoma and ovarian models, EZH2i and G9ai treatment effects are driven by a photodynamic stress (PDT) response, whereby PDT genes that are usually repressed by both enzymes are upregulated [107]. In particular, the cytokine IL24 was identified as a target of dual repression, and its upregulation was both essential and sufficient for cell death following treatment [107]. In contrast, de-repression of differentiation-associated genes was proposed to be central to the synergistic anti-tumour efficacy of dual EZH2i/G9ai treatment in prostate cancer models [113].

Another treatment effect of the EZH2i/G9ai combination with a growing body of support is that of immune stimulation, similar to DNMTi combination effects. In a multiple myeloma model, use of dual EZH2i/G9ai treatment activated IFN immune signalling [110], a finding that has been replicated by our group in ovarian high-grade serous carcinoma cell lines and murine models [155]. The mechanism of action underlying immune stimulation has been proposed to centre on the repetitive transposable elements (TEs) littered throughout the genome, and their potential to stimulate a response known as viral mimicry.

1.7 Transposable elements (TEs)

1.7.1 Introduction to transposable elements: nomenclature, abundance and evolutionary ages

Transposable elements are mobile sequences of DNA that make up approximately 45% of the human genome [156] and are found in all eukaryotic genomes [157]. TEs encompass elements with a wide range of different structures, abundances, evolutionary ages and sizes [157, 158]. Despite this complexity, they can be broadly divided into two major classes based on their original mechanism of transposition: class I TEs (retrotransposons) and class II TEs (DNA transposons) (figure 1.10a, figure 1.10b) [159, 160]. Class I TEs require an RNA intermediate for replication, which is reverse-transcribed into DNA followed by re-insertion into the genome. This "copy-and-paste" mechanism contrasts with the "cut-and-paste" mechanism of class II elements, which involves excision of the element prior to insertion elsewhere in the genome. Classes can be subdivided into subclasses, superfamilies and families/sub-families [161], although there remain conflicting strategies for classification of these abundant elements [161].

It is important to note that the endogenous retrovirus (ERV) superfamily can also be subdivided into classes (or families), creating a confusing nomenclature clash. ERVs may be class I, II or III (also

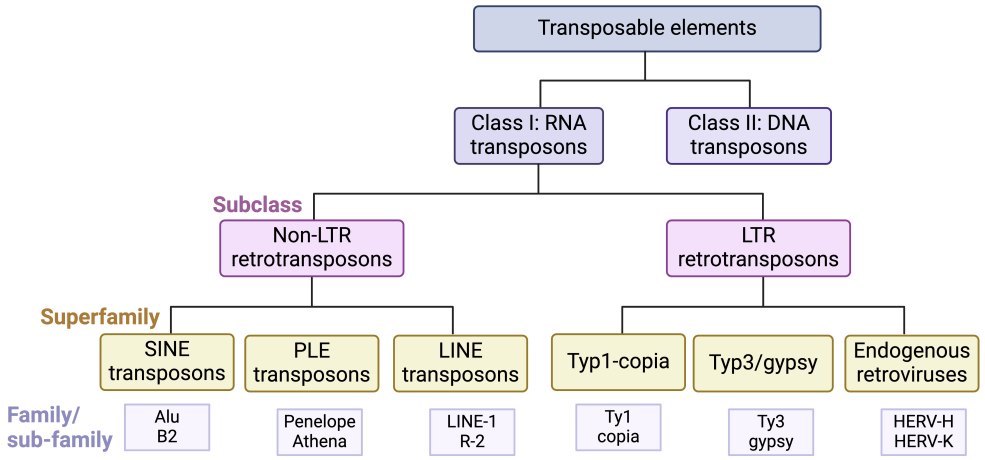
known as ERV1, ERVK and ERVL families respectively) (figure 1.10c) dependent on their closest resemblance to exogenous retrovirus categories [162]: class I ERVs are most similar to gamma- and epsilonretroviruses, class II to alpha-, beta- and deltaretroviruses and class III to the spumaviruses [162]. Whilst both mouse and human ERVs are classified in this way, the similarity between equivalent mouse and human classes is minimal [162] as is their abundance [163]. For example, class I and class III elements are much more prevalent in the human genome than the mouse genome (at 2.92% and 1.55% versus 0.68% and 0.58% respectively) whilst the opposite is true of class II elements (at 0.3% in human versus 3.14% in mouse) [163]. Note that ERVs are most often found in the genome as solo long terminal repeats (LTRs) due to recombination between LTRs and loss of the internal sequences [164].

Most ERVs integrated between 10-100 million years ago (mya) [165]. Defining the ages of ERV classes can be challenging, and numerous approaches exist [166]. For example, ERV ages can be measured via comparison of the sequence divergence between the 3' and 5' LTRs of an ERV; these should be identical at the time of integration, but accumulation of mutations over time results in sequence divergence [166]. The approach cannot measure ERV ages with complete accuracy, however can provide an estimation based on the assumption that $\sim$0.4% LTR divergence occurs per million years since integration [165].

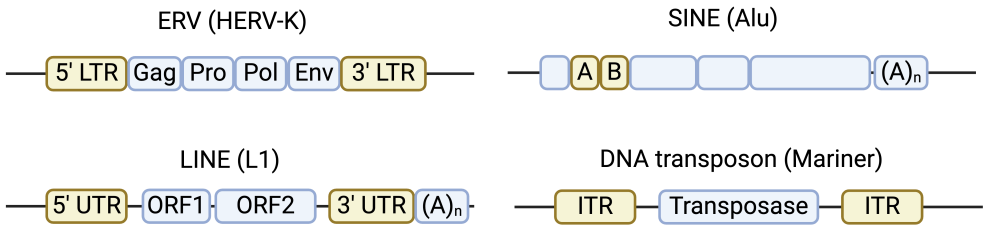
An alternative approach includes sequence comparison between orthologous insertions in closely related species, whereby the last common ancestor can be used to estimate the minimum time since ERV insertion [166]. This kind of analysis has demonstrated that some ERV elements exist in all jawed vertebrates, with several ERVs also found in jawless vertebrates, suggesting that the oldest ERVs are likely to have originated >450 mya [166]. However, an ERV insertion may be lost in a particular outgroup species and therefore its age can be underestimated using this inter-species comparison approach [166].

Many more ERV dating methods exist [166]. One approach that has recently demonstrated concordance with results from the LTR divergence approach is that of measuring CpG density [153, 165]. Given that CpG density decreases over time, those LTRs with the lowest CpG content are likely to be evolutionarily oldest [153]. Comparison between class I, II, and III ERVs demonstrated that class II ERVs (ERVK family) have both the lowest LTR divergence and the highest CpG content [153, 165], indicating they are evolutionarily young. Indeed, many ERVK element integrations are likely to have emerged in the last 5 million years [165], although insertion of others is dated between 15-40 mya [165, 167]. In contrast, class III ERVs (ERVL family) have the highest LTR divergence and the lowest CpG density and are therefore considered to be evolutionarily old, with many elements likely integrating more than 80-100 mya [165, 167]. Class I ERVs (ERV1 family) have intermediate levels of both LTR divergence and CpG density, and many of these elements are calculated to have

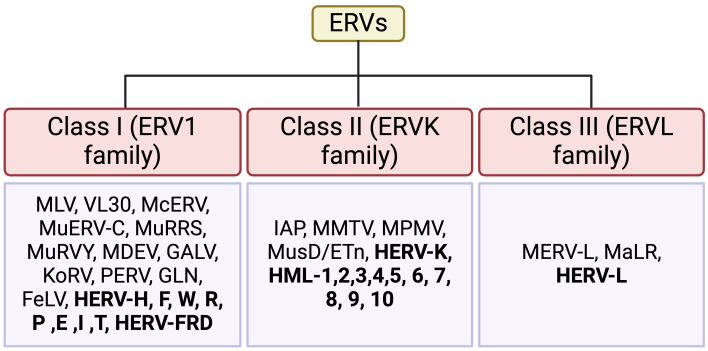
integrated between 30 and 90 mya [165, 167]. The wide distribution of LTR divergence levels in the ERV1 family make definitive distinction between these elements and those from the ERVL family difficult. Accordingly, the ERVK family are referred to as young ERVs throughout this thesis, whilst ERV1/ERVL are referred to as intermediate/old ERV elements.



(a)



(b)



(c)

Figure 1.10 **Classification and structure of transposable elements (TEs).** (a) The classification of TEs into class, subclass, superfamily and (sub-)family. LTR = long terminal repeat, SINE = short-interspersed nuclear element, PLE = Penelope-like elements, LINE = long-interspersed nuclear element. Adapted from [159] and [168]. (b) Example structures of TE families/sub-families from different TE superfamilies. N.B. ERVs are often found as solo LTRs. Adapted from [169] and [170]. (c) ERV classes (red boxes) and examples of ERV families/sub-families within each class (purple boxes). ERV classes are often also referred to as families. Human ERVs are shown in bold. Adapted from [162].

1.7.2 TEs in the genome: effects on genome structure and function

Despite the first suggestions by Barbara McClintock in 1956 of regulatory roles for TEs [171], it was not until the early 2000s that the hypothesis that TEs were "junk" or "selfish" DNA truly began to be overturned [172]. This is in part due to the difficultly in mapping TEs in the genome because of their repetitive nature [159]. Since this time, it has become clear that the evolution of eukaryotic genomes is intricately linked to TEs and the genetic diversity they supply [157].

The size and structure of the human genome are heavily affected by TEs. Mobilisation of TEs is directly mutagenic, via disruption of host genes [173]. Additionally, non-mobile elements influence genome structure due to the repetitive nature of TEs, which increases the recombination rate between sections of DNA [168] and can cause deletions, duplications, or inversions [159]. The balance between transposition and deletion events driven by TEs is a major contributor to genome size [159].

TEs are also involved in controlling higher-order chromosomal structures, via short- and long-range interactions. For example, both telomeres and centromeres in different species are known to depend on TEs or their products [159]. Additionally, TEs provide binding sites for anchoring proteins such as CTCF and TFs, thereby regulating chromatin loops and TAD boundaries [7, 174–176], whilst also contributing to stability of heterochromatin in the B compartment [176]. This is demonstrated by the human endogenous retrovirus sub-family H (HERVH), one of the most abundant ERVs in the human genome [177]. Transcriptionally active HERVH has been shown to be necessary and sufficient for demarcation of TADs in human stem cells [177, 178], with recruitment of CTCF and RNA polymerase II at HERVH sequences [177].

Alongside structural roles, TEs contribute to genome evolution as direct sources of coding and non-coding RNAs [159, 179]. For example, the *gag*, *pol* and *env* genes derived from ERVs have all been exapted to form genes in vertebrates [180]. One of the most well-described exaptation events is that of *env*-derived genes, called *syncytins* [180]. Syncytins have been exapted independently in several mammalian sub-lineages to aid in development of the placenta, via promoting fusion of trophoblast cells [180]. In addition to these protein-coding functions, up to 83% of lncRNAs contain at least one TE, occupying an average of 30% of the lncRNA sequence [181]. These TEs can contribute to the processing, stability, localisation and function of lncRNA as well as mRNA [181]. For example, in parallel to roles in stem cell TAD formation, HERVH transcripts may provide an enhancer-associated lncRNA that recruits both the Mediator complex – essential for transcription initiation – and pluripotency-associated TFs [177].

Each of the TE effects described so far is also associated with the crucial role of TEs as cis-regulatory elements. Roughly 25% of cis-regulatory elements have been suggested to derive from TEs [182], which are able to act as promoters, enhancers, insulators, silencers and transcription-factor binding sites [169, 176, 183, 184], alongside affecting transcription via regulation of TAD

boundaries and lncRNA effects (figure 1.11) [169]. Different classes and ages of TEs are more likely to act as different types of cis-regulatory element: older TEs show greater likelihood of performing cis-regulatory roles as both enhancers and promoters, with a particular over-representation of ERVs as enhancer elements and especially the class I/ERV1 family elements [185]. For example, the MER41 ERV1 family have been shown to act as enhancers in response to interferon gamma (IFNG) stimulation, promoting expression of IFNG target genes following binding of the Signal Transducer And Activator Of Transcription 1 (STAT1) transcription factor [184].

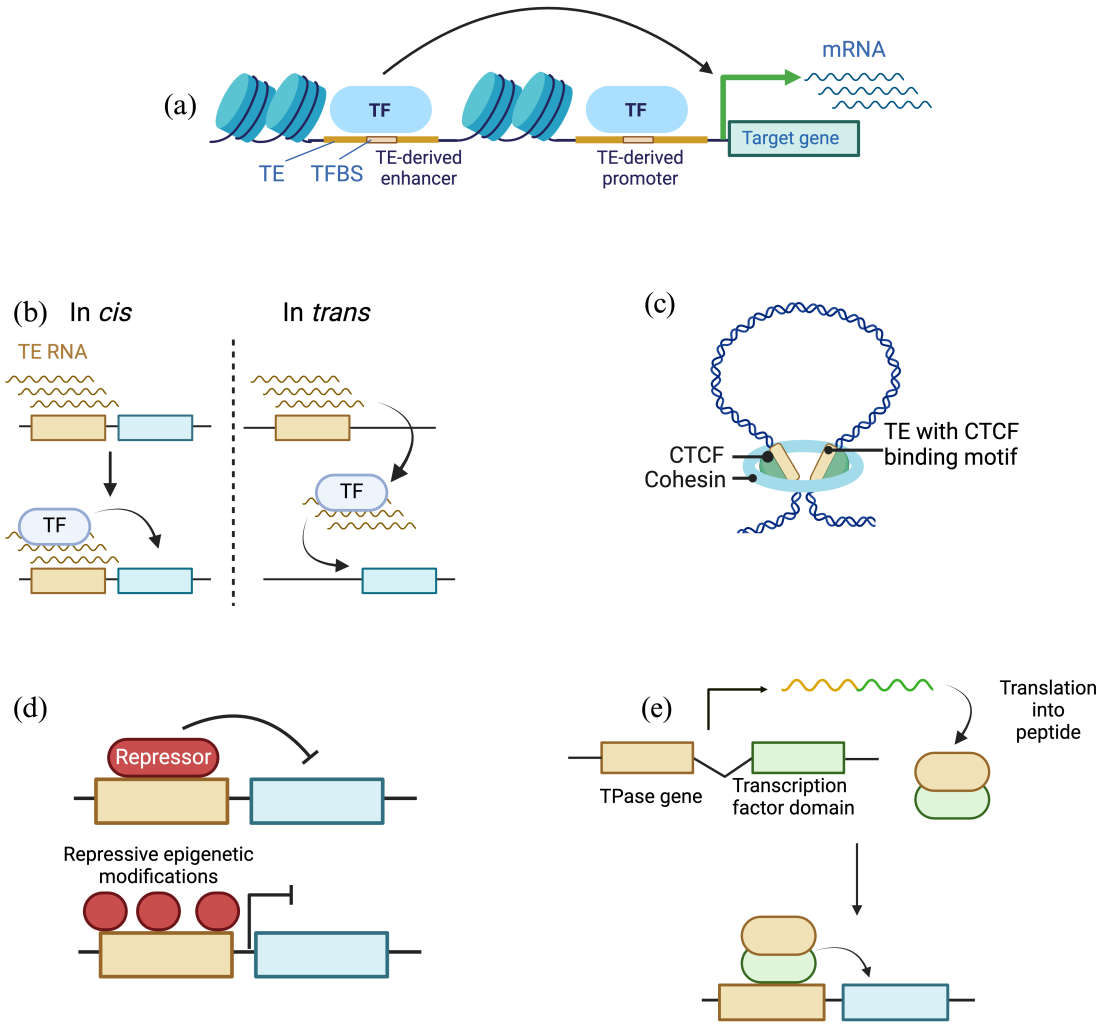


Figure 1.11 **Mechanisms by which transposable elements (TEs) regulate transcription.** (a) TEs can act as promoters or enhancers for host genes, with transcription factor binding sites (TFBSs) allowing binding of activating TFs. (b) TEs can be transcribed to produce lncRNAs, which can then bind TFs to recruit them to genes in *cis* or in *trans*. (c) TEs can regulate the formation of topologically associating domains (TADs), such as by providing a binding site for CTCF proteins. (d) TE silencing via repressive TFs or epigenetic modifications may silence nearby gene expression. (e) TEs provide transposase enzymes (TPase) with DNA-binding domains, that may fuse to host TF domains. Translation of this fusion gene results in a novel TF with a DNA-binding domain, able to bind DNA at cognate TE sites. Adapted from [169] with additions from [186]. Part (c) was adapted from the "Transcriptional Regulation by CTCF and Cohesin" template from BioRender.com (2024), retrieved from `https://app.biorender.com/biorender-templates`.

1.7.3 A crucial role for epigenetics in TE silencing

The large majority of TEs are not transcriptionally active, due to epigenetic silencing and the accumulation of mutations [187]. Two key hypotheses for the epigenetic modification of TEs exist: the genome defence hypothesis and the exaptation hypothesis [188]. The former proposes that epigenetic modifications repress TEs and protect against their disruptive genomic effects. In contrast, the exaptation hypothesis proposes that the co-option of TEs to become regulatory elements has resulted in recruitment of epigenetic modifications to support this role [188]. Evidence exists for both hypotheses and demonstrates that the function of TEs varies depending on TE age, type and cellular context [188].

Multiple different epigenetic factors contribute to TE silencing, including DNA methylation, histone modification, small RNAs, Krüppel-associated box (KRAB) domain-containing zinc finger proteins, splicing inhibition and RNA modifications [189]. DNMTs evolved at a very similar time to TEs and are thought to be the principal mechanism for suppression of TEs in the genome [190, 191]. However, DNA methylation occurs at CpG dinucleotides, the number of which decrease over time due to spontaneous cytosine deamination [153]. Additionally, DNA methylation marks are erased and reconfigured during development [87], highlighting the necessity for alternative epigenetic silencing mechanisms over TEs.

Histone methylation marks act in parallel to DNA methylation to silence transposable elements. Particularly, H3K9me2/3 is proposed to have evolved to silence transposons [27] and is strongly enriched for repetitive elements including TEs [27, 189]. H3K9 HKMTs such as G9a/GLP and SETDB1 are essential for ERV repression during early development [152]. Some reports suggest a similar role for EZH2; ERVs, long-interspersed nuclear elements (LINEs) and SINEs are all enriched in H3K27me3 marks [75, 192, 193] and it has been hypothesised that H3K27me and H3K9me2/3 may share a common ancestral role in silencing TEs [194]. Although evidence for a direct role for H3K27me3 in TE/ERV silencing is more limited than for H3K9me [153], compelling evidence exists for a combinatorial or compensatory role with DNA methylation or H3K9 methylation [153, 195–197]. Indeed, several groups have described an "epigenetic switch": loss of DNA methylation at TEs can be compensated for by repressive histone methylation, to ensure silencing of TEs with fewer CpG sites available for methylation, or of demethylated TEs such as during development or in disease [153, 195–197].

Importantly, the primary mechanism for epigenetic silencing of TEs varies depending on the age and type of TE, as well as the host species. Within the ERV superfamily, the different silencing mechanisms dependent on element age are particularly evident. Older/intermediate age ERVs are thought to rely more heavily on histone methylation rather than DNA methylation silencing mechanisms, due to the loss of CpG sites over time following cytosine deamination [153]. These effects

are still under investigation, and even for ERVs silenced by H3K9 methylation there is conflicting evidence over which classes of ERVs are silenced by which H3K9 HKMTs. For example, Montavon et al. showed that *G9a/GLP* knockout in mouse ESCs had no effect on class III ERVL elements whilst upregulating class II ERVK elements [198], whilst Maksakova et al. observed the inverse relationship [199].

1.7.4 A brief overview of TEs in cancer

TEs that are silenced in healthy tissues are often reactivated in cancer, the most frequently over-expressed TEs being ERVs [200]. For example, certain HERV-K and HERV-E elements are usually repressed in healthy tissue, but are expressed in various cancer types [201]. The aberrant epigenome that is characteristic of cancer [117] can contribute to the dysregulation of TE expression, via disruption of silencing marks and heterochromatin [200, 202]. However, the dysregulation of TEs is not simply a passenger effect: TE activity can directly contribute to carcinogenesis [203, 204].

Due to their physiological roles in structural and transcriptional control, TEs can promote tumourigenesis in a myriad of ways. TEs may generate chromosomal rearrangements, drive mutagenesis via new insertions, or drive dysregulated gene expression via their regulatory capacity and effects on chromatin conformation [203–205]. TEs have been described to contribute to several hallmarks of cancer in this way, such as genomic instability [206], immune evasion [207] and metastasis [208]. A clear example of a carcinogenic TE effect includes the insertion of a TE into DNA repair genes BRCA1 DNA repair associated (*BRCA1*) and BRCA2 DNA repair associated (*BRCA2*), driving mutagenesis of these genes that are so strongly implicated in breast and ovarian cancer [209]. In their regulatory capacity, TEs may drive overexpression of an oncogene as promoter [210] or enhancer elements [211], or disrupt entire transcriptional programs via providing a binding site for a transcription factor [212].

1.7.5 Exploitation of TEs in cancer via viral mimicry

Despite the many possible pro-tumourigenic effects of TEs, a growing body of literature supports a potential role for TEs in anti-tumour responses. Several groups have shown correlations between ERVs and immune activation [200, 213–215], and even between high ERV expression and improved prognosis [214, 216–219]. This has led to interest in targeting the de-repression of TEs – and particularly ERVs – as a therapeutic approach in cancer, often using epigenetic agents [202, 220, 221].

The immune response elicited by ERV or TE de-repression is known as viral mimicry [208]. Transcription of TEs, potentially accompanied by reverse-transcription, leads to accumulation of nucleic acids in various possible conformations in the cytoplasm. Similarly to infection with an

exogenous retrovirus, such nucleic acids can be recognised by intracellular nucleic acid receptors, resulting in the stimulation of a complex network of intracellular immune signalling pathways which converge on inflammatory transcription factors leading to cytokine and interferon responses (figure 1.12) [222]. Alternatively, TEs may be translated into peptide antigens and presented at the cell surface, promoting an immune response via immune cell recognition [223, 224]. Alongside stimulating an immune response, activation of these signalling pathways can induce cell-autonomous apoptotic and necroptotic cell death pathways [225–228].

Viral mimicry following treatment with epigenetic agents was initially identified using DNMTi, with induction of double stranded RNA (dsRNA) from ERVs [213, 229]. Subsequently, many other groups have replicated the induction of viral mimicry using DNMTi [214, 226, 230–234]. Viral mimicry has now been demonstrated in many tumour models, via both ERV and non-ERV TEs [234–237], via multiple different pathways including metabolic signalling or DNA damage signalling [226, 238], and using alterative epigenetic therapies including histone deacetylase inhibition [223, 231], lysine demethylase inhibition [239, 240] and histone methyltransferase inhibition [120, 152, 195, 225, 226, 241–246] as well as non-epigenetic approaches [232, 235–237, 247, 248].

1.7.6 Targeting EZH2 and/or G9a as a viral mimicry approach

Given the role for EZH2 and G9a in silencing TEs, both enzymes have been investigated in the context of viral mimicry. For example, in both small cell lung cancer and prostate cancer models, EZH2i has been shown to induce dsRNA and an IFN response [241, 243]. EZH2 activity has also been demonstrated to restrict viral mimicry effects in response to other treatments. For example, DNA hypomethylation at TEs occurs in the development of resistance to taxane treatment in triple-negative breast cancer, resulting in an epigenetic switch to rely on H3K27me3-mediated TE silencing [195]. Similarly, EZH2 appears to restrict the viral mimicry response to DNA hypomethylation in squamous cell carcinoma [246], explaining the heightened viral mimicry effect of EZH2 targeting in combination with DNMTi [153]. These findings suggest EZH2i may be a therapeutic vulnerability in multiple settings via viral mimicry induction, such as in the context of DNA hypomethylation following the use of other therapeutics.

G9a is another promising target for viral mimicry. Several groups have shown induction of viral defence or IFN genes following individual targeting of G9a [110, 153, 226]. However, this effect has been demonstrated to be stronger upon dual targeting of G9a and either DNMT [120, 152, 153] or EZH2 [110, 155], again demonstrating that the redundant and compensatory relationship between epigenetic marks recommends a dual inhibition approach. It is important to note that both G9a and EZH2 have been implicated in repression of immune genes alongside TEs [128, 142, 249, 250], for example H3K9me2 is a key suppressor of IFN and interferon-stimulated gene (ISG) expression, with

G9a ablation upregulating many of these genes [249]. Hence, the viral mimicry effects of targeting EZH2 and/or G9a may be amplified by direct de-repression of immune targets.

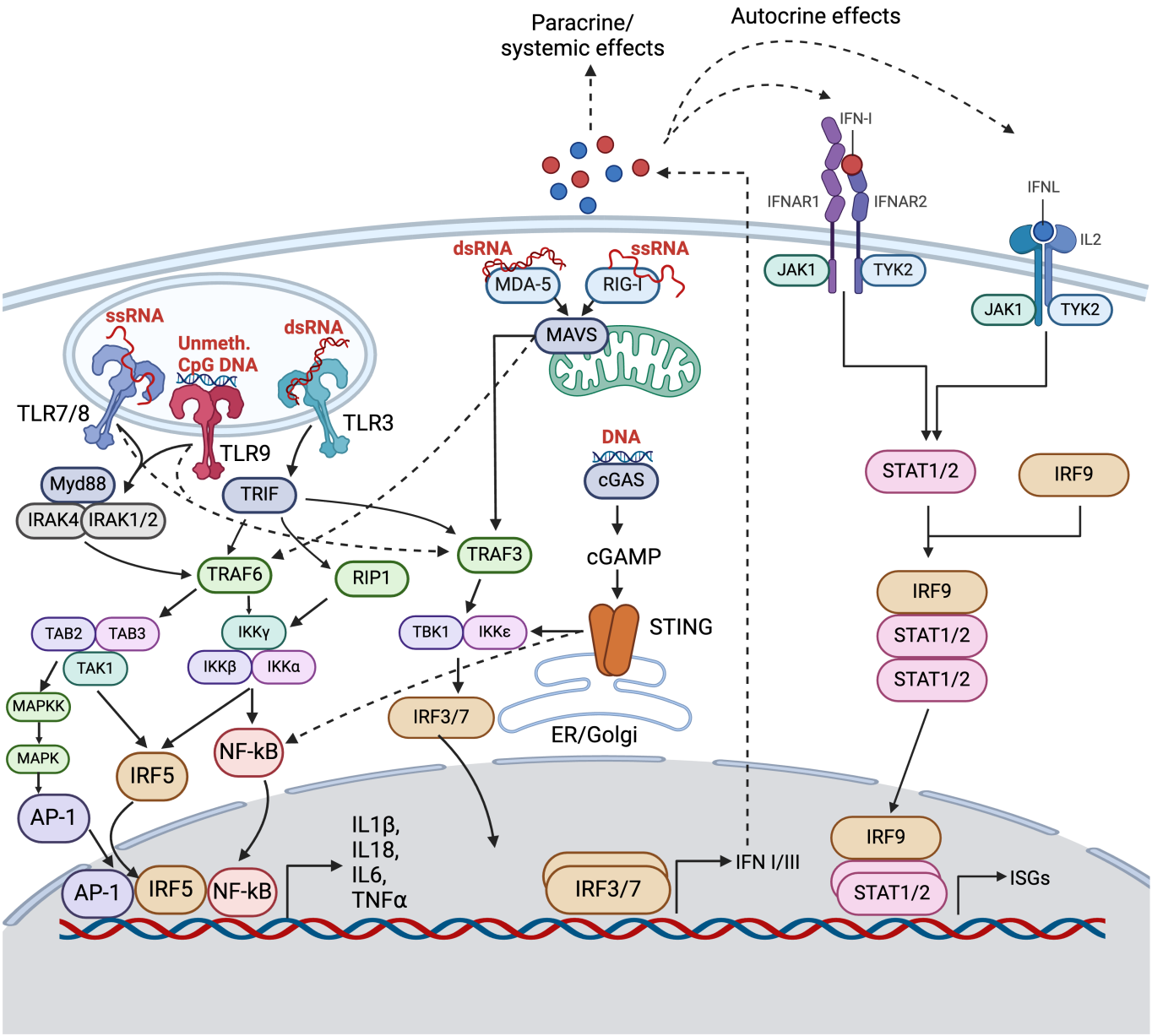


Figure 1.12 **Overview of intracellular innate immune signalling pathways downstream of nucleic acid (NA) recognition.** NAs may be recognised as single stranded RNA (ssRNA) or dsRNA in the endosome or cytoplasm, as unmethylated CpG DNA in the endosome, or as DNA localised in the cytoplasm. These NAs are bound by pattern recognition receptors (PRRs) such as Toll-like receptors (TLRs), cytosolic receptors including RIG-I-like receptors (RLRs) RIG-I and MDA-5, or by cytoplasmic cGAS. Downstream of recognition and signalling through a variety of potential pathways, there is convergence on transcription factors such as Activator Protein 1 (AP1), IRF5 and NF-KB or IRF3 and IRF7, resulting in the induction of inflammatory cytokines (e.g. IL1B, IL18, IL6, TNFA, IFNI and IFNIII). IFNs can be recognised at the cell surface by receptors resulting in further signalling through STAT1, STAT2 and IRF9, to activate transcription of interferon-stimulated genes (ISGs). Caption continued overleaf.

Figure 1.12 TRIF = TIR domain containing adaptor molecule 1, Myd88 = Myeloid Differentiation Primary Response Protein MyD88, IRAK = Interleukin 1 Receptor Associated Kinase, TRAF = TNF Receptor Associated Factor, RIP1 = Receptor-Interacting Protein 1, TAB = TGF-Beta Activated Kinase, TAK1 = MAPK37 = Mitogen-Activated Protein Kinase Kinase Kinase 7, MAPK = Mitogen-Activated Protein Kinase, MAPKK = Mitogen Activated Protein Kinase Kinase, AP1 = Activator Protein 1, IKK = Inhibitor Of Nuclear Factor Kappa-B Kinase Subunit, NF-KB = Nuclear Factor Kappa B subunit 1, TBK1 = TANK-Binding Kinase 1, MDA-5 = Melanoma Differentiation-Associated Protein 5, RIG-I = RIG-1 = DDX58 = Retinoic Acid Inducible Gene I/ DEAD (Asp-Glu-Ala-Asp) Box Polypeptide 58, MAVS = Mitochondrial Antiviral Signalling protein, cGAS = Cyclic GMP-AMP Synthase, STING = TMEM173 = MITA = Stimulator Of Interferon Response CGAMP Interactor, ER = endoplasmic reticulum, IFNAR1/2 = Interferon Alpha and Beta Receptor subunit 1/2 , JAK1 = Janus Kinase 1, TYK2 = Tyrosine Kinase 2, IL = Interleukin, TNFα = Tumour Necrosis Factor alpha, IRF = Interferon Regulatory Factor, STAT = Signal Transducer And Activator Of Transcription 1. Figure adapted from [251] with addition from [222].

1.7.7 Which features prime cells for viral mimicry?

The deregulation of TEs observed in cancer may contribute to the tumour-specificity of epigenetic treatments; aberrant genetic and epigenetic features of cancer cells may lead to increased basal expression of TEs, pushing cells closer to the threshold of tolerance and rendering them more prone to viral mimicry-induced death [220, 252]. For example, p53 deficiency has a priming effect for viral mimicry in response to DNMTi, most likely due to increased basal TE transcription [252]. Yet, the relationship between basal expression of ERVs and sensitivity to viral mimicry induction/immune stimulation is not yet clear; high basal ERV expression has been proposed both to increase [220] and to decrease [214] sensitivity to viral mimicry. The relationship may be confounded by the potential for cancer cells to become tolerant to, or evade, IFN induction in the context of chronically elevated dsRNA [220]. This literature suggests it may be valuable to consider more specific downstream priming events.

Additional factors that are demonstrated to sensitise cancer cells to viral mimicry include mutations in epigenetic modifiers, mutations in histone proteins and resistance to chemotherapy [220]. For example, mutations in the genes encoding the H3K36 or H3K27 lysine methyltransferases SETD2 or PRC2, or the DNMT DNMT3A, have all been shown to increase sensitivity to DNMTi-induced viral mimicry [253–256]. Interestingly, vitamin C, which is a cofactor for the DNA demethylation TET enzymes, is synergistic with DNMTi in induction of a viral mimicry response [228]. More recently, deregulation of the oncogene cMYC has been suggested to increase sensitivity to viral mimicry by increasing the rate of splicing errors and therefore pushing cells closer to a threshold for immune signalling [235].

1.8 cMYC

1.8.1 cMYC: structure, interactions and transcriptional activation effects

Alterations in cMYC – at the DNA, protein and RNA level – are found in over 70% of cancers [257]. cMYC is a basic helix-loop-helix-leucine zipper (bHLHZip) transcription factor (figures 1.13a and 1.13b), known as a "master regulator" due to its roles in regulating diverse cellular processes [258]. As many as 4,000 genes have been identified as cMYC-regulated [259]. This master regulator role allows aberrant cMYC activity to contribute to all hallmarks of cancer such as evasion of growth suppression, sustained proliferation, deregulation of metabolism and evasion of cell death [260, 261].

cMYC activity depends on dimerization with the MYC Associated Factor X (MAX) protein, via the coiling together of C-terminal bHLHZip domains on each protein (figure 1.13b) [257]. The cMYC-MAX complex can bind elements known as Enhancer-box (E-box) (CACGTG) or non-canonical E-box (CANNTG) [262] sequences to transactivate targets [257], including central regulators of proliferation such as cyclin-dependent kinase 2 (*CDK2*) and cyclin D2 (*CCND2*) [263]. However, distinct subsets of the vast number of potential E-box and non-canonical E-box sites are bound by cMYC in each individual cell type [258], suggesting the regulation of cMYC binding to DNA is not solely sequence-specific. The chromatin environment is now recognised as an additional important regulator of cMYC binding, with increased cMYC binding at open chromatin regions, as well as potential tethering of cMYC to target genes via chromatin-binding proteins [258]. The effect of cMYC on gene expression appears also to depend on its expression levels [264], contributing to the cell- and context-specific activation of target genes by cMYC.

Transactivation of target genes is dependent on the highly conserved N-terminal region of the cMYC protein – containing Myc Box I (MBI) and Myc Box II (MBII) – comprising the transactivation domain [257]. This domain allows interaction with a vast number of chromatin modifiers to facilitate cMYC transcriptional activation (figure 1.13c) [258]. For example, cMYC interacts with Transactivation/Transformation-Associated Protein (TRRAP) to alter chromatin accessibility via histone acetyltransferase activity and thereby increase transcription of target genes [258]. Alongside activation of specific genes, cMYC can act more broadly as a transcriptional amplifier: in tumour cells, cMYC accumulates at the promoters of genes that are already active and drives additional increases in transcription [258].

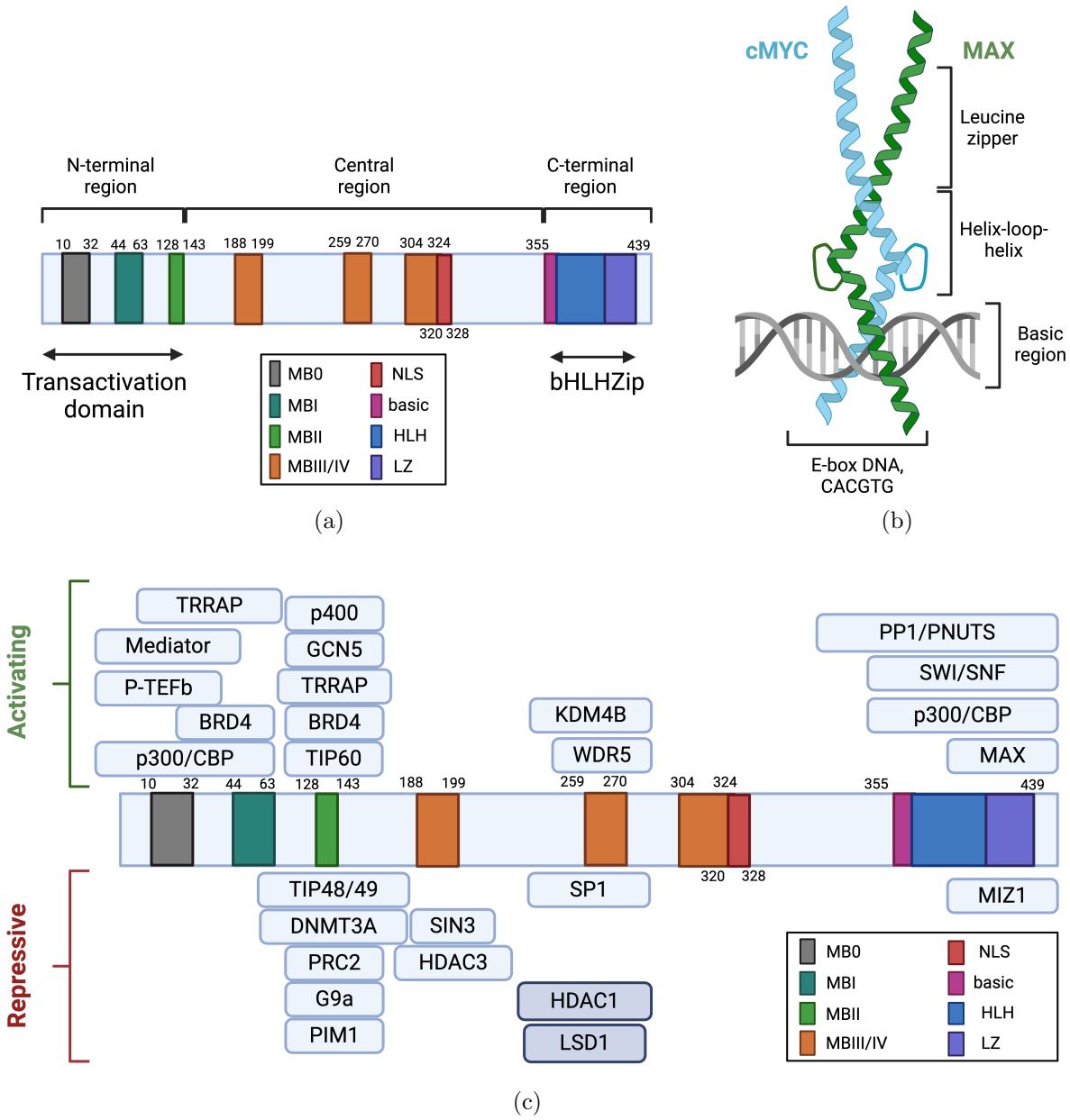


Figure 1.13 **cMYC structure and interactions.** (a) Domain structure of cMYC protein. MB = MYC Box, NLS = nuclear localisation signal, HLH = helix-loop-helix, LZ = leucine zipper. Adapted from [257]. (b) Crystal structure of cMYC (light blue) as part of a heterodimer with MAX (green), binding to E-box regions of DNA. Reproduced from [265]. (c) Interactions between cMYC domains and chromatin modifying factors. Activating modifying factors are pictured above, repressive modifying factors below. The HDAC1 and LSD1 interactions (darker blue) occur at unknown cMYC domains. Modified based on figure from [258]. TRRAP = Transactivation/Transformation-Associated Protein, P-TEFb = Positive Transcription Elongation Factor b, BRD4 = Bromodomain-Containing Protein 4, p300/CBP = E1A Binding Protein p300/CREB Binding Protein, p400 = E1A-Binding Protein p400, GCN5 = General Control of Amino Acid Synthesis Protein 5-Like 2, TIP60 = 60kDa Tat Interacting Protein, TIP48/49 = 48/49 kDa TBP-Interacting Protein, DNMT3A = DNA Methyltransferase 3a, PRC2 = Polycomb Repressive Complex 2, PIM1 = Pim-1 Oncogene/Pro-viral Integration Site 1, SIN3 = Paired Amphipathic Helix Protein, HDAC3 = Histone Deacetylase 3, WDR5 = WD Repeat Domain 5, KDM4B = Lysine (K)-Specific Demethylase 4, SP1 = Specificity Protein-1, PP1/PNUTS = Protein Phosphatase 1/protein phosphatase 1 regulatory subunit 10, SWI/SNF = SWItch/Sucrose Non-Fermentable, MAX = MYC Associated Factor X, MIZ-1 = Myc-Interacting Zinc Finger Protein 1, HDAC1 = Histone Deacetylase 1, LSD1 = Lysine-Specific Histone Demethylase 1A.

1.8.2 Transcriptional repression and immune evasion mediated by cMYC

A lesser-known function of cMYC is as a transcriptional repressor [266]. cMYC can interact with several co-repressor proteins and complexes to mediate this effect (figure 1.13c). Targets of cell-intrinsic repression by cMYC are numerous, including metabolic regulators [267, 268], negative regulators of cell cycle progression [268] and adhesion genes [266]. Additionally, there is a growing body of evidence for cell-intrinsic repression of immune signalling by cMYC [269–278].

Repression of immune signalling by cMYC has been demonstrated in multiple tumour types. Direct targets of cMYC include Janus Kinase 2 *(JAK2)*, *STAT1*, Interferon Regulatory Factors 5 and 7 (*IRF5/7*), Stimulator Of Interferon Response CGAMP Interactor (*STING*) and many interferon-stimulated genes (ISGs) [269, 272–275, 277, 278]. These effects contribute to defects in signalling of type I and II IFN signalling [269, 274, 275]. In addition to – and potentially downstream of – inhibition of these crucial innate immune signalling components, cMYC has been shown to reduce expression of antigen presenting machinery [270, 275, 276] and a range of inflammatory mediators such as Interleukin 1 beta (IL1B), CC Motif Chemokine Ligand 5 (CCL5) and CXCL10 [270, 272, 276, 278].

Importantly, cell-intrinsic immune repression is accompanied by global modulation of the immune response by cMYC. For example, cMYC promotes an immunosuppressive microenvironment with activation of pro-inflammatory cytokines including CC Motif Chemokine Ligands 2/9 (CCL2/9) and Interleukin 23 (IL23) [279], increased angiogenesis [279] and upregulation of cell surface checkpoints including Programmed Death-Ligand 1 (PDL1) [280]. These multimodal effects demonstrate the extensive cMYC-mediated immune effects in tumorigenesis, via both activating and repressive mechanisms.

cMYC can mediate transcriptional repression via interaction with several zinc finger TFs, including Myc-Interacting Zinc Finger Protein 1 (MIZ1), Specificity Protein-1 (SP1), Nuclear Factor Y (NF-Y), Ying Yang 1 (YY1) and Transcription Factor II I (TFII-I) [258]. Repressive cMYC activity has been best described via interaction with MIZ1, encoded by *ZBTB17*, which binds to cMYC via the helix-loop-helix region. Intriguingly, this interaction is proposed to be specific to tumourigenic over normal cells [281]. Independently from cMYC, MIZ1 is a transcriptional activator in association with p300. Hence, cMYC binding and sequestering of MIZ1 can result in repression via preventing the activation of MIZ1 gene targets, such as key regulators of cell cycle arrest, Cyclin Dependent Kinase Inhibitor 1A (*CDKN1A/p21CIP*) and Cyclin Dependent Kinase Inhibitor 2B (*CDKN2B/p15INK4B)* [258].

Additionally, MIZ1 may interact with cMYC-MAX complexes directly to modulate gene expression. In this case, the ratio of cMYC:MIZ1 has been proposed to determine the extent of repression versus activation of a target gene, with a ratio near 1:1 leading to transcriptional repression whilst

higher ratios lead to more gene activation by either cMYC or MIZ1 respectively [282]. Repression can be achieved via cMYC-MAX-MIZ1 complexes recruiting DNMTs and additional epigenetic modifiers to repress gene expression via chromatin modification [258]. However, over 60% of cMYC transcriptional repression is not explained by MIZ1-dependent mechanisms [282]. Recent literature has implicated several epigenetic modifiers as additional contributors to cMYC-dependent repression, independent of MIZ1 activity.

1.8.3 Interactions between cMYC and G9a

G9a and cMYC are interconnected on various levels and alterations in the two proteins significantly co-occur in cancer [268] (appendix figure A.2). G9a has been shown to promote cMYC expression via *MYC* promoter binding [137] and protein stabilisation [283], and consequently inhibition of G9a can downregulate cMYC at the mRNA and protein level [137, 139, 267, 283]. Additionally, cMYC has been shown to occupy the *G9a* promoter [268, 283] and cMYC targeting can reduce mRNA and protein levels of G9a and reduce global H3K9me2 [268]. Thus, G9a and cMYC sit in a positive feedback loop, with bidirectional regulation.

Aside from this relationship between cMYC and G9a expression and stability, cMYC and G9a have also been shown to physically interact to repress target genes [284–286], via the MBII region [268]. Targets of cMYC-G9a repression are various, including genes involved in lipid metabolism, cell cycle regulation and signal transduction [268]. In the context of chronic inflammation, the cMYC-G9a complex has been shown to repress pro-inflammatory immune and apoptosis genes in a cell-intrinsic manner [285]. Importantly, the cMYC-G9a complex binds to cMYC-repressed genes but not cMYC-activated genes [268, 283] and inhibiting G9a can therefore promote de-repression of targets of cMYC repression and provide anti-tumour efficacy [268, 283], whilst cMYC activation targets are unaffected or even downregulated [137, 139, 268]. Of note, most cMYC-G9a repression is independent of MIZ1, but a subset of cMYC-G9a targets is also MIZ1-bound suggesting that G9a targeting might also increase expression of cMYC-MIZ1 repression targets [268].

1.8.4 Interactions between cMYC and EZH2

In a similar manner to G9a, EZH2 and cMYC exist in a positive feedback loop and alterations in each are found to co-occur significantly in cancer (appendix figure A.2). EZH2 has been shown to regulate cMYC at the mRNA and protein level [287–289], potentially via direct *MYC* promoter binding [289–291], or indirectly via the downregulation of microRNAs [287] or regulation of cMYC-stabilising proteins [288, 292]. Consequently, inhibition of EZH2 can reduce cMYC signalling in multiple models [287, 289, 293, 294]. The inverse relationship also exists, whereby cMYC can regulate

EZH2 and H3K27me3 levels: cMYC can bind to the EZH2 promoter [295, 296], suppress negative regulators of EZH2 including microRNAs [250, 297] and stabilise EZH2 at the post-transcriptional level [298]. Several intricate feedforward loops exist between EZH2 and cMYC that amplify these effects. For example, PRC2 inhibits the microRNA *miR-101* in a cMYC-dependent manner, yet *miR-101* inhibits *EZH2* expression [299].

The relationship between cMYC and EZH2 has also been demonstrated to include physical interaction and co-regulation of genes [292, 299–301]. As with cMYC-G9a, the cMYC-EZH2 interaction occurs at least in part via the MBII region of cMYC [292] and can result in recruitment of EZH2 to the promoter of target genes to promote repression [297]. Targets of transcriptional regulation include genes involved in cancer stemness maintenance [289], survival pathways [297] and resistance to chemotherapy [297]. cMYC has also been shown to contribute to non-canonical immune-repressive effects of EZH2; its overexpression leads to EZH2 activation, by blocking several upstream inhibitory pathways [250]. This leads to non-canonical PRC2-independent EZH2 effects that promote reduced IFN signalling and increased cell viability and proliferation [250]. Depleting inhibitors of EZH2, but not those with catalytic inhibitory activity only, were able to restore immune signalling in this context [250].

Some of the cMYC-EZH2 effects include transcriptional activation. EZH2 has been shown to contribute to the activating oncogenic cMYC effects in a non-canonical manner, without H3K27me3 deposition [300, 301]. Here, EZH2 interacts with cMYC and uses a hidden transactivation domain to recruit coactivators such as p300, resulting in co-localisation of activating marks such as H3K27Ac and H3K4me3 [300, 301]. Catalytic inhibition of EZH2 fails to target this non-canonical cMYC-EZH2 axis or to achieve cMYC depletion; instead, efficacy against the cMYC-EZH2 axis has been shown to be much greater using EZH2 protein depletion approaches [292, 300, 301].

Additionally, cMYC-EZH2 effects may depend on interaction with other epigenetic modifiers, particularly BRD4 [293, 294, 296, 302]. Multiple groups have demonstrated synergy between BRD4 and EZH2 inhibition [293, 294, 302]. This may be in part due to BRD4i-mediated disruption of the cMYC-EZH2 relationship. BRD4 has been shown to upregulate *MYC* and consequently *EZH2* transcription, with the use of BRD4i restricting cMYC recruitment to the *EZH2* promoter and therefore reducing EZH2 mRNA and protein levels [296]. Hence, EZH2 cooperates with cMYC in both canonical and non-canonical ways to promote tumourigenesis, directly and indirectly, including both activation and repression of cMYC target genes.

1.8.5 cMYC interactions with other H3K27 and H3K9 epigenetic modifiers

The extensive regulation of genes by cMYC is reliant on interactions with additional H3K27 and H3K9 epigenetic modifiers, aside from G9a and EZH2. For example, H3K9me3 can be placed by SETDB1 as well as G9a; SETDB1 has been shown to increase cMYC protein and signalling, whilst cMYC can bind to the promoter of *SETDB1* and increase its transcription [303]. In addition, the KDM4 H3K9 demethylation family are implicated in cMYC effects (figure 1.13c) [304]. KDM4B directly interacts with cMYC and its target genes and has been found to be indispensable for transactivation of certain metabolic and proliferation targets [304]. Considering acetylation rather than methylation, H3K9 is regulated by a number of complexes including HDAC3 [305]. cMYC can recruit HDAC3 to deacetylate the promoters of target genes in natural killer (NK) immune cells, reducing the cytotoxic capacity of these cells in small cell lung cancer (SCLC) (figure 1.13c) [305]. Hence, cMYC cooperates with numerous H3K9 modifiers to regulate its transcriptional effects.

Several cMYC-H3K27 axes aside from that of cMYC-EZH2 have also been described. For example, cMYC demonstrates a strong preference for open chromatin and may be dependent on H3K27Ac for chromatin binding [306]. Moreover, *MYC* overexpression has been suggested to result in a gain of H3K27Ac peaks at super-enhancer regulatory loci (clusters of enhancers with strong enrichment for transcriptional activators), feeding forward into increased cMYC binding and aberrant transcription in sarcoma cell lines [306]. The H3K27 acetylation complex p300/CBP has also been shown to interact directly with the transactivation domain region of cMYC (figure 1.13c) [307]. This p300/CBP-cMYC relationship is bidirectional, with p300/CBP regulating the stability and turnover of cMYC, whilst cMYC can recruit p300 to gene promoters to promote transcription [307]. This literature demonstrates that regulation of H3K27 acetylation as well as methylation is important for cMYC-mediated effects.

1.8.6 Therapeutic targeting of cMYC: current approaches

There have been numerous attempts to target cMYC pharmacologically [257, 259, 261, 308]. Initially, the ubiquitous and essential nature of cMYC in non-malignant cells suggested targeting may be accompanied by excessive toxicity [308]. Additionally, the nuclear localisation of cMYC and the lack of a binding pocket suitable for small molecule inhibitors led to the labelling of cMYC as "undruggable" [308].

However, a dominant negative approach to cMYC inhibition (Omomyc) challenged this conclusion. In a lung adenocarcinoma model, Omomyc was shown to cause regression of lung tumours [309]. It was well-tolerated with reversal of side-effects upon withdrawal, with most of these side

effects occurring in rapidly regenerating tissues such as the skin and the gut [309]. Many alternative approaches to targeting cMYC and its activity are in progress (figure 1.14, table 1.2) [257]. Yet, despite such promising advances, cMYC targeting therapy has yet to reach the clinic [257].

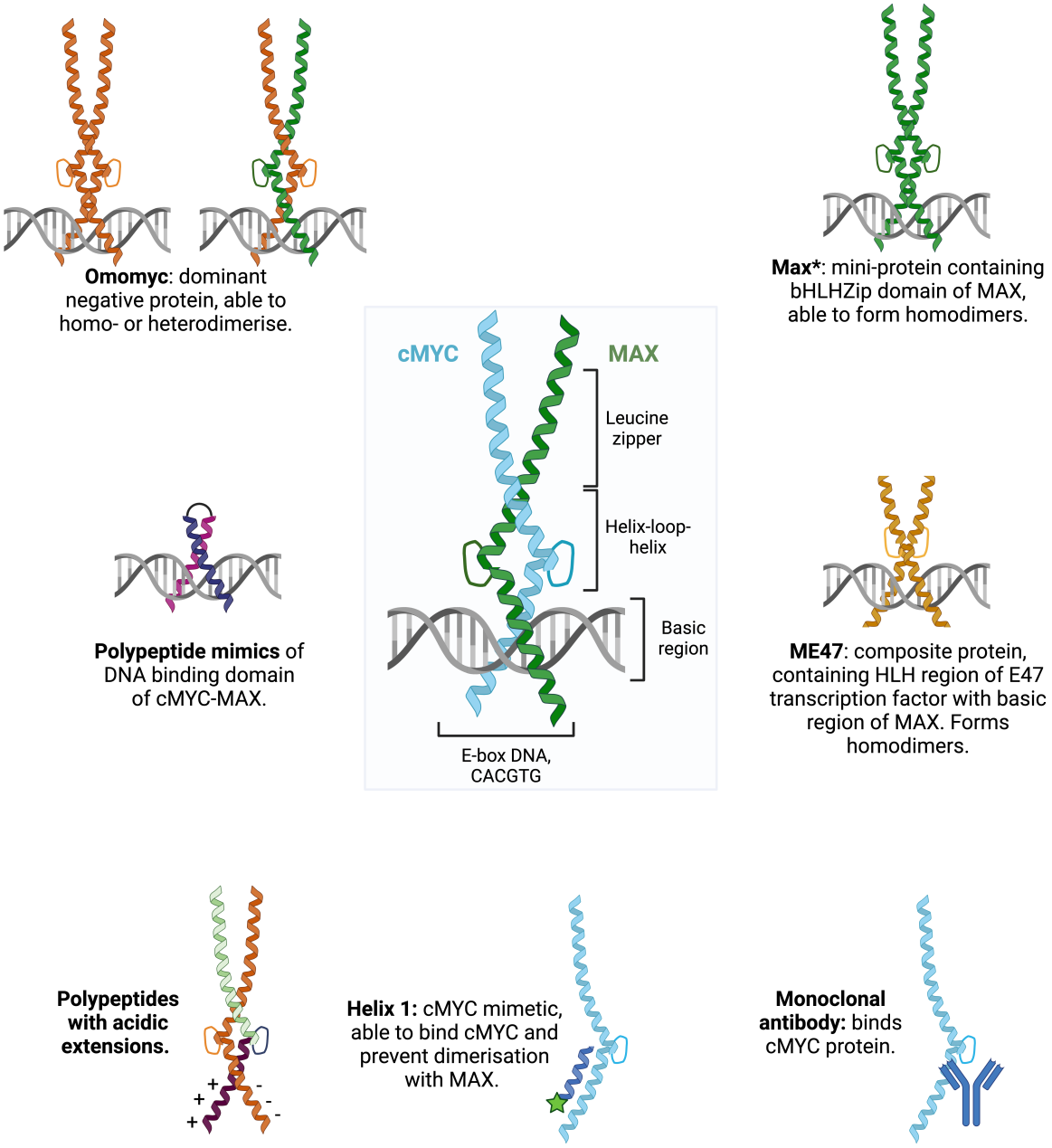


Figure 1.14 **cMYC inhibition approaches.** Visualisation of peptide and protein inhibitors of cMYC (outside), with the cMYC-MAX heterodimer in the centre box. Figure adapted from [257]. Inhibitors described in table 1.2.

Small molecule inhibitor	
Example	**Description**
MYCi975	Binds HLH domain of cMYC and increases phosphorylation of cMYC at threonine 58, leading to increased proteasomal degradation of cMYC [310].
Protein/peptide inhibitor	
Example	**Description**
Omomyc	A dominant negative protein, able to homodimerize or heterodimerize with cMYC and MAX. Various mechanisms are proposed: Omomyc may sequester cMYC/MAX proteins away from the canonical heterodimeric complex, compete for E-box sites on DNA, and promote degradation of cMYC.
Max bHLHZip (Max*)	A mini-protein with the bHLHZip domain of MAX, able to form homodimers and compete for E-box DNA sequences.
Mad	A mini-protein based on an endogenous cMYC-antagonist, MXD1. Able to recruit the epigenetic co-repressor mSin3a, part of a large HDAC complex.
ME47	A composite protein combining the HLH region of the E47 transcription factor with the basic region of MAX, able to homodimerize and compete for E-box sequences.
Monoclonal antibody	An anti-cMYC monoclonal antibody that prevents the binding of cMYC-MAX complexes to DNA.
Helix 1 c-MYC mimetic	A short peptide sequence able to bind to cMYC and prevent its dimerization with MAX. Helix 1 most likely also disrupts formation of cMYC-MAX-MIZ1 complexes.
Polypeptides with acidic extensions	Peptides with the basic region of MAX replaced by an extension of acidic residues. These extend the binding interface of MAX with cMYC, preventing DNA binding by the heterodimer.
Polypeptide mimics of the DBD of cMYC-MAX	Peptides that bind E-box regions to prevent binding of cMYC-MAX to DNA.

Table 1.2 **MYC inhibition approaches.** Table summarising current small molecule and protein/peptide approaches to cMYC targeting. Table reproduced from [257] with addition from [310]. HLH = helix-loop-helix. DBD = DNA-binding domain.

1.8.7 Epigenetic compounds: a novel therapeutic approach for cMYC-deregulated cancer?

Targeting cMYC via disruption of epigenetic interactions may provide an alternative strategy to mitigate cMYC-mediated oncogenicity. As detailed above, both EZH2 and G9a have been implicated in cMYC-mediated oncogenicity and exist in positive feedback loops with *MYC* expression. Targeting either enzyme has proven effective in reducing cMYC activity in multiple models [137, 139, 267, 283, 287–289].

Additionally, cMYC mediates cell-intrinsic repression of immune pathways, in part via collaboration with chromatin modifiers [269, 272–274, 277, 278, 285, 311]. EZH2 or G9a inhibition has been shown to stimulate immune responses via induction of viral mimicry [110, 120, 152, 153, 155, 226, 241, 243], with early evidence suggesting contribution of EZH2 and G9a to cMYC-mediated immune repression [250, 285].

MYC deregulation may decrease the threshold for viral mimicry effects, acting as a "priming" event; cMYC is involved in splicing regulation [235] and acts as a global amplifier of transcription [258], which can contribute to increased accumulation of dsRNAs [312]. Additionally, *MYC* deregulation can increase DNA damage and more generally sensitise cells to apoptosis [279], with basal levels of *MYC* shown to relate to sensitivity to epigenetic therapy [311]. Hence, G9ai/EZH2i-induced viral mimicry may be amplified in *MYC*-deregulated cancer cells, via: 1) the priming effects of *MYC* deregulation and 2) the direct de-repression of cMYC-EZH2/G9a-repressed immune genes. Additionally, the feedback loops that exist between cMYC and EZH2/G9a suggest that EZH2/G9a inhibition may target cMYC repression and activation axes more broadly. Thus, epigenetic targeting of EZH2/G9a activity may help to overcome the difficulties of targeting cMYC activity directly and have potential as a therapeutic approach in cMYC-deregulated tumours (figure 1.15).

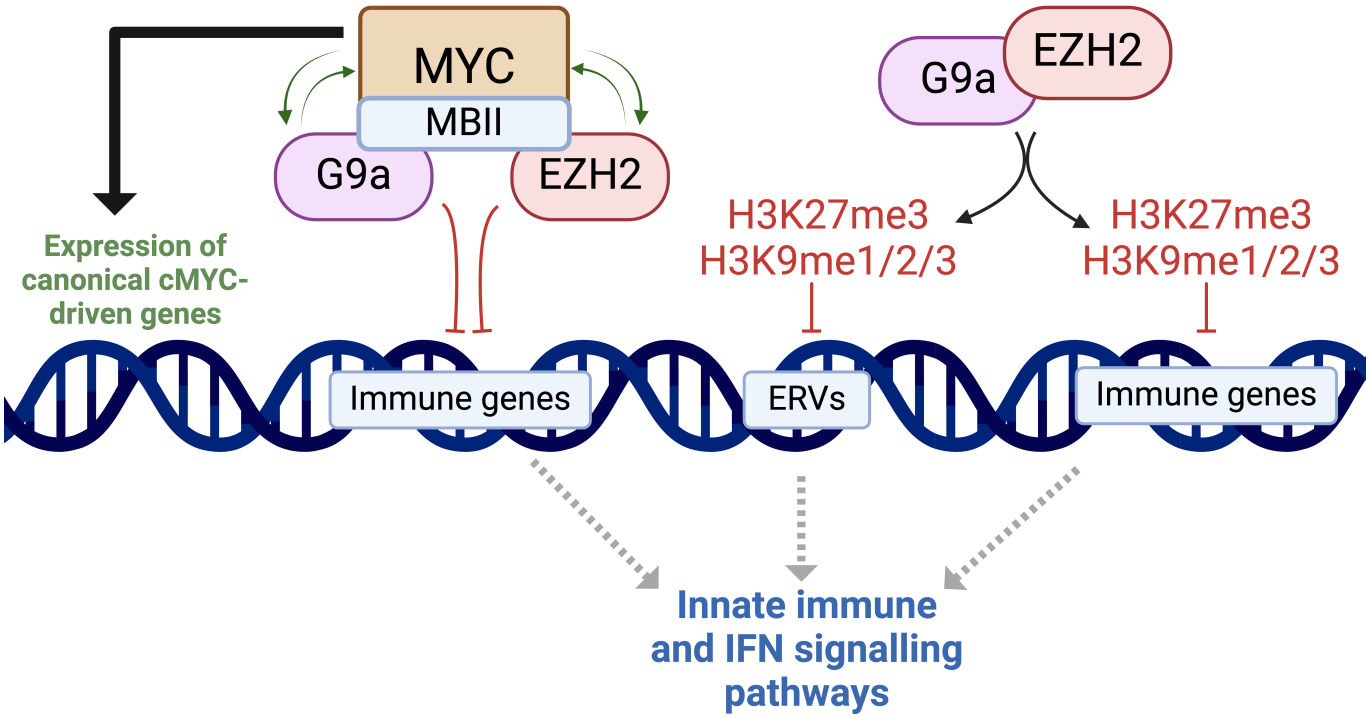


Figure 1.15 **cMYC-G9a/EZH2 immune repression schematic.** Schematic illustrating cMYC oncogenic activity and the interdependency of these effects with EZH2 and G9a. cMYC (brown box, left) exists in a positive feedback loop with EZH2/G9a, both at the mRNA/protein level (green arrows) and via direct interaction using the Myc Box II (MBII) region. cMYC drives transactivation of canonical target genes such as those involved in proliferation and survival. However, it also coordinates transrepression of genes, in part via interaction with epigenetic modifiers such as EZH2 and G9a. The mechanism of action of this repression is not well-defined; it may be mediated by deposition of repressive histone marks, or via non-canonical effects. Targets of cMYC-mediated repression include innate immune and interferon (IFN) signalling genes amongst others. EZH2 and G9a also repress immune signalling independently from cMYC, both directly and indirectly. Indirect effects are mediated via repression of TEs such as ERVs, that can induce an immune response if re-expressed.

1.9 cMYC-deregulated cancer: ovarian high-grade serous carcinoma (HGSC)

1.9.1 Overview

Epithelial ovarian cancer (EOC) is the deadliest gynaecological cancer [313]. It is the 5th leading cause of cancer-related deaths in women in Western nations [314] and in contrast to most solid tumour types has shown little improvement in 5-year survival over the last 40 years [315]. Ovarian high-grade serous carcinoma (HGSC) is the most common EOC subtype, representing roughly 70% of EOC cases [316] and accounting for 70-80% of EOC-associated death [317]. These disheartening statistics are driven in part by the frequent advanced stage diagnosis of HGSC, which occurs due to rapid dissemination and absence of symptoms in early stage disease [315].

HGSC is characterised by almost ubiquitous mutation in *TP53*, the gene encoding the p53 tumour suppressor protein [318]. Accompanying *TP53* mutation are genetic and epigenetic alterations in homologous recombination (HR) DNA repair genes in up to 50% of primary tumours [313, 319]. This includes processes such as methylation or mutation of *BRCA1/BRCA2*, alongside other HR genes such as *RAD51C*, *BRIP1* and *PALB2* [313, 319, 320]. However, aside from these genes, single nucleotide variants and other short variants are rare in comparison to other cancer types: instead, HGSC is a copy number driven cancer with significant chromosomal instability (CIN), due to the genomic instability caused by the prevalent *TP53* and HR alterations [316]. Copy number changes result in frequent amplification of various cancer-associated genes, such as *CCNE1*, *MYC* and *MECOM* [313].

1.9.2 EZH2 and G9a: contributions to HGSC carcinogenesis and chemoresistance

Although not often mutated in EOC, *EZH2* and *G9a* have both been shown to be amplified and overexpressed in EOC compared to normal ovarian tissue (see figures 1.8 and 1.9) [54, 134, 321, 322]. Overexpression of *EZH2* or *G9a* is associated with higher grade disease in EOC [321–324] and higher expression of either enzyme has been shown to predict poorer prognosis [322, 323]. Both enzymes are implicated in metastasis and resistance to treatment [32, 322, 324–327], this latter effect potentially via their roles in maintenance of a stem-like population [328–330]. Importantly, some of EZH2's pro-tumourigenic roles in EOC appear to be via non-canonical activity, independent of its methyltransferase effects [331].

1.9.3 HGSC treatment

First-line treatment for HGSC includes treatment with a cytotoxic chemotherapy, comprising six cycles of a platinum-based agent and a taxane [332]. Additionally, patients undergo surgery with cytoreduction, aiming to minimise residual disease; the success of this debulking surgery has a substantial impact on patient outcome [315]. Cytoreductive surgery is either performed at the time of diagnosis, or after 3-4 cycles of neoadjuvant chemotherapy [333].

The addition of another cytotoxic drug to this regimen has so far proven largely ineffective at improving survival [334]. However, addition of the anti-VEGF-A monoclonal antibody bevacizumab [335] with chemotherapy allows a modest increase in progression-free survival (PFS), although with no impact on overall survival (OS) [334, 335]. Additionally, the use of a Poly-ADP Ribose Polymerase inhibitor (PARPi) as maintenance therapy following initial chemotherapy cycles improves PFS in all patients, with greatest benefit in those with *BRCA1/BRCA2* mutations [336]. Consequently, maintenance therapy based on different combinations of PARPis and bevacizumab are now included in treatment regimens, the specifics of which depend on the HR status of patients as determined by testing of the primary tumour [335].

Despite a high 60-80% response rate to initial chemotherapy in HGSC [336], up to 70% of patients will relapse within 3 years [335]. Upon recurrence, patients can be treated with another platinum-based regime if they are considered platinum-sensitive (defined as recurrence more than 6 months after initial therapy) [332]. This may be combined with PARPi maintenance therapy, or bevacizumab [335]. Importantly, there are no validated predictors of response to platinum re-challenge and this definition of platinum sensitivity is debated [335, 336]. Surgery is also considered for platinum-sensitive recurrent HGSC in some cases, with potential improvements in OS and PFS [335].

If platinum treatment is not appropriate, single-agent treatment with other chemotherapies (such as gemcitabine, paclitaxel, or topotecan) may be used, as may bevacizumab [335]. However, there are no robust data from clinical trials to guide the choice of chemotherapy nor the duration of treatment for this patient cohort and the objective response rate to single-agent treatment at this stage is only $\sim$10-15% [335]. As such, there is a desperate need for more effective treatments at both diagnosis and relapse in HGSC.

1.9.4 The immune system in HGSC

Several clinicopathological features affect prognosis in HGSC, including age and stage at diagnosis, residual disease following cytoreduction and primary platinum resistance [337]. Additionally, genomic features including HR deficiency due to mutation in genes such as *BRCA1/BRCA2* or *RAD51C*, or due to epigenetic changes, are associated with greater response to platinum or PARPi and improved

OS [338, 339]. Aside from these features, evidence has accumulated for a strong immune association with prognosis.

A consensus of four subtypes of HGSC based on gene expression signatures has been defined: immunoreactive, differentiated, proliferative and mesenchymal [340]. This classification demonstrated that the immunoreactive subtype – with high levels of T cell infiltration – was associated with the best prognosis [340]. Indeed, the infiltration of CD8+ T cells (tumour-infiltrating lymphocytes, or TILs) is highly correlated with improved survival in HGSC, regardless of residual disease following cytoreduction, treatment with standard agents, or germline *BRCA1* mutation status [341, 342]. Other immune cell types have also been related to disease outcomes, such as NK cells [343]. NK cells often co-infiltrate with CD8+ T cells and their frequency in the ascitic fluid found in HGSC correlates with longer overall survival [343]. Additionally, the combination of CD8+ T cells and CD20+ B cells is positively prognostic, with greater prognostic value than CD8+ TILs alone [344]. Yet despite the extensive evidence for immune importance in HGSC, responses to immunotherapies such as PD1 or PDL1 blockade are disappointing, with a recent phase II study demonstrating an objective response rate of only 9% using the anti-PD1 pembrolizumab [340].

Extensive research is ongoing to identify mechanisms of immune evasion and alternative targets for immune stimulation. For example, several inflammatory mediators have been identified that beneficially modulate the microenvironment in HGSC, including the interferon-induced chemokines CCL5, CXCL9 and CXCL10 [345]. However, it has been demonstrated that immune evasion and loss of CD8+ T cell infiltration in EOC can be achieved via epigenetic silencing of *CCL5*, *CXCL9* and *CXCL10* genes [128, 346]. Additionally, recent data suggest that upstream innate immune signalling can influence the tumour microenvironment (TME) and patient outcome [347–349]. IFN and STAT1 signalling in HGSC tumours – with downstream CXCL9-11 chemokine expression – is associated with increased CD8+ T cell response and chemosensitivity [350]. These IFN pathways may be initiated by Stimulator Of Interferon Response CGAMP Interactor (STING) signalling in the TME, which is also demonstrated to contribute to the CD8+ T cell response in solid tumour models [347, 349] and to the response to PARPi [348]. Whilst most literature describes STING/IFN signalling in non-tumour cells, tumour cell-intrinsic STING activation has also been demonstrated to contribute to CD8+ T cell and NK cell infiltration in EOC [351, 352]. Hence, targeting of the STING-IFN-STAT1 axes lying upstream of chemokine expression and immune cell infiltration may improve patient outcome and treatment response.

It must be noted that STING signalling and IFN can have pleotropic roles in ovarian cancer [353]. Whilst promoting immunoreactivity in an acute inflammatory setting, constitutive inflammatory signalling may promote pro-tumourigenic immune resistance, as observed in *BRCA1*-mutant models of ovarian cancer [351]. The genotype of EOC is also suggested to affect the anti-tumour efficacy of

STING signalling [354]. Consequently, further investigations into innate immune and inflammatory signalling pathways are essential to harness anti-tumour immune effects in HGSC.

1.9.5 Epigenetic targeting of EZH2 and G9a in HGSC: an anti-tumour immune boost?

Numerous epigenetic therapies are under investigation in HGSC, including several in clinical trials [332]. As a cancer characterised by frequent *MYC* amplification, with prognostic significance of alterations in EZH2/G9a as well as immune factors, HGSC has clear potential for anti-tumour epigenetic targeting. Particularly, targeting of EZH2 or G9a enzymes in HGSC may provide beneficial immune stimulation and anti-tumour efficacy, due to their contribution to cMYC-mediated effects and their direct and indirect suppression of immune responses. Within HGSC cohorts, *MYC*-amplified HGSC may represent a particularly targetable subtype with high levels of immune repression, a decreased threshold for viral mimicry and an increased reliance on EZH2/G9a activity.

The interconnected nature of EZH2 and G9a, alongside their enzymatic redundancy and promiscuity in deposition of repressive marks, suggest dual targeting of both enzymes simultaneously may be most effective. In support of this hypothesis, our group has shown that dual EZH2/G9a targeting in a murine model of HGSC (*Trp53*$^{-/-}$ ID8) stimulates an immune response including increased effector CD8+ T cell and NK cell infiltration, accompanied by a survival benefit [155]. These initial data in HGSC support further investigation of dual EZH2/G9a targeting in *MYC*-aberrant cancer, to validate immune stimulation and therapeutic effects.

1.10 cMYC-deregulated cancer: lung adenocarcinoma (LuAd)

1.10.1 Overview

Lung cancer is the most common cause of cancer death [355], with lung adenocarcinoma (LuAd) comprising roughly 40% of all lung cancer cases (figure 1.16) [356]. LuAd is frequently diagnosed at a late stage and suffers from a poor five-year survival rate with little improvement over the last few decades [357]. Several important genetic drivers have been identified in LuAd; the most common mutations are in in the oncogenes *KRAS* and *EGFR* and in the tumour suppressor genes *TP53, KEAP1, STK11* and *NF1* [355]. Additionally, aberrant cMYC signalling is now recognised as an important contributor to LuAd [358].

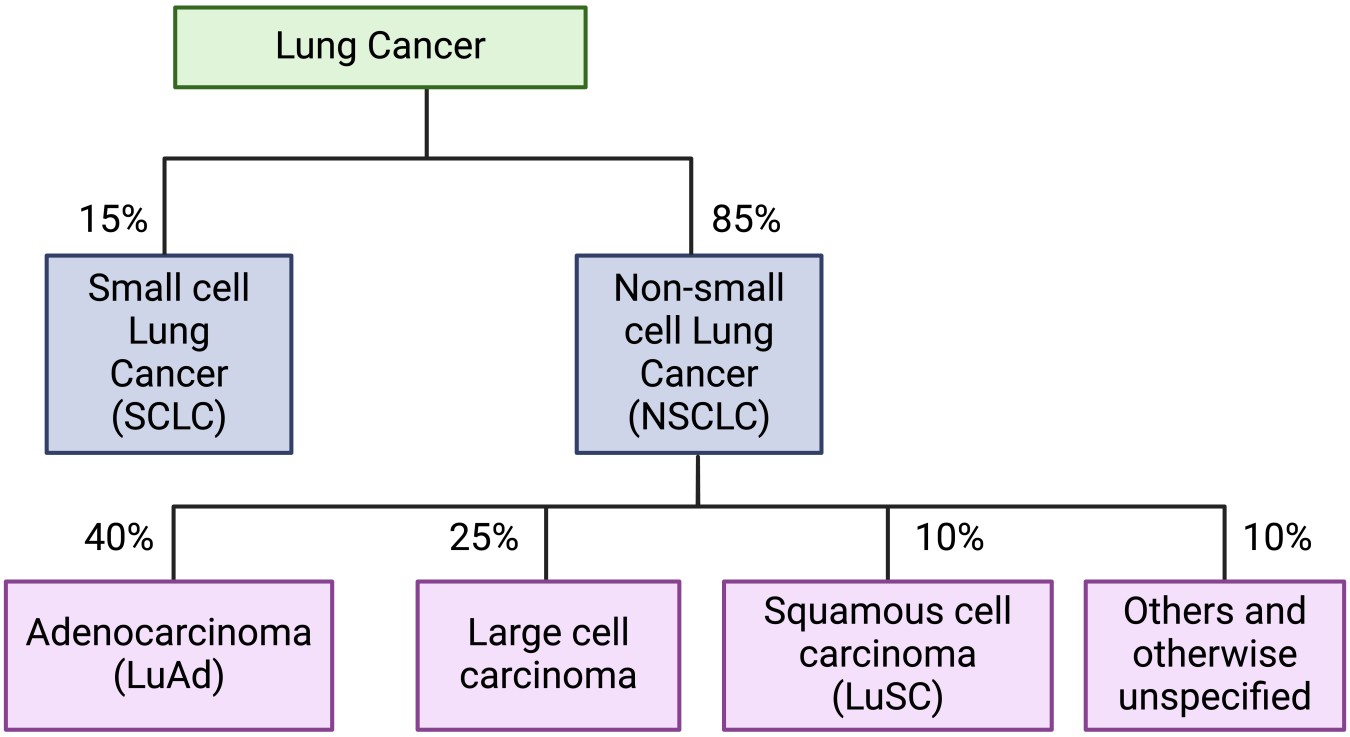


Figure 1.16 **Histological classifications of lung cancer and the proportion of total lung cancer incidence they constitute.** Figure reproduced from [356].

1.10.2 cMYC deregulation and the immune system in LuAd

Dysregulation of *MYC*, including overexpression or amplification, is frequently found in both small cell and non-small cell lung cancer subtypes [359]. *MYC* copy number gain has been shown to predict poor overall and disease-free survival [360]. It is recognised as a driver of progression and resistance to therapy, with roles in metastasis, metabolic reprogramming, angiogenesis and genetic instability, leading to extensive cMYC targeting efforts [359]. Especially, a large subset of LuAd tumours with the $KRAS^{G12D}$ mutation is dependent on cMYC activity to sustain tumourigenesis [309].

cMYC activity is implicated in immune evasion in lung cancer [359]. It has been shown that cMYC can upregulate immune checkpoint ligands [280, 361] and that cMYC in cooperation with oncogenic RAS drives a highly immunosuppressive microenvironment [362]. In a mouse lung model of adenoma driven by a $KRAS^{G12D}$ mutation, co-activation of cMYC results in the conversion of adenoma to a proliferative, invasive and immune-suppressed adenocarcinoma [362]. These tumours exclude adaptive T and B cells and innate immune NK cells, but strikingly the deactivation of cMYC reverses immune exclusion and promotes tumour regression [362]. cMYC has also been shown to repress interferon signalling in another *KRAS*-driven mouse model of LuAd, with the $KRAS^{G12C}$ mutation [363]. These data suggest that the large group of LuAd tumours with *KRAS* mutations and *MYC* alteration represent a relevant model system for investigating the disruption of cMYC-mediated immune repression via dual EZH2/G9a targeting.

1.11 Introduction to HKMTi-1-005

1.11.1 Identification of a compound with dual activity against EZH2 and G9a

As described, inhibition of both EZH2 and G9a simultaneously is required to reverse epigenetic silencing of certain gene targets [109]. Additionally, growth inhibition of tumour cells is markedly stronger when both enzymes are inhibited compared to either enzyme alone [109]. Yet the use of two single inhibitors simultaneously is considerably more complex to implement clinically than the use of a single compound with dual activity [364]. The following paragraphs describe the work undertaken by Curry et al. to identify compounds with dual effects against EZH2 and G9a activity, resulting in the creation of HKMTi-1-005 [109].

Given that the H3K9 and H3K27 lysine residues lie within similar ARKS motifs, and G9a is able to methylate H3K27, it was hypothesised that the histone substrate-binding pocket of EZH2 and G9a/GLP were similar. BIX-01294 is a G9a/GLPi that can bind this substrate-binding pocket, providing a rationale for the creation of a compound library based on the BIX-01294 quinazoline template. The library was screened to identify compounds with additional EZH2i activity with the aim of identifying peptide substrate-competitive dual EZH2/G9a inhibitors.

Initial EZH2i activity was determined by re-expression of two target genes of EZH2 repression, *KRT17* and *FBXO32*, in the breast cancer cell line MDA-MB-231. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) was used to identify three compounds capable of upregulating these genes in a dose-dependent manner: HKMTi-1-005, HKMTi-1-022 and HKMTi-1-011 (figure 1.17). The related single G9ai compounds BIX-01294 and UNC0638 had variable effects on these genes, whilst the single EZH2i GSK343 had no effect on any target genes. Chromatin immunoprecipitation (ChIP) was subsequently used to demonstrate a decrease in H3K27me3 and H3K9me3 at these target genes, accompanied in some cases by an additional increase in activating chromatin marks such as H3K4me2/3 or H3K27/H3K9Ac.

Having looked at specific gene targets, Curry et al. expanded to broader gene expression analyses using a microarray approach, focusing on HKMTi-1-005 as a representative hit compound. This was once again performed in the MDA-MB-231 cell line, with 24h treatments using either HKMTi-1-005, the EZH2i GSK343, or the G9ai UNC0638. A greater reversal of EZH2-mediated gene silencing – defined as those genes upregulated upon short interfering RNA (siRNA) knockdown of *EZH2* in MDA-MB-231 – was demonstrated by HKMTi-1-005 than either of the individual inhibitors, although both compounds did demonstrate significant reversal of repression. However, investigation of another EZH2 geneset that was defined using siRNA knockdown of *EZH2* in a different breast cancer cell

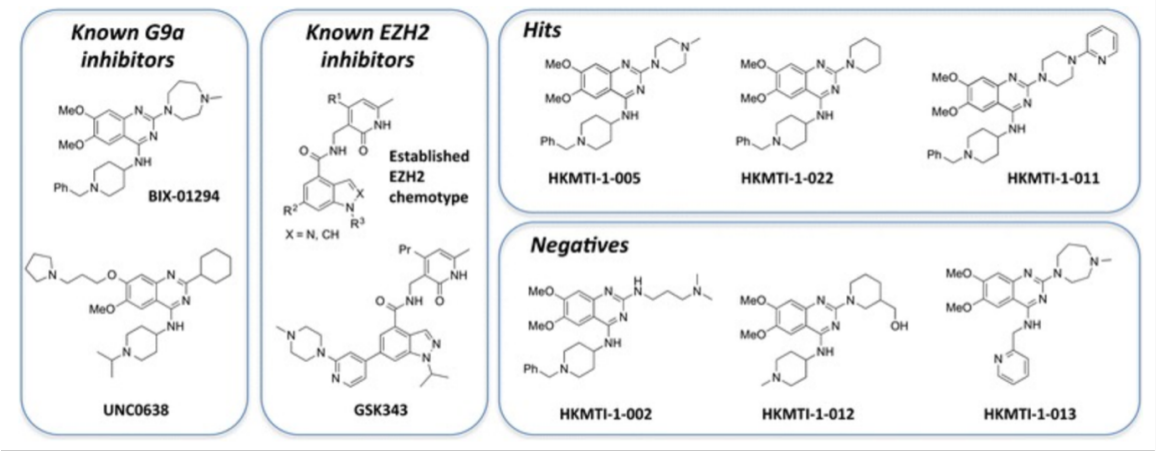


Figure 1.17 **Chemical structure of HKMT inhibitors.** Figure taken from [109], showing the chemical structure of known G9a and EZH2 inhibitors and the results of the library screen for dual inhibitors of both enzymes, including the HKMTi-1-005 hit.

line demonstrated no upregulation by any of the compounds. These data suggested cell type-specific EZH2 targets. Therefore, a meta-analysis was performed to identify consensus EZH2 repression targets using 18 siRNA EZH2 studies; this meta EZH2 repression geneset acts as a pharmacodynamic biomarker of EZH2i activity and showed highly significant upregulation in MDA-MB-231 following HKMTi-1-005 treatment.

In addition to global gene expression analysis, histone marks were also examined globally in MDA-MB-231 cells. Following HKMTi-1-005 treatment, western blot analysis demonstrated that H3K27me3 and H3K9me3 were reduced in a dose-dependent manner. This was achieved using a concentration of 2.5μM HKMTi-1-005, as compared to the cell viability half-maximal inhibitory concentration (IC_{50}) value of 4.3μM in these cells, supporting activity of the compound at concentrations below those inducing loss of cell viability. Additionally, HKMTi-1-005 treatment upregulated many apoptosis-associated genes and induced apoptotic cell death in a dose-dependent manner, whilst UNC0638 and GSK343 failed to do so. Notably, the IC_{50} values for BIX-01294, UNC0638 and GSK343 in terms of both cell viability and clonogenic growth were consistently higher than for HKMTi-1-005.

Given the known sensitivity of *EZH2*-mutant cell lines to EZH2i [145], the sensitivity of a panel of lymphoma and breast cancer lines with different *EZH2* statuses were examined. These data suggested that HKMTi-1-005 is particularly effective against *EZH2*-mutant cell lines, as observed with the GSK343. However, the results also demonstrated that, in contrast to GSK343 sensitivity, alternative *EZH2* deregulation such as amplification or increased expression may also increase sensitivity to the HKMTi-1-005 [109]. Importantly, normal breast epithelial cells were markedly less sensitive to HKMTi-1-005 treatment than breast cancer cells.

Finally, the biochemical properties of the hit compounds were investigated. First, the inhibitory activity against EZH2 was examined. To do so, a minimal trimeric PRC2 complex of EZH2, EED and

SUZ12 was used to mimic the *in vivo* PRC2 complex, with a biotinylated-H3 as the peptide substrate. The data revealed similar inhibitory concentrations to those found in the cell-based assays, of 24, 12 and 16μM for HKMTi-1-005, HKMTi-1-011 and HKMTi-1-022 respectively. However, cell-free assays must be accompanied by a caveat; the interpretability of these results is limited by the use of a minimal trimeric complex in place of the multimeric PRC2 complex found in cellular environments, and the use of a peptide substrate as a substitute for the hugely intricate chromatin network that would usually act as the EZH2 substrate. Additional biochemical analyses were undertaken to confirm the mechanism of action of the HKMTi-1-005 hit, demonstrating a peptide substrate-competitive mechanism distinct from other known EZH2 inhibitors. Selectivity for EZH2 and G9a activity was also assessed, verifying minimal activity against 15 other methyltransferases including DNMT1, SUV39H2 and SETDB1 at up to 100μM. Note that although direct activity of HKMTi-1-005 against GLP was not measured, the BIX-01294 scaffold from which the compound derives does inhibit both G9a and GLP due to the high homology between the SET domains of each enzyme [32]. However, BIX-01294 has a roughly 20x selectivity for G9a over GLP [365]. Hence, it is assumed that HKMTi-1-005 achieves some inhibitory activity against GLP, although to a much smaller degree than that of G9a.

These data suggest that HKMTi-1-005 is a compound with dual activity against the EZH2 and G9a methyltransferases in breast cancer cells, with selectivity for its target enzymes over other HKMTs, and for tumour cells over non-tumour cells. They further demonstrate substrate-competitive activity of the compound and suggest direct EZH2i activity. Yet, the models investigated in this proof-of-concept publication are largely breast cancer cell lines and the cell-free assays must be interpreted with caution. Hence, prior to clinical evaluation of HKMTi-1-005, several aspects of the pre-clinical work must be strengthened.

1.12 Aims and objectives

An overview of the broad aims and objectives of this thesis are summarised as follows:

- To consolidate and characterise the disruption of H3K9 and H3K27 methylation maintenance by HKMTi-1-005, in HGSC cell lines and a murine model of HGSC.
- To investigate the induction of viral mimicry by HKMTi-1-005 in HGSC models, including induction of transposable elements and the nature of immune stimulation.
- To examine cell-intrinsic cMYC-mediated immune repression and its relationship to HKMTi-1-005 effects and sensitivity.

Chapter 2

Methods

2.1 Overview of bioinformatic tools

An overview of the core bioinformatic tools used in this thesis are included in tables 2.1 and 2.2. These tools were largely used via the Imperial College London High-Performance Computing facilities or using R [366] and RStudio [367].

2.2 HGSC cell lines

All HGSC cell culture, treatment and nucleic acid extraction were performed by Dr Ian Garner. An overview of the analyses performed on HGSC models is given in figure 2.1.

2.2.1 HGSC cell line details

PEO1, PEO4, PEA2, OVCAR5, OVCAR8 and IGROV1 HGSC cell lines were obtained from Drs Langdon and Melton, Institute of Genetics and Molecular Medicine, Edinburgh University. PEO1 and PEO4 are derived from the same patient at different sites and different stages in treatment: PEO1 from the peritoneal ascites of a patient with *BRCA2*-mutated HGSC after initial treatment with cisplatin, chlorambucil and 5-fluorouracil, and PEO4 from pleural effusions of the same patient but after acquired resistance to these chemotherapies, with a secondary mutation in *BRCA2* that restored the open reading frame [368]. PEA2 is derived from the ascites of a patient with HGSC after relapse from cisplatin and prednimustine therapy [368]. OVCAR5 is from pre-treatment ascites of a patient with ovarian adenocarcinoma, whilst OVCAR8 is from the primary tumour of a patient with mixed carcinoma but after treatment with high-dose carboplatin [369]. IGROV1 was initially thought to be a HGSC cell line, however recent analyses have suggested it is instead endometroid/clear cell ovarian line and it was originally derived from an endometrioid tumour with areas of serous, clear

cell and undifferentiated carcinoma [370].

Package/tool suite	Function/command	Usage
AnnotationDbi	Many, e.g. org.Mm.eg.db, org.Hs.eg.db etc.	Interface for querying many different annotation data packages, e.g. allows access to transcriptome and genomic data packages
BEDTools	merge, intersect, subtract	"Genome arithmetic": conversion between genomic file formats and intersect/merge/count genomic intervals across files
biomaRt	useMart, getLDS	Conversion between mouse and human orthologous gene annotations
Bowtie2	bowtie2-build, bowtie2	Aligning assay for transposase-accessible chromatin with sequencing (ATACseq) reads to a reference genome (HGSC lines)
BWA-MEM	bwa index, bwa mem	Aligning ATACseq sequencing reads to a reference genome (all ID8 samples)
car	vif, avPlots	Calculate variance inflation factors ahead of Cox proportional hazards (CoxPH) analyses, create added-variable plots
ChIPseeker	annotatePeak, plotAnnoBar, plotDistToTSS	Annotation and visualisation of genes/TEs in ATACseq data, and distance from peaks to genomic features
CIBERSORTx	Online web tool	Infer absolute immune composition based on bulk patient RNA data
Consensus^{TME}	consensusTMEAnalysis	Infer relative immune composition based on bulk patient RNA data
deepTools	alignmentSieve	Read shifting +4bp for positive strand reads and -5bp for negative strand reads to account for Tn5 insertion, which creates a 9bp duplication
DESeq2	Many, e.g. DESeqDataSetFromMatrix, DESeq, results, rlog, plotPCA etc.	Detection of differentially expressed genes, visualisation of data and results
DiffBind	Many, e.g. dba.overlap, dba.peakset, dba.count, dba.plotPCA, dba.plotProfile etc.	Analysis of ATACseq data using differentially accessible chromatin region (DAC) method (see 3.15) and visualisation of results
edgeR	DGEList	Assemble count matrix
EnhancedVolcano	EnhancedVolcano	Create volcano plots for visualisation of differential expression and DAC results
EnrichR	Online web tool	Perform over-representation analysis (ORA) on mouse RNAseq data
FastQC	fastqc	Quality assessment of sequencing data
fgsea	fgsea	Fast gene set enrichment analysis (GSEA) on preranked data
finalfit	finalfit	Create results tables for CoxPH model outputs
geneSCF	geneSCF	Perform ORA on HGSC cell line RNAseq data
GenomicRanges	GRanges, findOverlaps, additional basic operations	Intersect ATACseq results with TE reference files, manipulate genomic range data within R
ggpubr	stat_compare_means, stat_cor, ggbarplot, ggboxplot, ggviolin	Customize plots and add statistical comparisons
ggstatsplot	ggbarstats	Plot bar chart with Fisher's exact test results

Table 2.1 **Core bioinformatic tools: part 1.** Table listing main bioinformatic tools used in analyses. Continued overleaf in part 2.

Package/tool suite	Function/command	Usage
HOMER	annotatePeaks.pl, findMotifsGenome.pl	Annotate ATACseq peaks of accessibility, find enriched TF motifs in ATACseq peaks
Macs2	callpeak	Call peaks of accessibility in ATACseq data
msigdbr	msigdbr	Retrieve hallmark/GOBP genesets from the Molecular Signatures Database (MSigDB) mouse and human collections
MultiQC	multiqc	Collate quality control (QC) analyses into a single report
pheatmap	pheatmap	Drawing of heatmaps to demonstrate sample clustering in sequencing data
Picard	CollectInsertSizeMetrics, MarkDuplicates, SortSam	Processing and QC of ATACseq data, including collection of inzert size distribution metrics and removal of duplicate reads
Rsubread	featureCounts	Assign mapped reads to genomic features (e.g. genes, TEs)
rtracklayer	import, liftOver	Retrieve reference GTF files, lift genomic coordinates between genome builds
samtools	view, index, sort, quickcheck	Genomic file manipulation e.g. indexing, filtering, sorting by genomic coordinate
singscore	rankGenes, multiScore, plotDispersion, plotRankDensity	Score patient data and cell lines for genesets, visualise results
STAR	STAR	Create genome indices and align reads to reference genome for RNAseq data
stats	Many, e.g. lm, setNames, fisher.test etc.	Statistical testing in R
SuperExactTest	supertest, plot	Statistical testing and visualisation of multiset interactions
survival	Surv, coxph, cox.zph	Test proportional hazards, build CoxPH model, visualise results
survminer	ggsurvplot, ggcoxzph, ggcoxdiagnostics	Visualisation of survival analysis results as Kaplan-Meier curves, graphical testing of CoxPH assumptions
TCGABioLinks	Many, e.g. GDCquery, GDCdownload, getResults, GDCprepare etc.	Download and manipulate TCGA data
Telescope	telescope assign	Estimation of TE expression at a locus-specific level, ahead of differential expression analysis
TEtranscripts	TEtranscripts	Differential expression analysis of TEs at a subfamily level
tidyverse	ggplot2, dplyr, tidyr, readr, tibble, tidyr, stringr, forcats	Set of packages in R used for many analyses to read in, wrangle, tidy and visualise data
Trimmomatic	trimmomatic	Read trimming, adapter removal
WebGestalt	Online web tool	Classical GSEA for RNAseq data

Table 2.2 **Core bioinformatic tools: part 2.** Table listing main bioinformatic tools used in analyses.

2.2.2 Cell culture and HKMTi-1-005 treatment

Cell lines were maintained in Roswell Park Memorial Institute 1640 medium (RPMI; Sigma Life Science, Missouri, USA), supplemented with 10% foetal calf serum (FCS) and 2mM L-glutamine. Incubation of cells was in a 10% CO2 incubator at 37°C, with sub-culturing when cells reached 70% confluency. Ahead of harvesting cells for RNAseq and ATACseq protocols, cells were seeded at a density of 250,000 cells per well for 24h, followed by treatment with HKMTi-1-005 at their respective IC_{50} values for 72h in duplicate (PEA2, OVCAR5, OVCAR8, IGROV1) or triplicate (PEO1, PEO4). IC_{50} values were determined by Dr Ian Garner via MTT assays, a cell viability analysis based on the ability of cells to metabolise the MTT reagent into a purple product that can subsequently be measured via colorimetric assay. IC_{50} values for PEA2, PEO1, PEO4, OVCAR5, OVCAR8 and IGROV1 were 9.5µM, 5.7µM, 8.9µM, 10.7µM, 9.6µM and 7.6µM respectively (Dr Ian Garner, personal communication).

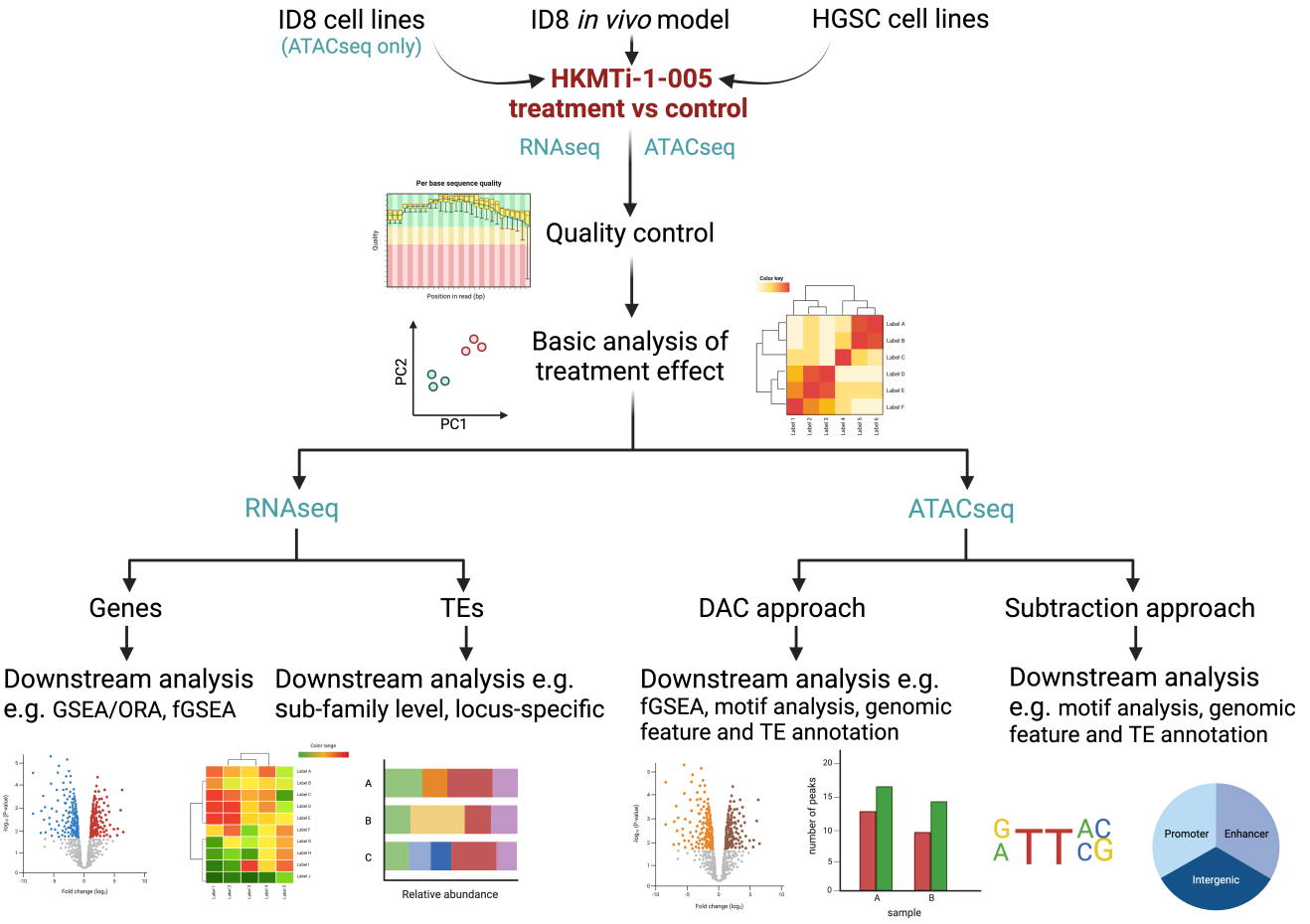


Figure 2.1 **Schematic overview of RNAseq and ATACseq analyses of HKMTi-1-005 treatment effect.** Adapted from templates by BioRender.com (2024), retrieved from `https://app.biorender.com/biorender-templates`.

2.2.3 RNAseq

RNAseq library preparation and sequencing

RNAseq preparation was performed according to the New England Biolabs protocol (accessible at `https://www.neb.com/en-gb/products/e7760-nebnext-ultra-ii-directional-rna-library-prep-kit-for-illumina`), with rRNA depletion. Paired-end sequencing was performed at the Institute of Cancer Research (ICR) ahead of downstream processing.

RNAseq data processing: genes

Sequencing quality of FASTQ files was assessed using FastQC [371] and MultiQC [372] to collate reports. Reads were trimmed using Trimmomatic [373], with the following parameters: LEADING:5 TRAILING:5 SLIDINGWINDOW:4:12 MINLEN:70 (MINLEN:30 for PEO1 and PEO4 where reads were only 50bp). Adapters were removed using the NexteraPE-PE.fa file supplied by Trimmomatic, via the ILLUMINACLIP:NexteraPE-PE.fa:2:30:10 parameter. STAR was used for genome alignment [374]. First, genome indices were created with STAR --runMode genomeGenerate using the GRCh38 (hg38) release 79 reference genome (`https://ftp.ensembl.org/pub/release-79/gtf/homo_sapiens/`, [375, 376]). STAR was then run to allow alignment to this reference, with the following parameters: --outSAMtype BAM SortedByCoordinate --quantMode TranscriptomeSAM --outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread0 --outFilterMatchNmin 0. QC was once again performed using FastQC and MultiQC to assess quality of alignment [371, 372]. Blacklisted regions were removed using BEDTools [377] (bedtools intersect -abam) and the hg38 blacklist file from the Boyle Lab collection (available at `https://github.com/Boyle-Lab/Blacklist/tree/master/lists/Blacklist_v1`). Samtools [378] was used to index bam files. featureCounts [379] from the RSubread package was run with the following parameters to achieve reverse-stranded, paired-end counting with exclusion of chimeric fragments: strandSpecific=2,isPairedEnd=TRUE, countChimericFragments=FALSE. Subsequently, differential expression analyses were performed using DESeq2 [380] which uses the median of ratios (MoR) normalisation method for counts. Gene annotations were added using the AnnotationDbi [381] function mapIDs ahead of visualisation of results and enrichment analyses. IGROV1 was excluded from downstream analyses due to limited treatment effect and unclear tumour origin, as described in 3.2.1.

RNAseq data processing: TEs

To account for the repetitive TE elements, FASTQ files were re-aligned using STAR to allow for multimapping of reads, as recommended [382] (--outFilterScoreMinOverLread0--outFilterMatchNminOverLread0--outFilterMatchNmin0--outFilterMultimapNmax100--winAnchorMultimapNmax100)

[374]. Files were indexed and sorted using samtools [378]. TEtranscripts [382, 383] was then used to perform differential expression at the subfamily level using the --stranded reverse parameter and the hg38_rmsk_TE.gtf from the Hammell lab repository (`https://labshare.cshl.edu/shares/mhammelllab/www-data/TEtranscripts/TE_GTF/`). For locus-specific analyses, Telescope [384] was used with the telescope assign command and --attribute transcript_id parameter to create count files. These count files were then collated for input into DESeq2 for differential expression analyses.

2.2.4 ATACseq

ATACseq library preparation and sequencing

PEO1 and PEO4 cell lines were included in ATACseq analysis. The protocol described by Brunton et al. [385] was followed after cell harvest for ATACseq. Samples were sequenced at the ICR ahead of downstream processing. A diagrammatic explanation of the two ATACseq approaches described below is given in the relevant results section (see figure 3.15).

ATACseq data processing

FASTQ files were quality checked and trimmed using FastQC, MultiQC and Trimmomatic as described for RNAseq data [371–373]. Genome indices were downloaded and unzipped using wget and unzip commands, from `https://genome-idx.s3.amazonaws.com/bt/GRCh38_noalt_as.zip`. Reads were aligned to the genome with bowtie2 using the --very-sensitive -X2000 parameters [386]. Aligned sam files were converted to bam files using samtools [378], followed by FastQC and MultiQC [371, 372] to assess quality of alignment. A threshold of minimum 25M reads, at least 90% total alignment and a maximum 55% multiple alignments was used to filter out low quality samples, resulting in the removal of the third PEO4 untreated replicate. Files were indexed with samtools [378] and mitochondrial reads were removed using a grep command, followed by collection of insert size distribution data using the collectInsertSizeMetrics tool from Picard, with subsequent removal of PCR duplicates using the Picard MarkDuplicates tool (`http://broadinstitute.github.io/picard/`). Further filtering included removal of unmapped reads, low quality reads (MAPQ < 30, reflecting low mapping quality), non-primary reads and any duplicate reads using samtools, to retain only properly paired reads as recommended by ENCODE [387]. Reads were shifted using alignmentSieve from the deepTools suite to account for Tn5 insertion [388]. Samtools was used to index bam files [378]. Finally, following sorting and indexing of bam files, the Macs2 callpeak function was used to call peaks of accessibility with the following parameters to specify paired-end mode, human genome and to keep all reads: -f BAMPE -g hs -B --keep-dup all [389]. Blacklist regions and mitochondrial alignments were removed using BEDTools [377] with the hg38 blacklist file described, before the total

number of peaks in each treatment condition were counted.

ATACseq: DAC approach

DACs were calculated using DiffBind [390], a tool originally created for ChIPseq data but with additional functionality for ATACseq. A detailed documentation of the package and pipeline can be found in the vignette (`https://bioconductor.org/packages/release/bioc/html/DiffBind.html`). Briefly, the filtered narrowPeak files from Macs2 output were read in and a consensus peakset was created (with a minimum overlap of 2 samples for each cell line). Reads that overlap the consensus peakset intervals were counted, counts were normalised and DESeq2 was used to establish differentially accessible sites with associated false discovery rate (FDR) and log2 fold-change (log2FC) values. The default normalisation was used (DESeq2 method), however multiple options are available for the choice of reference reads including either the use of all background peaks ("background") or the use of peaks that overlap consensus peaks only, known as reads in peaks (RiP). Calculation of the fraction of reads in peaks (FRiP) demonstrated consistently lower values in treated versus untreated samples for the PEO1 and PEO4 lines, due to the increased number of total reads in treated samples (table 2.3). To reduce bias of normalisation based on total library size, the RiP reference read approach was chosen. Internal functions from the DiffBind packages were used for plotting results in the form of principal component analysis (PCA) plots, heatmaps and volcano plots. Following DiffBind, peaks were annotated with gene labels using ChIPseeker [391]. TEs were annotated using the GenomicRanges findOverlaps package to assess DAC overlap with the hg38_rmsk_TE.gtf reference used in RNAseq analyses.

ID	Condition	Replicate	Reads	FRiP
PEO1_UT1	Control	1	6696378	0.44
PEO1_UT2	Control	2	9983775	0.40
PEO1_UT3	Control	3	11711522	0.39
PEO1_T1	Treated	1	15605391	0.28
PEO1_T2	Treated	2	50780410	0.24
PEO1_T3	Treated	3	14317591	0.32
PEO4_UT1	Control	1	24075442	0.36
PEO4_UT2	Control	2	9202281	0.38
PEO4_T1	Treated	1	44993085	0.25
PEO4_T2	Treated	2	33309695	0.23
PEO4_T3	Treated	3	33806772	0.20

Table 2.3 **FRiPs and total reads in PEO1 and PEO4 ATACseq samples.**

ATACseq: subtraction approach

Using BEDTools [377], peak files for each condition were sorted and merged, then untreated peaks were subtracted from treated peaks. Peaks were sorted and merged again to generate bed files with peaks of accessibility unique to the HKMTi-1-005-treated condition. TE annotation was performed using the GenomicRanges findOverlaps package to assess peak overlap with the hg38_rmsk_TE.gtf reference. Genes in peaks were annotated using the annotatePeaks.pl tool from the HOMER suite [392], with the hg38 genome.

2.3 ID8 *in vivo* samples

2.3.1 ID8 *in vivo* model details

Further details on the ID8 model of HGSC are available in [393, 394]. In brief, the model is a syngeneic murine model [393] with bi-allelic deletion of *Trp53* [394].

2.3.2 HKMTi-1-005 treatment and sample collection

ID8 *in vivo* experiments were performed by Dr Pavlina Spiliopoulou, in accordance with the U.K. Home Office Animal (Scientific Procedures) Act 1986, under the project licences 70/7997 and PA780D61A. Treatment, collection and preliminary processing of the ID8 *in vivo* RNA and ATAC data are described in [155]. To summarise, 6-week-old female C57BL/6J mice were injected intraperitoneally with 5 x 10^6 *Trp53*$^{-/-}$ cells per mouse to initiate tumour formation. Mice were treated with 20mg/kg HKMTi-1-005 intraperitoneally, twice daily from days 21-34 after injection with *Trp53*$^{-/-}$ cells, or with vehicle (1% Tween, 3.4% DMSO in 0.9% NaCl) twice daily for the same duration. The dosage of 20mg/kg was based on pharmacokinetic data from previous simulation studies with HKMTi-1-005 and tolerability studies in mice. Tolerability studies had shown no acute toxicity and less than 10% weight loss in mice at this dose. Tumours were harvested on day 35 for ATACseq and RNAseq experiments. For survival analyses, mice were killed when a humane endpoint was reached. Raw RNAseq and ATACseq data were deposited at the European Nucleotide Archive (ENA), available under the primary accession number PRJEB44851, secondary accession number ERP12894.

2.3.3 RNAseq

RNAseq data processing: genes

FastQC and MultiQC analyses performed in initial analyses [155] demonstrated sufficient quality of FASTQ files from RNA sequencing runs, without read trimming. Reads were aligned us-

ing STAR [374] following creation of genome indices with --runMode genomeGenerate and the GRCm39 (mm39) release 104 (`http://ftp.ensembl.org/pub/release-104/gtf/mus_musculus/`, [375, 376]). Alignment used the following parameters: --outSAMtype BAM SortedByCoordinate --quantMode TranscriptomeSAM --outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0 --outFilterMatchNmin 0. FastQC and MultiQC [371, 372] were used to assess and collate alignment quality, ahead of removal of blacklisted regions using BEDTools [377]. As the mm39 blacklist file was not yet available at the time of writing, the liftOver file based on mm10 coordinates was used (downloaded from `https://github.com/Boyle-Lab/Blacklist/issues/25`). Read counting was performed using featureCounts from the RSubread package [379] with the following parameters to ensure reverse-stranded, paired-end reads were counted: strandSpecific=2, isPairedEnd=T, requireBothEndsMapped=T, allowMultiOverlap=T. DESeq2 [380] was used to assess differential expression of genes with annotation using the mapIDs function from the AnnotationDbi package [381] ahead of enrichment analyses and visualisation of results.

RNAseq data processing: TEs

STAR alignment for TE analysis utilised different parameters to allow for multimapping of reads: --outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0 --outFilterMatchNmin 0 --outFilterMultimapNmax100 --winAnchorMultimapNmax 100 [374]. Following indexing and sorting of output files using samtools, TE analyses were first performed at the subfamily level using TEtranscripts [382, 383], with the --stranded reverse parameter and using the mm39 RepeatMasker annotation file (available at `https://labshare.cshl.edu/shares/mhammelllab/www-data/TEtranscripts/TE_GTF/`). Subsequent locus-level analyses utilised the Telescope package [384], using the telescope assign command on the same alignment input used for TEtranscripts with the mm39 RepeatMasker annotation. Output files were utilised in DESeq2 analysis to calculate differential expression of TEs.

2.3.4 ATACseq

ATACseq: data processing

Previously generated MultiQC data [155] from raw files were assessed, resulting in exclusion of the H090003 (treated) sample with fewer than 20M reads. FASTQ files were trimmed with Trimmomatic [373] and the following parameters to remove adapter contamination and short reads: ILLUMINACLIP:NexteraPE-PE.fa:2:30:10LEADING:5TRAILING:5SLIDINGWINDOW:4:12MINLEN:30. FastQC and MultiQC [371, 372] were run to assess data quality ahead of read alignment. BWA-MEM [395] was used in place of Bowtie2 due to space limitations on the server. Genome indices were generated with default parameters for the GRCm39 genome, ahead of bwa-mem alignment using

ID	Condition	Replicate	Reads	FRiP
ID8_T1	Treated	1	13423234	0.18
ID8_T2	Treated	2	106861956	0.11
ID8_UT1	Control	1	68322726	0.19
ID8_UT2	Control	2	83872747	0.16
ID8_T3	Treated	3	23252628	0.17
ID8_T4	Treated	4	22448938	0.14
ID8_T5	Treated	5	36831294	0.16
ID8_UT3	Control	3	60313466	0.21
ID8_UT4	Control	4	40075064	0.16
ID8_UT5	Control	5	62803029	0.11
ID8_UT6	Control	6	34385272	0.23

Table 2.4 **FRiPs and total reads in ID8 *in vivo* ATACseq samples.**

default parameters [395]. Sam output files were sorted by coordinate and converted to bam files using picard SortSam (`http://broadinstitute.github.io/picard/`). Duplicates were removed using the picard MarkDuplicates tool and insert size distribution was measured using InsertSizeMetrics from the same suite. Files were indexed and further filtered according to the same criteria described for PEO1 and PEO4 samples: unmapped reads, low quality reads (MAPQ < 30, reflecting low mapping quality), non-primary reads and any duplicate reads were removed using samtools, to retain only properly paired reads as recommended by ENCODE [387]. Reads were indexed and shifted to account for the insertion of Tn5 transposase using alignmentSieve [388], files were sorted and indexed using samtools, and peaks were called. Macs2 was once again used for peak calling [389], with the same parameters as for HGSC cell lines but with the mouse effective genome size parameter (-g mm). Blacklist regions were removed using BEDTools [377] and a grep command, as were regions aligning to the mitochondria.

ATACseq: DAC and subtraction approaches

DiffBind was used for DAC analysis of ID8 *in vivo* ATACseq data [390]. The same pipeline was used as described for HGSC cell lines. Exceptions included the use of a minimum of 4 overlapping samples for creation of the consensus peak file (minOverlap=4) and the necessary use of bUseSummarizeOverlaps = F when reads aligning to the peakset were counted. FRiPs and total number of reads for mouse data are included in table 2.4. The "subtraction method" analysis of ATACseq data was identical to that used for HGSC lines, with the use of the mouse mm39 reference genome in place of the hg38 human reference.

2.4 ID8 *in vitro* samples

All ID8 *in vitro* cell culture, treatment and nucleic acid extraction were performed by Dr Ian Garner.

2.4.1 ID8 cell line details

The ID8 *in vitro* cell lines are spontaneously immortalised lines originally derived from the mouse ovarian surface epithelium, with similarity to human epithelial ovarian cancer (Cellosaurus accession: CVCL_IU14). Four genotypes were used in ATACseq experiments: ID8 *Trp53*$^{-/-}$; *Brca1*$^{-/-}$ (referred to as BRCA1) [396], ID8 *Trp53*$^{-/-}$;*Brca2*$^{-/-}$ (referred to as BRCA2) [394], ID8 *Trp53*$^{-/-}$ (referred to as F3) [394] and the ID8 CRISPR control (referred to as C3) [394].

2.4.2 Cell culture and HKMTi-1-005 treatment

Cell lines were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 4% FCS and ITS Media Supplement, with incubation of cells in a 10% CO2 incubator at 37°C and sub-culturing when cells reached 70% confluency. Ahead of harvesting cells for RNAseq and ATACseq protocols, cells were seeded at a density of 250,000 cells per well for 24h. Cell lines were treated with HKMTi-1-005 at their IC_{50} values for 72h in duplicate, as described for HGSC cell lines. IC_{50} values for BRCA1, BRCA2, F3 and C3 were 5.7µM, 5.3µM, 7.2µM and 6.3µM respectively.

2.4.3 ATACseq

ATACseq protocol

Harvested cells were processed using the ATACseq protocol described by Brunton et al. [385], followed by sequencing at the ICR.

ATACseq: data processing

Data processing was performed according to an identical pipeline to that described for the ID8 *in vivo* ATACseq samples. Due to a low read count (<25M), one treated C3 sample was excluded, which resulted in only one treated replicate for this genotype. Therefore, C3 was not included in downstream DAC or subtraction analyses.

ATACseq: DAC and subtraction approaches

DAC analyses utilised DiffBind as described for ID8 *in vivo* samples [390]. A minimum of 2 overlaps were required between replicates for creation of the consensus peakset. Given the *in vitro* data did

ID	Condition	Replicate	Reads	FRiP
F3_UT1	Control	1	63782231	0.48
F3_UT2	Control	2	45320829	0.45
F3_T1	Treated	1	49039501	0.50
F3_T2	Treated	2	75026274	0.50
BRCA2_UT1	Control	1	50967620	0.48
BRCA2_UT2	Control	2	67964370	0.48
BRCA2_T1	Treated	1	65848019	0.42
BRCA2_T2	Treated	2	55133665	0.48
BRCA1_UT1	Control	1	39728165	0.32
BRCA1_UT2	Control	2	38200035	0.31
BRCA1_T1	Treated	1	38327795	0.39
BRCA1_T2	Treated	2	43950761	0.36

Table 2.5 **FRiPs and total reads in F3, BRCA1 and BRCA2 ID8 *in vitro* ATACseq samples.**

not show a skew towards lower FRiPs or more total reads in treated samples (see table 2.5), reference reads included all background peaks for normalisation.

2.5 RNAseq: downstream analyses

For visualisations of the distance between replicates of a sample, the pheatmap package was used. PCA plots were made using the DESeq2 plotPCA package [380], whilst EnhancedVolcano was used to create volcano plots [397]. For the HGSC cell lines, ORA was performed using geneSCF [398] with entrezIDs and the Gene Ontology Biological Process (GOBP) database, including all genes that had a non-zero baseMean value in the DESeq2 output as the background for geneSCF runs. For the ID8 *in vivo* data, ORA used the EnrichR platform [399–401]. Input genes for all ORA analyses were those with adjusted p-value (p-adj) <0.05 in either direction. Classical GSEA was performed using WebGestalt [402] for both human and mouse analyses, once again using the GOBP database with an additional run using the Hallmark database. The enrichment statistic (p) parameter was set to 0. Input for GSEA analyses was the stat column (representing the Wald statistic, log2FC divided by the standard error) from DESeq2 output. Fast gene set enrichment analysis (fGSEA) was created as an implementation of GSEA, but tends to outperform classical GSEA [403], is much faster and is more suited for accurate estimation of small p-values, especially where sampling depth is low [404]. fGSEA analyses were performed on the same gene inputs as for classical GSEA, using the fgsea package fgseaMultiLevel function with the following parameters: sampleSize = 101, minSize = 5, maxSize = 1100, eps = 0, scoreType = c("std", "pos, "neg"), nproc = 0, gseaParam = 1, BPPARAM = NULL, nPermSimple = 10000, absEps = NULL. P-value outputs from fGSEA are corrected for multiple testing using the Benjamini-Hochberg (BH) adjustment. The msigdbr package was used to

access GOBP and pathways. Other genesets used are described in table 2.6.

2.6 ATACseq: downstream analyses

To assess the number of control versus HKMTi-1-005-treated peaks of accessibility, filtered peaks were counted using a wc -l command on narrowPeak output files from Macs2 and a mean was taken across replicates. fGSEA analyses were performed using the fgsea package and the same parameters as described for RNASeq data, using log2FC*-log10(p-adj) from the DAC ATACseq analyses as ranking input (due to the lack of a "stat" column output from DiffBind). For plotting accessibility over regions associated with specific histone marks, ENCODE coordinates derived from ChIPseq assays in H1 ESCs were downloaded, available at `https://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeBroadHistone/`. The equivalent histone mark-associated regions from mouse neural tube from 12.5-day embryos were also downloaded, available at `https://www.encodeproject.org/reference-epigenomes/ENCSR273IZZ/`. H3K9me1/2 tracks were not available for these ChIPseq data. Coordinates were converted from hg19/mm10 to hg38/mm39 coordinates using the rtracklayer liftOver tool [405], followed by creation of bed files of coordinates and heatmap plotting over the associated regions using DiffBind [390]. Plotting used the following parameters and the dba.plotProfile function to plot an additional 10% upstream and downstream of each histone region of interest, with each region divided across 100 bins: style="percentOfRegion", nOfWindows=100, distanceAround=10. Bivalent regions were defined using bedtools intersect [377] to find the overlap of regions associated with different histone marks. Enhancer elements were defined using the Fantom5 database [406] in hg19/mm10 coordinates and once again converted to hg38/mm39 using liftOver [405]. Motif enrichment was performed using the findMotifsGenome.pl tool from the HOMER suite with the default parameters [392]. ChIPseeker [391] was used to annotate the genomic features (including promoter-TSS, exon, 5'UTR, 3'UTR, intron and intergenic regions) associated with DACs, as well as the distance from the transcription start site (TSS), with the TSS region defined as +/-1kb from the TSS. Volcano plots of changes in gene or TE accessibility were made using EnhancedVolcano [397].

2.7 The GDSC2 cell line panel

2.7.1 GDSC2 cell line details and data access

GDSC2 data were downloaded from the publicly available resource: `https://www.cancerrxgene.org/downloads/bulk_download` (accessed 30/09/22). 1001 cell lines were included in the Genomics of Drug Sensitivity in Cancer 2 (GDSC2) panel, derived from over 50 different tissue types. Of

these, 968 lines had associated gene expression data and 969 had drug sensitivity data for up to 295 compounds. Sensitivity was measured using cell viability assays based on metabolic activity to generate half-maximal inhibitory concentration (IC_{50}) values, for each combination of cell line and compound. Additional details for calculation of IC_{50} values are available in the accompanying documentation hosted at the GDSC2 site. HKMTi-1-005 was included in the GDSC2 screen under the DRUG_ID 2010, with sensitivity data recorded for 741 cell lines.

Gene expression data in the form of Robust Multi-Array Average (RMA)-normalised reads, amplification data and mutation data were downloaded from the COSMIC Cell Lines Project v96, release 31/05/22 [407]. Growth rate measured as doubling time was downloaded from the Cell Model Passports hub hosted by The Sanger Institute (`https://cellmodelpassports.sanger.ac.uk/downloads`, 07/09/22 release). Amplification is defined by COSMIC as average genome ploidy <= 2.7 and total copy number >= 5, or average genome ploidy > 2.7 and total copy number >= 9 [407]. To determine overexpression (OE), z-scores were calculated using the scale function in R. OE was defined as a z-score >1. Plots were created using ggplot2 [408].

2.8 Patient cohorts

2.8.1 Patient cohort details

International Cancer Genome Consortium (ICGC)

International Cancer Genome Consortium (ICGC) data were downloaded from `https://dcc.icgc.org/releases/current/Projects/OV-AU` (ICGC-OV), accessed 20/06/2023. Expression data were in the form of Fragments Per Kilobase of transcript per Million mapped reads (FPKM)-normalised reads (exp_seq.OV-AU.tsv file) and accompanying clinical information was found in the donor.OV-AU.tsv.gz file at the same site. Copy number data were not available at this site; although consensus calls for gene level data were available elsewhere (`https://dcc.icgc.org/releases/PCAWG/consensus_cnv/`), only 33 samples had copy number data for *MYC*. Of these, only 7 samples were *MYC*-amplified according to the COSMIC criteria described in section 2.7.1, hence ICGC data were not stratified by *MYC* status for downstream analyses. FPKM data were filtered to include only those counts with >0.5 FPKM in at least half the samples ahead of all downstream analyses.

The Cancer Genome Atlas (TCGA)

For ovarian, breast and LuAd cohorts, TCGA RNA data were downloaded using the GDCquery function from within the R package TCGABioLinks [409, 410], accessed 03/05/2023. Counts were downloaded as unstranded transcripts per million (TPM) data. Clinical data were accessed via

cBioPortal [121, 123], including the *MYC* status of patients. TPM data were filtered to include only those counts with >0.5 TPM in at least half the samples ahead of all downstream analyses. Where multiple samples were obtained from a single patient, a mean of gene expression across these samples was taken. Before aggregation across multiple samples, there were n=429 ovarian samples, n=1232 breast samples and n=600 LuAd samples. After aggregation there were n=422, n=1095 and n=517 samples respectively.

2.8.2 Patient cohort analyses

Immune inference

The CIBERSORTx "impute cell fractions" module was used to infer absolute immune infiltration (`https://cibersortx.stanford.edu/`). CIBERSORTx was run in absolute mode with 100 permutations, batch correction and with quantile normalisation disabled. The LM22 signature matrix file was used alongside the LM22 Source GEP. For calculation of relative immune infiltration, the Consensus^{TME} (cTME) package in R was used [411]. Count data were converted into a matrix and cTME was run with the statMethod = "ssgsea" parameter, with cancer = "OV", "BRCA" or "LUAD" as appropriate.

Survival analyses

The ICGC-OV cohort was used for an overview of association between 7ISG scores and survival, independent of *MYC* status. Overall survival (OS) in this cohort was defined as time of survival from initial diagnosis in days (donor_survival_time). ICGC-OV samples are a cohort of primary pre-treatment tumours and were not filtered: all patients were grade 2 or 3 [319], Fédération Internationale de Gynécologie et d'Obstétrique (FIGO) stage III/IV, with no data available regarding the radiotherapy/neoadjuvant chemotherapy status of patients, nor the extent of residual disease after debulking. Age was stratified at 50 years old, as used elsewhere for HGSC cohorts [412].

The TCGA-OV cohort was used for assessment of 7ISG associations with survival metrics in *MYC*-amplified versus non-amplified samples. Disease-specific survival (DSS) was defined by TCGA as the date of initial diagnosis to date of death specifically from disease, whilst OS was the date of initial diagnosis to date of death from any cause. DSS was used as the primary survival metric going forward. TCGA-BRCA (n=68 samples with DSS, with only 6 events) and TCGA-LUAD cohorts (n=29 samples with DSS, with only 8 events) were excluded due to the low number of events in each cohort preventing fitting of a CoxPH model. TCGA cohorts include radiotherapy status and neoadjuvant chemotherapy status; TCGA-OV data were filtered to include only grade 2 or 3 patients who had not received radiotherapy or neoadjuvant chemotherapy and with annotated FIGO stage (all

stage IIC-IV), resulting in 109 samples with *MYC* amplification and 212 without for use in survival analyses. Debulking status was not included, as a substantial proportion of patients were missing these data (n=43) and the vast majority of patients had residual disease (n=339 versus n=116 with no residual disease). Age was stratified at 50 years old, as used elsewhere for HGSC samples [412].

Before fitting a Cox proportional hazards (CoxPH) regression model, proportional hazards assumptions were tested for each dataset using the cox.zph function from the survival package and plotted using the ggcoxzph function [413]. Variance inflation factors (VIFs) were tested using the vif function from the car package to assess multicollinearity of predictor variables [414]. Survival objects were created using the Surv function and a CoxPH model was fitted using the coxph function, both from the survival package [413]. Within the CoxPH model, the 7ISG was included as a continuous variable to reduce the chance of false positives and increase statistical power. For visualisation purposes, 7ISG scores were split at the median value within each cohort to plot Kaplain-Meier curves. Curves were plotted using the ggsurvplot function from the survminer package [415].

2.9 Genesets and scoring

2.9.1 Genesets

The principal genesets used in scoring analyses are described in table 2.6. Genesets were downloaded from MSigDB [416] except where stated. Equivalent mouse genesets were downloaded from the mouse MSigDB collection where possible. Where no equivalent mouse geneset was available, the AnnotationDbi mapIDs function and [381] the biomaRt useMart and getLDS functions [417] were used to map human genes to orthologous mouse genes.

2.9.2 Singscore

The singscore package was used for geneset scoring. Singscore is a rank-based single sample scoring approach that mitigates the effect of differing sample sizes and of sample composition bias [418]. This package is described in detail in the vignette (`https://bioconductor.org/packages/release/bioc/html/singscore.html`). The following parameters were used with the singscore multiScore function to score all samples with scores centred around 0, using the mean absolute dispersion ("mad") function and allowing scoring of genesets of known directionality or mixed up- and down-regulated genes: subSamples = NULL, centerScore = TRUE, dispersionFun = mad, knownDirection = TRUE.

Geneset	Description	Source	Number of genes in geneset
(GOBP_)INNATE_IMMUNE_RESPONSE	Innate immune responses are defense responses mediated by germline encoded components that directly recognize components of potential pathogens. [GO_REF:0000022, GOC:add, GOC:ebc, GOC:mtg_sensu]	MSigDB v2023.1.Hs	919
7ISG	A reduced signature of immune genes strongly and consistently upregulated by HKMTi-1-005 treatment in HGSC cell lines (see appendix figure A.1 for full criteria for gene selection).	See 4.2.1 and appendix figure A.1.	7
DANG_MYC_TARGETS_DN	Genes down-regulated by MYC [GeneID=4609] and whose promoters are bound by MYC, according to MYC Target Gene Database.	MSigDB v2023.1.Hs	31
DANG_MYC_TARGETS_UP	Genes up-regulated by MYC [GeneID=4609] and whose promoters are bound by MYC, according to MYC Target Gene Database.	MSigDB v2023.1.Hs	130
DANG_REGULATED_BY_MYC_DN	Genes down-regulated by MYC [GeneID=4609], according to the MYC Target Gene Database.	MSigDB v2023.1.Hs	251
DANG_REGULATED_BY_MYC_UP	Genes up-regulated by MYC [GeneID=4609], according to the MYC Target Gene Database.	MSigDB v2023.1.Hs	68
Dual_EZH2G9a_targets	Defined by Casciello et al. in breast, melanoma and ovarian cancer cell lines, genes that are differentially expressed (in either direction) following simultaneous inhibition of EZH2 and G9a, but not by single inhibition of either enzyme individually.	Casciello et al., Cancer Res. 2022 Apr 1;82(7):1208-1221. doi: 10.1158/0008-5472.CAN-21-2218. PMID: 35149587.	597
EZH2_REPRESSION_TARGETS	Derived from a meta-analysis of 18 independent microarray studies performed by Curry et al., consensus genes that are upregulated by EZH2 knockdown in a variety of cell lines.	Curry et al., Clin. Epigenetics. 2015 Aug 21;7(1):84. doi: 10.1186/s13148-015-0118-9. PMID: 26300989; PMCID: PMC4545913.	300
GOBP_HISTONE_H3_K27_METHYLATION*	The modification of histone H3 by addition of one or more methyl groups to lysine at position 27 of the histone.	MSigDB v2022.1.Hs	18
GOBP_HISTONE_H3_K27_TRIMETHYLATION*	The modification of histone H3 by addition of three methyl groups to lysine at position 27 of the histone.	MSigDB v2022.1.Hs	10
GOBP_HISTONE_H3_K9_DIMETHYLATION*	The modification of histone H3 by addition of two methyl groups to lysine at position 9 of the histone.	MSigDB v2022.1.Hs	7
GOBP_HISTONE_H3_K9_METHYLATION*	The modification of histone H3 by addition of one or more methyl groups to lysine at position 9 of the histone.	MSigDB v2022.1.Hs	35
GOBP_HISTONE_H3_K9_TRIMETHYLATION*	The modification of histone H3 by addition of three methyl groups to lysine at position 9 of the histone.	MSigDB v2022.1.Hs	10
HALLMARK_INTERFERON_ALPHA_RESPONSE	Genes up-regulated in response to alpha interferon proteins.	MSigDB v2023.1.Hs	97
HALLMARK_INTERFERON_GAMMA_RESPONSE	Genes up-regulated in response to IFNG [GeneID=3458].	MSigDB v2023.1.Hs	200
HALLMARK_MYC_TARGETS_V1	A subgroup of genes regulated by MYC - version 1 (v1).	MSigDB v2022.1.Hs	200
LU_EZH2_TARGETS_DN	Genes down-regulated in SKOV3ip1 cells (ovarian cancer) upon knockdown of EZH2 [GeneID=2146] by RNA interference (RNAi).	MSigDB v2022.1.Hs	400
LU_EZH2_TARGETS_UP	Genes up-regulated in SKOV3ip1 cells (ovarian cancer) upon knockdown of EZH2 [GeneID=2146] by RNAi.	MSigDB v2022.1.Hs	289
MYC_G9A_REPRESSION_TARGETS	Defined by Tu et al., genes that had decreased cMYC binding accoding to ChIPseq whilst also being upregulated at the gene expression level by RNAseq analysis, in MCF10A.PM and MDA-MB-231 cells (breast cancer) treated with short hairpin targeting G9a.	Tu et al., Cancer Cell. 2018 Oct 8;34(4):579-595.e8. doi: 10.1016/j.ccell.2018.09.001. PMID: 30300580.	38
NUYTTEN_EZH2_TARGETS_DN	Genes down-regulated in PC3 cells (prostate cancer) after knockdown of EZH2 [GeneID=2146] by RNAi.	MSigDB v2022.1.Hs	1022
NUYTTEN_EZH2_TARGETS_UP	Genes up-regulated in PC3 cells (prostate cancer) after knockdown of EZH2 [GeneID=2146] by RNAi.	MSigDB v2022.1.Hs	1033
PRC2_EZH2_UP.V1_DN	Genes down-regulated in TIG3 cells (fibroblasts) upon knockdown of EZH2 [GeneID=2146] gene.	MSigDB v2023.1.Hs	192
PRC2_EZH2_UP.V1_UP	Genes up-regulated in TIG3 cells (fibroblasts) upon knockdown of EZH2 [GeneID=2146] gene.	MSigDB v2023.1.Hs	192
YU_MYC_TARGETS_DN	Genes down-regulated in B cell lymphoma tumors expressing an activated form of MYC [GeneID=4609].	MSigDB v2023.1.Hs	60
YU_MYC_TARGETS_UP	Genes up-regulated in B cell lymphoma tumors expressing an activated form of MYC [GeneID=4609].	MSigDB v2023.1.Hs	43

Table 2.6 **List of genesets.** List of main genesets used in scoring analyses. Genesets marked with an asterisk have been deprecated since the time of use.

2.10 Statistical analyses

Statistical analyses were mostly performed in the R environment (R version 4.3.1 (2023-06-16)) [366] within RStudio [367], excepting the stated subset performed in GraphPad PRISM v10.

Correlation analyses utilised Pearson's analysis where data were parametric, or Spearman's analysis where data were non-parametric. Tests for normal distribution of data were performed using D'Agostino & Pearson test, Anderson-Darling test, Shapiro-Wilk test or Kolmogorov-Smirnov test in GraphPad PRISM (if >1 test failed, data were considered non-parametric). GraphPad PRISM was used to generate data for correlation tiles and correlation bar plots, ahead of plotting with ggplot2 in RStudio [367, 408]. X-Y correlation plots were generated in RStudio, with correlation strength and significance calculated using the ggpubr stat_cor function [415]. Groupwise comparisons were performed using t-tests or Mann-Whitney (also known as Wilcoxon) tests, for parametric or non-parametric data respectively, using ggpubr functions [415].

The SuperExactTest package [419] was used for global comparison of expression changes (RNAseq) versus accessibility changes (ATACseq DAC method and ATACseq subtract method) of genes. SuperExactTest was also used for comparison of results from the two differing ATACseq approaches (DAC and subtraction), for comparison of upregulated TEs across cell lines (both at the subfamily and locus-specific levels) and between upregulated TEs and more accessible TEs.

Multiple linear regression was performed using the lm function from the stats package in R [366], with added-variable plots visualised using the avPlots function. Statistical analyses of 7ISG relationships with survival are described separately in section 2.8.2.

2.11 Collaborative projects

2.11.1 CRUK Scotland Institute: lung adenocarcinoma experiments

All lung adenocarcinoma experiments were performed at the CRUK Scotland Institute (previously known as the CRUK Beatson Institute) by Dr Sarah Laing and Professor Daniel J Murphy.

KM^2 LuAd mouse model and *in vivo* procedures

Mouse experiments were performed in accordance with Home Office Licence number PE47BC0BF (CRUK BICR, UK). The *LSL-KRas*^{G12D} (RRID:IMSR_JAX:008179) and*Rosa26*$^{DM.lsl-MYC}$ (RRID :IMSR_JAX:033805) allelic mice have been described elsewhere [420, 421]. Mice were maintained on a mixed FVBN/C57Bl6 background, housed on a 12hr light/dark cycle, and fed and watered ad libitum. Recombinant Adenovirus expressing CRE recombinase (Ad5mSPC-CRE) was purchased from the University of Iowa gene therapy facility. For Ad5mSPC-CRE instillation, sedation of young

adult mice (8- to 10-week-old) was achieved via intraperitoneal injection of a mixture of medetomidine and ketamine. The calcium phosphate precipitation method, described previously [422], was used for intranasal administration of $1x10^8$ pfu Ad5mSPC-CRE for all experiments. Mice were randomly assigned to treatment groups, balanced for sex, with both males and females used in all studies. To assess HKMTi-1-005 effect on tumour characteristics, the treatment schedule was 3 days of treatment followed by four days of recovery each week, with HKMTi-1-005 at 40mg/kg or vehicle control (0.9% NaCl) administered twice daily by intraperitoneal injection, from 4 weeks post-induction to 8 weeks post-induction. To assess the potential survival benefit of treatment, HKMTi-1-005 (35mg/kg) or vehicle control (0.9% NaCl) was administered twice daily by intraperitoneal injection using the same treatment schedule. Mice were monitored by facility staff without knowledge of experimental parameters, and were humanely culled when clinical endpoint was reached.

KM^2 LuAd mouse model: immunohistochemistry and tissue analysis

4µm, formalin-fixed, paraffin-embedded sections were used for immunohistochemical analysis, following heating to 60°C for 2 hours. Peroxidase blocking was performed over 10 minutes in 1% H_2O_2 diluted in H_2O. Subsequently, enzyme-mediated antigen retrieval was performed. 1% BSA was used for 1h at room temperature to block non-specific antibody binding. Ki67 (Cell Signalling 12202) was used at a dilution of 1:1000, followed by Rabbit Envision (Aligent) as a secondary antibody. The horseradish peroxidase (HRP) signal was detected using liquid DAB (Aligent and Invitrogen). Sections were counterstained with hematoxylin and cover-clipped using DPX mount (CellPath). HALO Software (Indica Labs) was used to quantify tumour burden and Ki67+ cells.

Lung cancer cell lines and treatment

The A459 and H358 cell lines were obtained from ATCC and were maintained in RPMI, supplemented with 10% foetal bovine serum (FBS) and penicillin-streptomycin. The effect of HKMTi-1-005 on cell viability was measured using dose-response curves from Incucyte analysis. 10µM HKMTi-1-005, 100nM Trametinib, or H_2O (control) was used to treat cell lines for 24h ahead of cell harvest for gene expression analysis. Total RNA was extracted using Qiagen RNeasy kit as per the manufacturer's protocol. DNA depletion was performed using the RNase-Free DNase Set (Qiagen). Total RNA was eluted in 30µl of RNase-Free H_2O, followed by measurement of the concentration of RNA using NanoDrop 2000c (Thermo Fisher). cDNA was generated using the QuantiTect Reverse Transcription kit (Qiagen), with 2µg RNA in each reaction. Expression of a set of ISGs (*IRF7, IRF9, STAT2*) as well as the housekeeping comparison *GUSB* was quantified by qRT-PCR using SYBR Green method (VWR QUNT95072). Primers used are detailed in table 2.7. Expression of ISGs was normalised to expression of *GUSB*.

Gene	Forward Primer	Reverse Primer
GUSB	cctgtgacctttgtgagcaa	aacagatcacatccacatacgg
IRF7	ctgcagtcacacctgtagcc	gtggactgagggcttgtagc
IRF9	tcctccagagccagactact	caatccaggctttgcacctg
STAT2	accattctggacatggctgg	ctccgactcacaaagcccat

Table 2.7 **Primers used in qRT-PCR of lung cancer cell lines.**

2.11.2 AstraZeneca: biochemical investigations into HKMTi-1-005

The series of biochemical experiments assessing HKMTi-1-005 activity were carried out by AstraZeneca (AZ). This included a fluorescence resonance energy transfer (FRET) assay in G401 cells to assess H3K27me3 levels based on proximity of probes that bind H3 and K27me3 (utilising the Cisbio kit: `https://uk.cisbio.eu/epigeneous-h3k27me3-cellular-kit-40453`). In brief, cells were seeded on plates and left for 48h, followed by lysis and addition of an HTRF antibody conjugate. Subsequently, overnight incubation was performed and a Pherastar reader was used to measure fluorescence output. Cell viability assays were also performed by AZ, using the Promega Cell Titre Glo luminescence cell viability assay. Cell-free assays to determine PRC2 inhibitory effect of HKMTi-1-005 utilised the MTase-Glo™methyltransferase assay from Promega, with a 5-component complex (EZH2, EED, SUZ12, RBBP4 (RbAp48) and AEBP2 subunits) (figure 2.2). Varying concentrations of peptide and nucleosome substrates were used to assess substrate-competitive inhibition at different concentrations of HKMTi-1-005 to generate PRC2 inhibition curves [423].

2.12 Graphics

All graphics were created with BioRender.com.

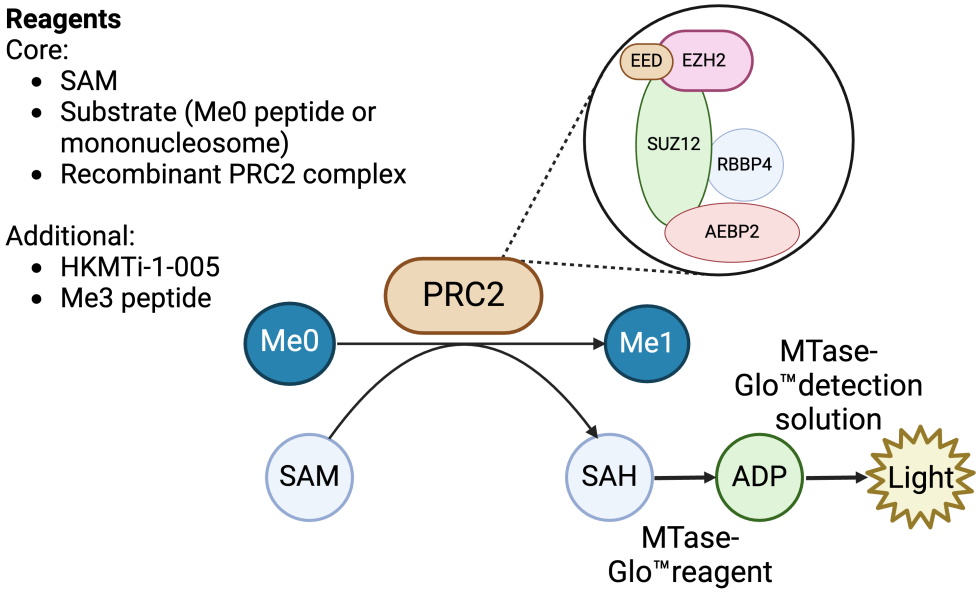


Figure 2.2 **Schematic describing the cell-free PRC2 inhibition assay performed by AZ.** The MTase-Glo™methyltransferase assay measures the accumulation of S-adenosyl homocysteine (SAH). SAH is the product of the methyltransferase reaction catalysed by the PRC2 complex, which uses SAM as a substrate. SAH concentration correlates with the luminescence signal produced following a luciferase reaction, which can be used to determine PRC2 activity in different contexts (e.g. varying peptide concentrations or in the presence of different compounds such as HKMTi-1-005), via comparison to a SAH standard curve. The PRC2 complex used is a recombinant 5-component complex comprising EZH2, EED, SUZ12, RBBP4 (RbAp48) and AEBP2 subunits. Figure based on schematic available at `https://www.promega.co.uk/products/epigenetics/methylation-analysis/mtase_glo-methyltransferase-assay/?catNum=V7601`, created with BioRender.com.

Chapter 3

Results Section 1: HKMTi-1-005 has EZH2i- and G9ai-like properties

3.1 Introduction

As detailed in section 1.4.3, EZH2 and G9a demonstrate a complex interplay and redundancy in their histone methyltransferase activities. Single EZH2/G9a inhibitors have proven insufficient for induction of certain target genes and repetitive elements [109, 153]. The use of two single EZH2 and G9a inhibitors simultaneously may overcome some of the EZH2-G9a interplay described, but this approach is limited by the additional complication in the clinical and pharmacodynamic evaluation of two single inhibitors, as compared to a single compound with dual activity [364]. HKMTi-1-005 – based on the template of the single G9ai BIX-01294 – was synthesised to address this issue [424].

HKMTi-1-005 was chosen from an initial screen of over 200 compounds [109]. The choice of HKMTi-1-005 over other hits was based on its cell growth inhibition, its ability to reduce H3K9/H3K27 methylation, and its specificity for the EZH2 and G9a enzymes over other methyltransferases [109]. Additionally, the pharmacological and pharmacokinetic properties of HKMTi-1-005 were superior to selected other hits tested from the screen, including absorption, distribution, metabolism and excretion (ADME) properties (unpublished data, in collaboration with the ICR). Tolerability of HKMTi-1-005 treatment at twice daily doses of 40mg/kg over 5 days has been demonstrated in ovarian cancer xenograft models, as has significant tumour growth inhibition in these models (appendix figure A.2). Treatment with HKMTi-1-005 also reduced tumour burden in mice bearing *Trp53*$^{-/-}$ ID8 ovarian tumours, whilst providing a survival benefit [155].

EZH2i- and G9ai-like activity of HKMTi-1-005 has been evidenced [109]. Its effects and mechanism of action were demonstrated to be distinct from currently available single inhibitor compounds, with observation of direct enzymatic inhibition of both G9a and EZH2 [109]. However, most of these initial analyses were performed in a single breast cancer cell line, MDA-MB-231. Additionally,

cell-free assays utilised a minimal 3-component PRC2 complex to demonstrate EZH2i effects, which cannot accurately recapitulate the cellular environment.

To advance HKMTi-1-005 and to progress understanding of dual EZH2/G9a targeting as a treatment modality, several additional pre-clinical investigations are required. I first looked to consolidate EZH2i- and G9ai-like effects of HKMTi-1-005 in other cancer types *in vitro*, alongside validation of these effects in *in vivo* models. To this end, the effect of HKMTi-1-005 treatment on a panel of ovarian high-grade serous carcinoma (HGSC) cell lines was assessed. HGSC was chosen as an appropriate model due to the importance of both epigenetic and the immune factors in patient outcome (sections 1.9.2 and 1.9.4). Gene expression and chromatin accessibility were investigated using RNAseq and ATACseq respectively, with a focus on changes at targets of EZH2 and G9a repression. Additional ATACseq data derived from HKMTi-1-005-treated mouse cell line models of HGSC were analysed to corroborate chromatin accessibility effects. In parallel to these *in vitro* investigations, I assessed evidence for G9a and EZH2 inhibitory effects of HKMTi-1-005 in an *in vivo* setting. This involved re-analysis of prior RNAseq and ATACseq data derived from a *Trp53*$^{-/-}$ ID8 mouse model of HGSC treated with HKMTi-1-005 [155]. Alongside examining specific EZH2 and G9a targets in these HGSC models, I looked to provide a more global understanding of which chromatin contexts and genomic elements were most affected by HKMTi-1-005 treatment.

Following characterisation of the HKMTi-1-005 treatment effect on HGSC models, I examined the sensitivity of different tumour types to HKMTi-1-005. For this analysis, the publicly available Genomics of Drug Sensitivity in Cancer 2 (GDSC2) cell line panel was utilised [425]. The GDSC2 contains sensitivity data of >1000 pan-cancer cell lines to over 250 compounds, one of which is HKMTi-1-005. These data were compared to sensitivity data for single inhibitors of EZH2 or G9a, alongside the DNA damage agent cisplatin. This analysis aimed to assess whether sensitivity profiles supported the EZH2 and G9a inhibitory activity of the compound and to examine its distinction from single inhibition approaches. Additionally, the most sensitive tumour types were identified to guide further investigation into drivers of HKMTi-1-005 sensitivity and potential targets for clinical efficacy.

Finally, I looked to validate the initial cell-free assay demonstrating EZH2i activity of HKMTi-1-005, given the caveats of the assay and the G9ai-based origin of the compound. To do so, I evaluated the results of a series of biochemical assays performed by AstraZeneca that investigated the direct EZH2i activity of HKMTi-1-005.

Taken together, this section of results aims to consolidate and characterise the disruption of H3K9 and H3K27 methylation maintenance by HKMTi-1-005 *in vitro* and *in vivo*, whilst providing an insight into tumour types that may be sensitive to dual EZH2/G9a targeting by HKMTi-1-005.

The aims of this chapter are to address the following questions:

- Do biochemical assays confirm EZH2i activity of the compound, given its derivation from a single G9ai?
- Do we see evidence for the dual activity of the compound against EZH2- and G9a-regulated targets in HGSC models, both *in vitro* and *in vivo*? Is this effect specific to EZH2/G9a targets over those of other epigenetic modifiers?
- Globally, what effect does HKMTi-1-005 have on gene expression and chromatin accessibility in HGSC models? Are there particular chromatin contexts and genomic elements that are responsive to HKMTi-1-005, that may be relevant to the anti-tumour treatment effect?
- Does sensitivity to the compound support EZH2i- and G9ai-like behaviour of the compound across pan-cancer cell lines? Which tumour types are most sensitive to HKMTi-1-005 treatment?

3.2 *In vitro* and *in vivo* transcriptome analysis: HKMTi-1-005 treatment effect on HGSC

3.2.1 RNAseq data processing: HGSC cell lines

The following six cell lines were included to assess HKMTi-1-005 treatment effect *in vitro*: PEO1, PEO4, PEA2, OVCAR5, OVCAR8 and IGROV1. In a series of experiments performed by Dr Ian Garner, cells were treated at their respective IC_{50} values for 72h, followed by RNA extraction with ribosomal RNA depletion. Subsequently, paired-end RNA sequencing was conducted at the ICR. Experiments were run in triplicate for PEO1 and PEO4 cell lines, with the remaining lines run in duplicate. On receipt of the raw data from this sequencing run, I performed downstream analyses to assess treatment effect.

I first performed a quality control (QC) analysis of each sample. FastQC [371] was used for an overview of data quality, including parameters such as Phred score (reflecting sequence quality), per-base N content (reflecting the number of low confidence base calls) and adapter contamination. Trimmomatic [373] was used to trim reads as an additional QC measure, aiming to remove any remaining adapters and improve per-base sequence content. After trimming, reads were aligned to the reference human GRCh38 genome (release 79) downloaded from the Ensembl database [375, 376], using STAR [374]. A repeat FastQC run was performed, followed by compilation of all quality control reports by MultiQC [372] (figures 3.1 and 3.2). The reports indicated a satisfactory output from read alignment for all samples except PEA2 treated sample 2 (PEA2 T2), which showed a low read count (figures 3.1d and 3.2). The data for this sample were retained to assess its suitability for use in further analyses, based on downstream quality parameters.

Subsequently, bam files were filtered using bedtools [377] to remove blacklisted regions following alignment. These regions consist of anomalous, unstructured, or high-signal regions, requiring removal due to their potential to bias the downstream results and lead to inaccurate quantification [426]. Following filtering, the featureCounts function [379] was used to count reads ahead of differential expression analyses with the DESeq2 package [380].

The results of differential expression analyses were visualised in RStudio [367] to determine the distance between samples as a final QC measure. PCA plots demonstrated a clear segregation of samples based on treatment status for PEO1, PEO4, PEA2, OVCAR5 and OVCAR8, suggesting a strong HKMTi-1-005 treatment effect (figures 3.3a and 3.3b). The high similarity between the PEA2 T1 and T2 samples and the segregation of treated from untreated samples allowed the T2 sample to be retained for downstream analyses, despite the low read count.

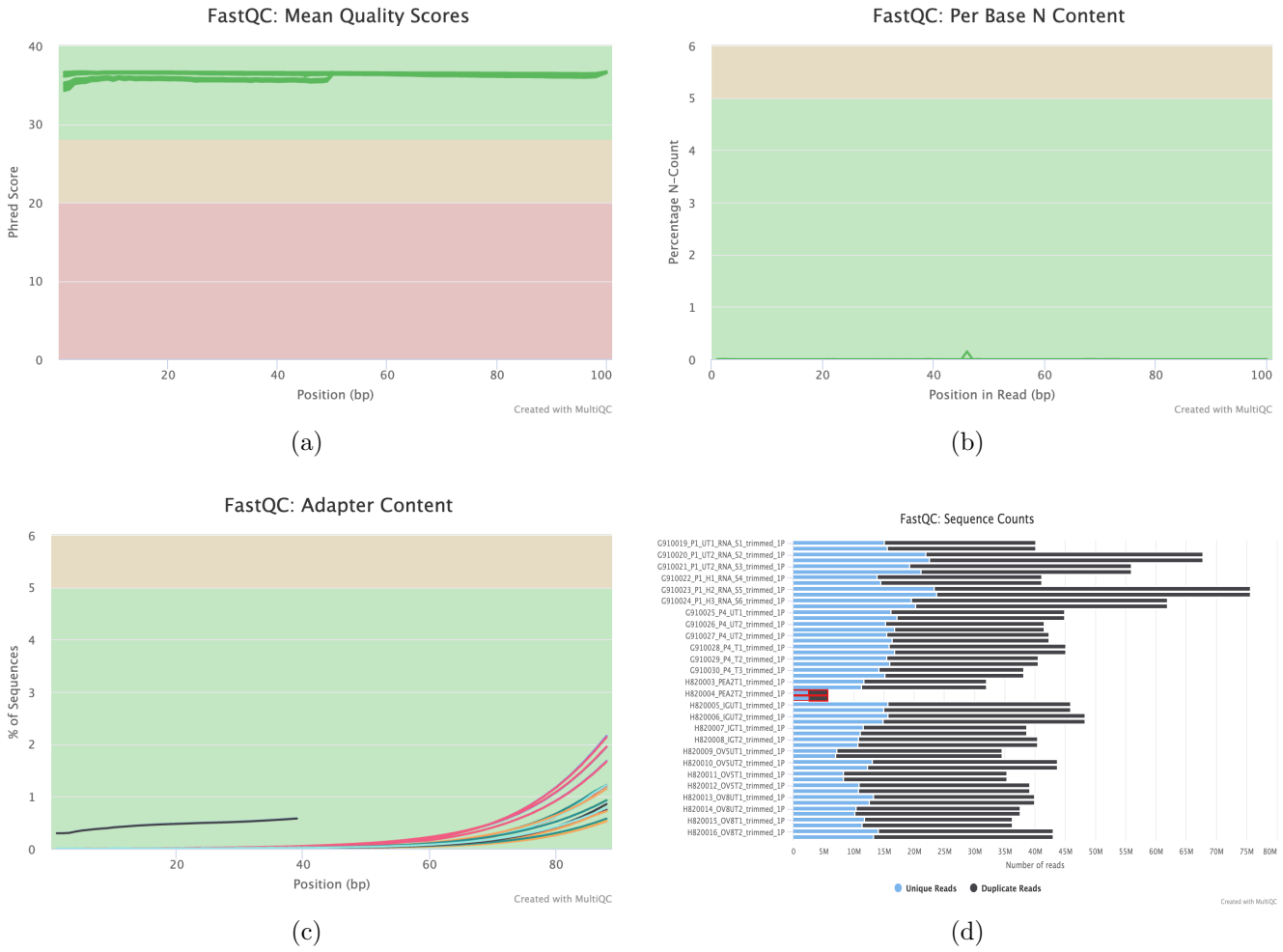


Figure 3.1 **QC for *in vitro* HGSC RNAseq data.** QC metrics for RNAseq data derived from PEO1, PEO4, PEA2, OVCAR5, OVCAR8 and IGROV1 HGSC cell lines, control and HKMTi-1-005-treated, after read trimming. (a) Phred scores, reflecting a measure of the accuracy of base calling. A Phred score of >30 reflects a 99.9% base call accuracy. (b) Per-base N content, a quality metric based on the percentage of bases with no base call at each read position. (c) Adapter contamination (cumulative) across the read library. (d) FastQC sequence count for each sample, with an estimate of the proportion of duplicate reads. Low read counts were observed for the second PEA2 treated replicate, indicated in red. P1 = PEO1 and P4 = PEO4, run in triplicate. P2 = PEA2, OV5 = OVCAR5, OV8 = OVCAR8, and IG = IGROV1, run in duplicate. Treated samples are labelled as H1/H2/H3 or T1/T2/T3, whilst untreated samples are labelled UT1/UT2/UT3. PEO1 and PEO4 reads were 50bp long, whilst PEA2, OVCAR5, OVCAR8, and IGROV1 were 100bp long. bp = base pairs.

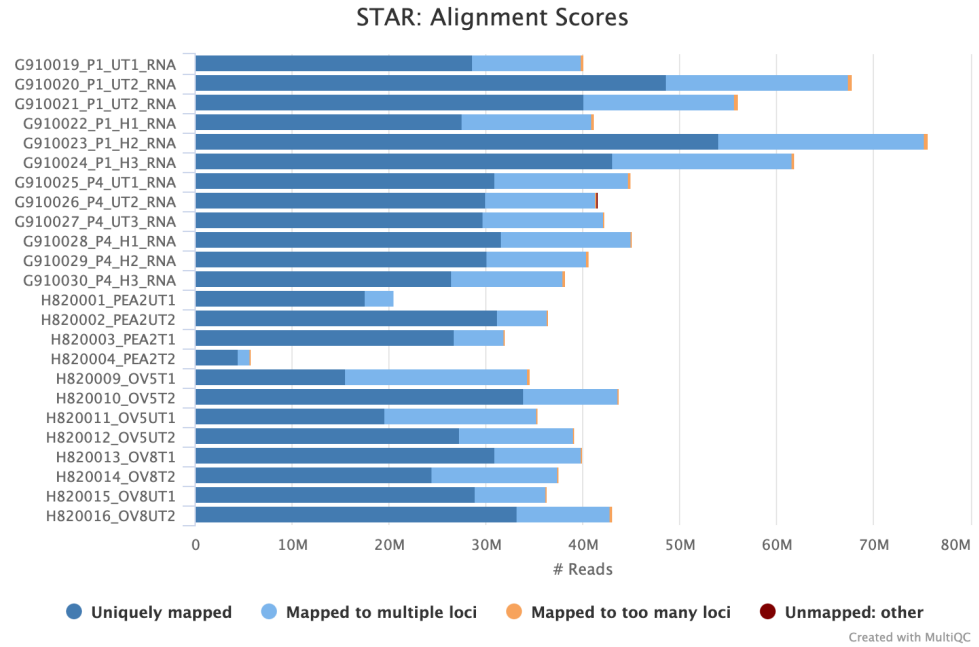
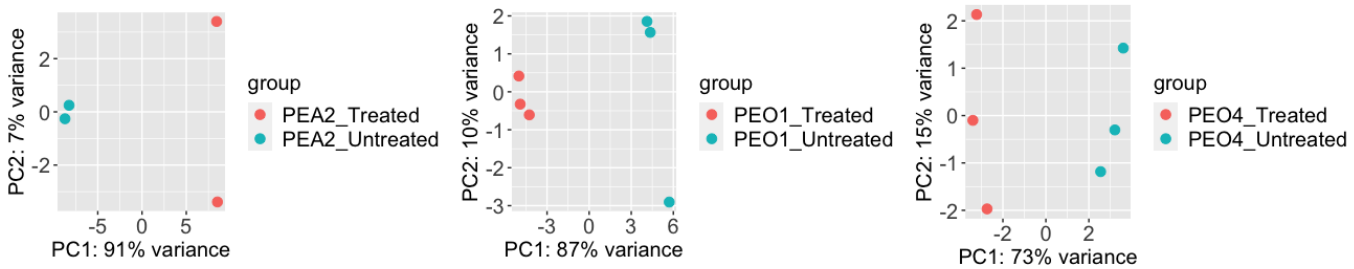


Figure 3.2 **Alignment metrics for *in vitro* HGSC RNAseq data.** Alignment metrics for RNAseq data from PEO1, PEO4, PEA2, OVCAR5, OVCAR8 and IGROV1 HGSC cell lines, control and HKMTi-1-005-treated. Reads were aligned to the reference human GRCh38 genome (release 79). P1 = PEO1 and P4 = PEO4, run in triplicate. P2 = PEA2, OV5 = OVCAR5, OV8 = OVCAR8 and IG = IGROV1, run in duplicate. Treated samples are labelled as H1/H2/H3 or T1/T2/T3, whilst untreated samples are labelled UT1/UT2/UT3.

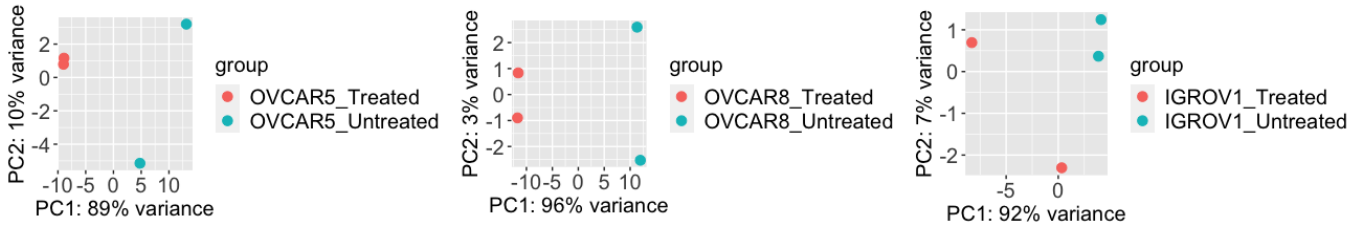
However, these final QC measures demonstrated a lack of segregation by treatment status in IGROV1 samples (figures 3.3b and 3.3c). Instead, the second IGROV1 treated replicate clustered with untreated samples. The IGROV1 cell line is now considered to be poor model of HGSC, with a "hypermutator" phenotype and a flat copy number profile; the cell line appears more similar to endometrioid and clear cell ovarian cancers, and in fact was originally derived from a mixed tumour type [370]. Given the nature of the cell line and the limited treatment effect on the second replicate, IGROV1 was excluded from further analyses. All other samples were retained.

3.2.2 RNAseq data processing: ID8 *in vivo* murine model of HGSC

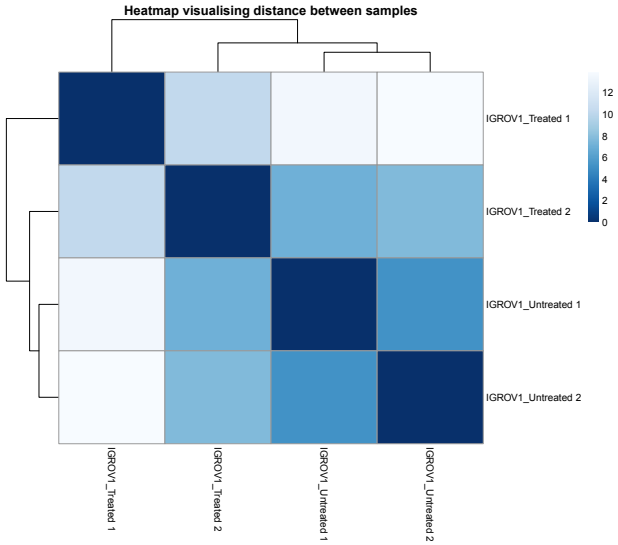
In addition to the cell line analysis, I performed a re-analysis of *in vivo* RNAseq data obtained from a prior experiment performed by our group [155]. In the initial experiment, RNAseq was performed on *Trp53*$^{-/-}$ ID8 tumours from mice treated with HKMTi-1-005 (20 mg/kg intraperitoneally twice a day for 14 days, n=7), or those treated with vehicle (1% Tween/3.4% DMSO in 0.9% NaCl intraperitoneally twice a day, n=7). Prior downstream processing had included alignment to the GRCm38 (mm10) genome version with a focused analysis on specific gene and pathway changes relating to immune signalling. Hence, I re-analysed the raw data obtained from ID8 tumours to include alignment to an updated mouse genome version (GRCm39 release 104 (mm39)) and a broader analysis of HKMTi-1-005 effects.



(a)



(b)



(c)

Figure 3.3 **QC following differential expression analyses, *in vitro* HGSC RNAseq data.** QC metrics reflecting similarity between samples after differential expression analysis, for RNAseq data derived from PEA2, PEO1, PEO4, OVCAR5, OVCAR8 and IGROV1 HGSC cell lines, control and HKMTi-1-005-treated. PCA plots of (a) PEA2, PEO1, PEO4, and (b) OVCAR5, OVCAR8, and IGROV1 replicates (left to right). PEO1 and PEO4 run in triplicate, all other samples run in duplicate. (b) Heatmap visualising the sample-to-sample distance between IGROV1 replicates, based on regularized logarithm (rlog)-transformed counts. Samples are clustered hierarchically based on sample-to-sample distances, giving an overview of similarity between replicates: darker blue colour indicates a smaller distance between samples.

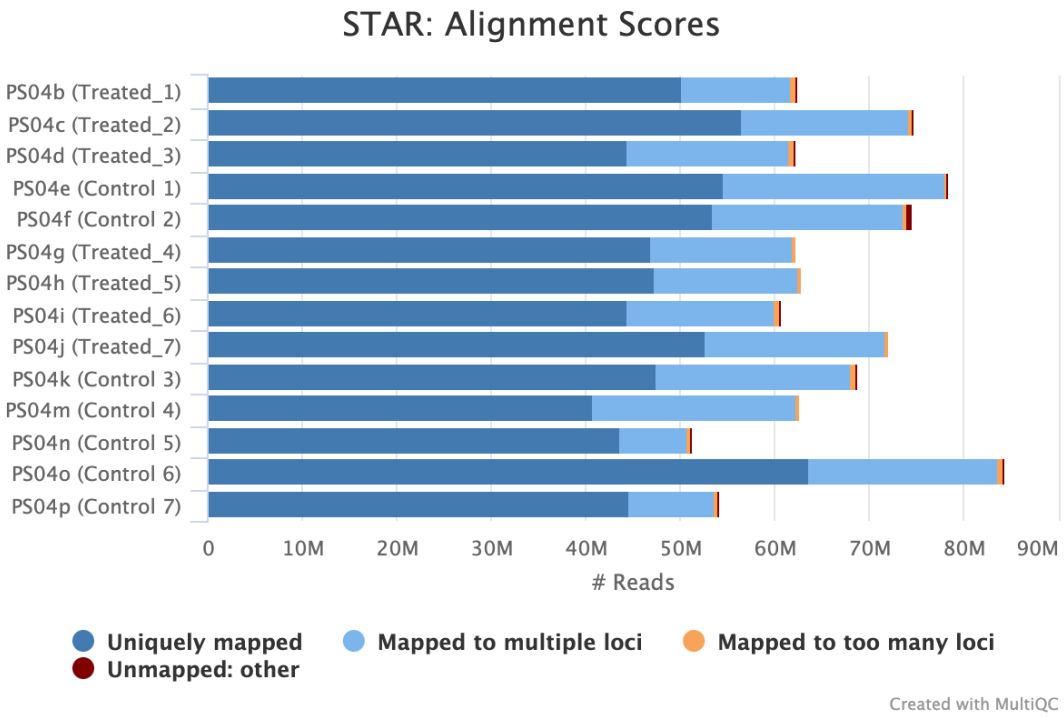
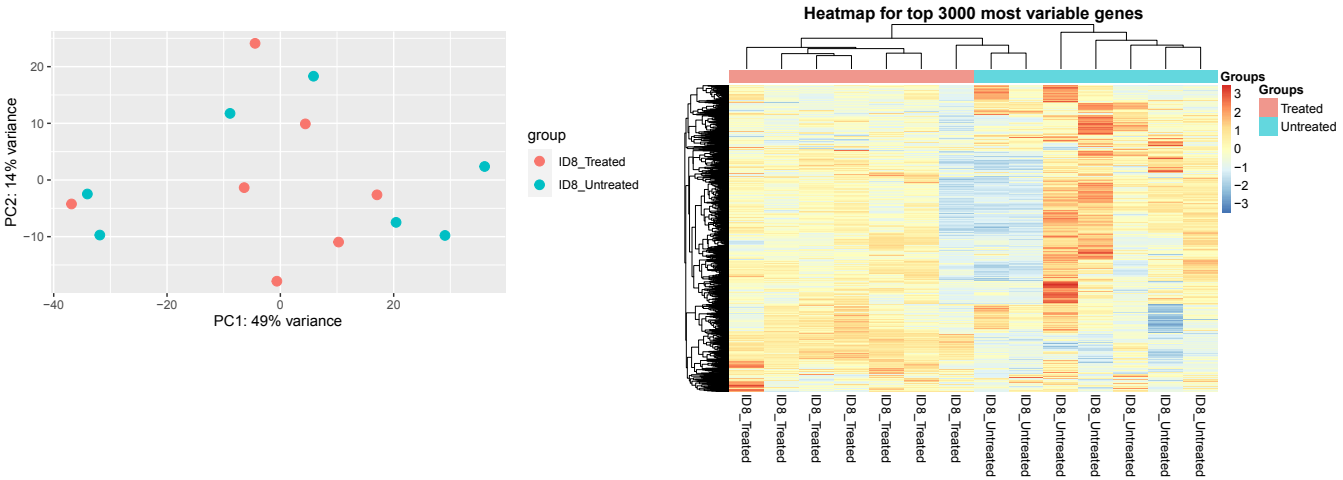


Figure 3.4 **Alignment metrics for ID8 *in vivo* RNAseq data.** Alignment metrics for RNAseq data derived from ID8 *in vivo* model, control and HKMTi-1-005-treated. Reads were aligned to the reference mouse GRCm39 genome (release 104) using STAR. N=7 treated (PS04b, PS04c, PS04d, PS04g, PS04h, PS04i, PS04j) and n=7 untreated (PS04e, PS04f, PS04k, PS04m, PS04n, PS04o, PS04p) replicates.

The pipeline for analysis of ID8 *in vivo* RNAseq data was matched as closely as possible to that described for the HGSC cell lines. Quality control using FastQC [371] had been performed previously and demonstrated satisfactory data quality, hence raw FASTQ files were aligned to the mm39 genome using STAR [374]. MultiQC [372] was used to collate quality metrics for aligned files; files demonstrated satisfactory read alignment, resulting in the retention of all samples for downstream analysis (figure 3.4). Blacklist regions were removed using bedtools [377, 426]. FeatureCounts [379] and DESeq2 [380] allowed read counting and differential expression respectively, followed by visualisation of the distance between samples in RStudio [367]. Treated and untreated samples segregated less clearly in this instance (figure 3.5a), an effect that may be expected in tumours as compared to cell lines given the bulk nature of sequencing. However, when narrowing analyses to the genes with the highest variability, samples did cluster largely by treatment status (figure 3.5b). Of note, differential expression analyses demonstrated a strong and significant overlap between the initial mm10 and updated mm39 analyses (appendix table A.3).



(a) (b)

Figure 3.5 **QC following differential expression analyses, ID8 *in vivo* RNAseq data.** QC metrics reflecting similarity between samples after differential expression analysis, for RNAseq data derived from the ID8 *in vivo* HGSC mouse model, control and HKMTi-1-005-treated. (a) PCA plot of ID8 *in vivo* replicates, for dimensionality reduction of data. (b) Heatmap demonstrating clustering of ID8 *in vivo* treated and untreated samples, based on regularised log (rlog)-transformed counts of the 3000 most variable genes. Gene counts scaled by row: red indicates higher count and blue indicates lower count.

3.2.3 Overview of HKMTi-1-005 transcriptional effects

To give a global view of treatment effect, volcano plots for each of the five retained HGSC cell lines and the ID8 tumours were generated (figure 3.6). As anticipated given the repressive effects of H3K27me and H3K9me, these demonstrated a greater number of up- versus down-regulated genes in each sample. A total of between 846 (PEO4) and 4473 (OVCAR8) differentially expressed genes (DEGs) were identified across samples, with DEGs defined as genes with an absolute log2 fold-change (log2FC) of > 0 and an adjusted p-value (p-adj) of < 0.05. PEO4 – derived from the same patient as PEO1, but after the emergence of cisplatin resistance [427] – demonstrated the fewest DEGs following treatment, suggesting a weaker response to HKMTi-1-005 than other lines.

Each sample was analysed using both over-representation analysis (ORA) and gene set enrichment analysis (GSEA). The database used to perform these analyses was the Gene Ontology database, specifically the Biological Process sub-ontology (GOBP) containing genes involved in biological processes that are annotated by the same ontology term. For GSEA, Hallmark pathways were also analysed; these consist of genes that are co-ordinately regulated, and represent "well-defined biological states or processes" [416].

ORA can be used to determine whether certain genesets are represented by DEGs more than would be expected by chance, providing a non-directional enrichment result. Using this approach via

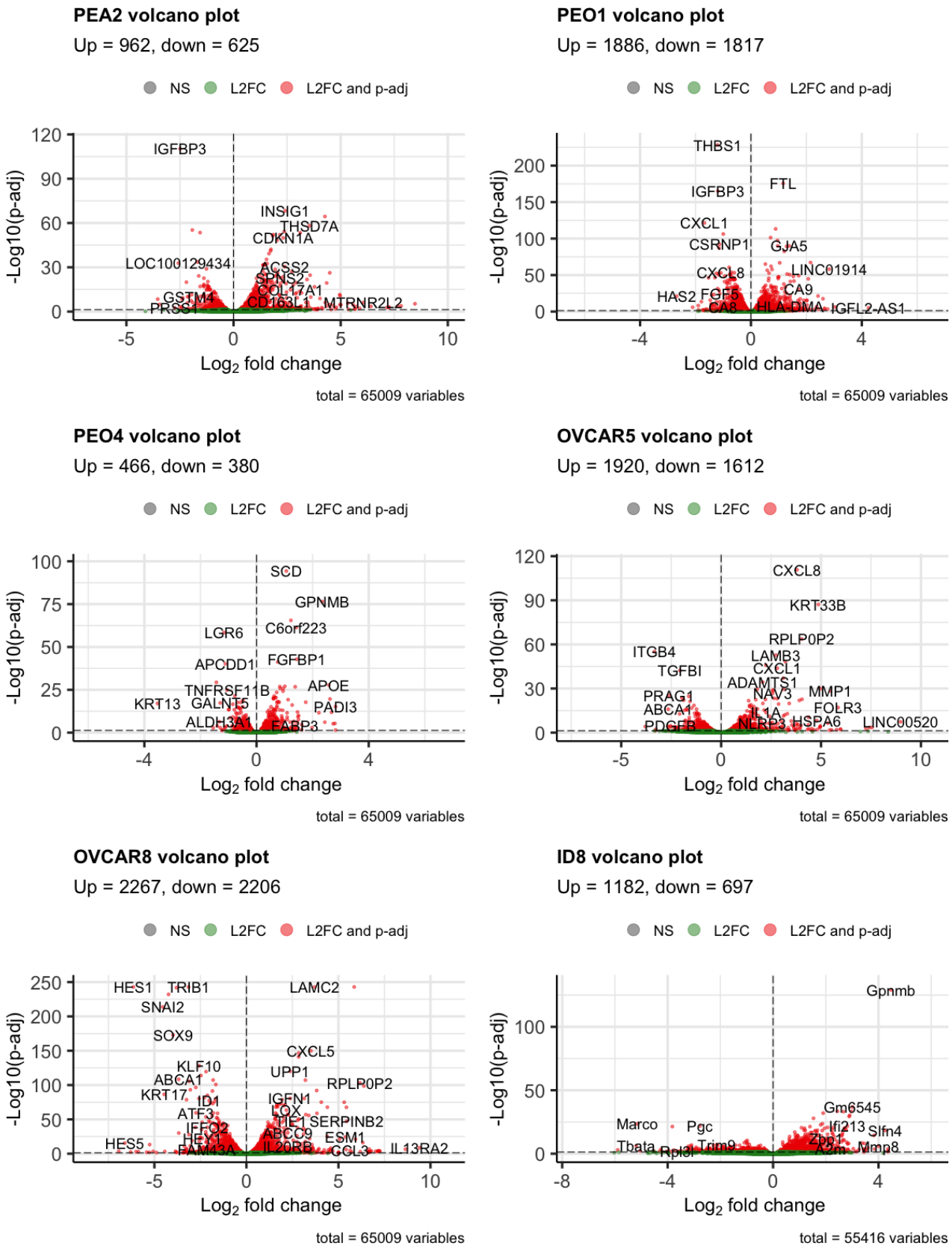
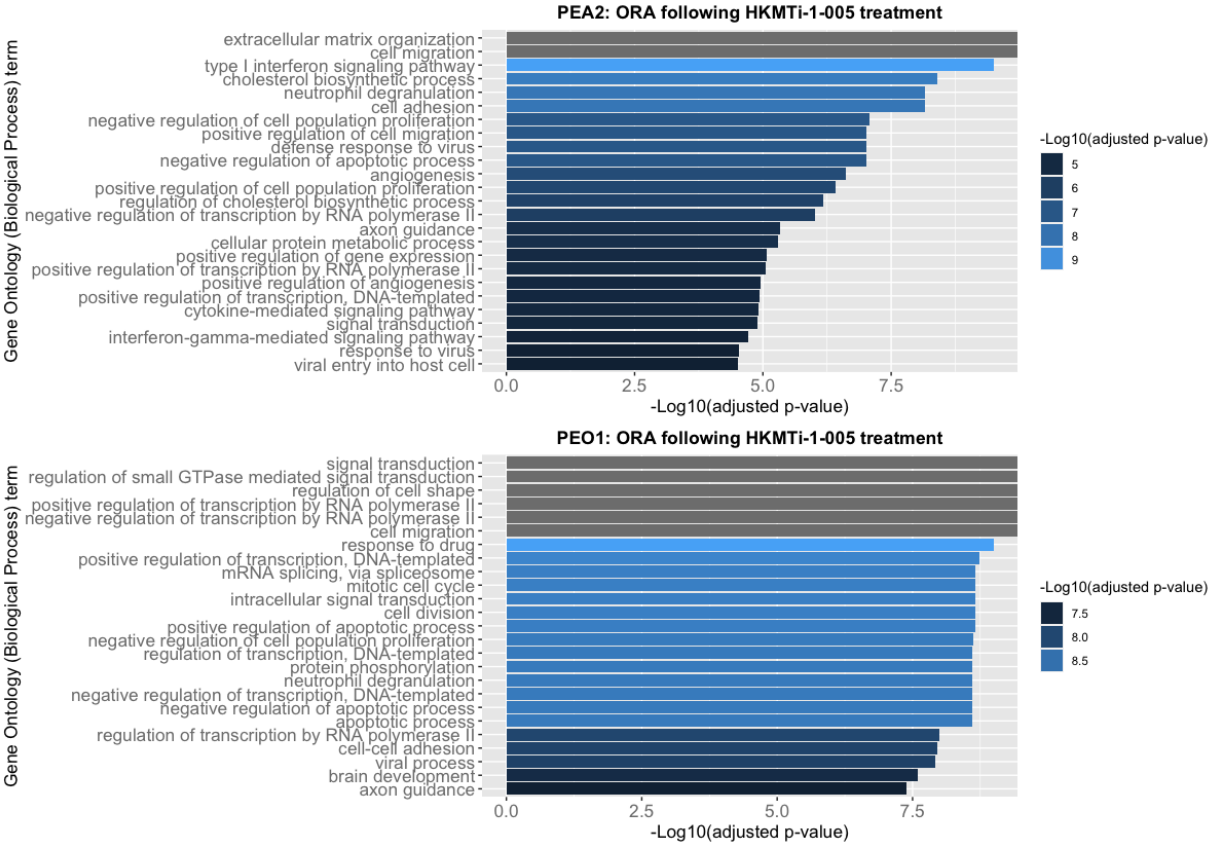


Figure 3.6 **RNAseq volcano plots, HGSC samples.** Volcano plot visualisation of differentially expressed genes in PEA2, PEO1, PEO4, OVCAR5, OVCAR8 cell lines and the ID8 *in vivo* sample (left to right, top to bottom), control vs HKMTi-1-005-treated. DEGs are coloured red (absolute log2FC > 0, p-adj < 0.05). Genes coloured in green = absolute log2FC > 0, p-adj > 0.05. For visualisation purposes, only genes with absolute log2FC > 2 and p-adj < 0.05 are labelled by name.

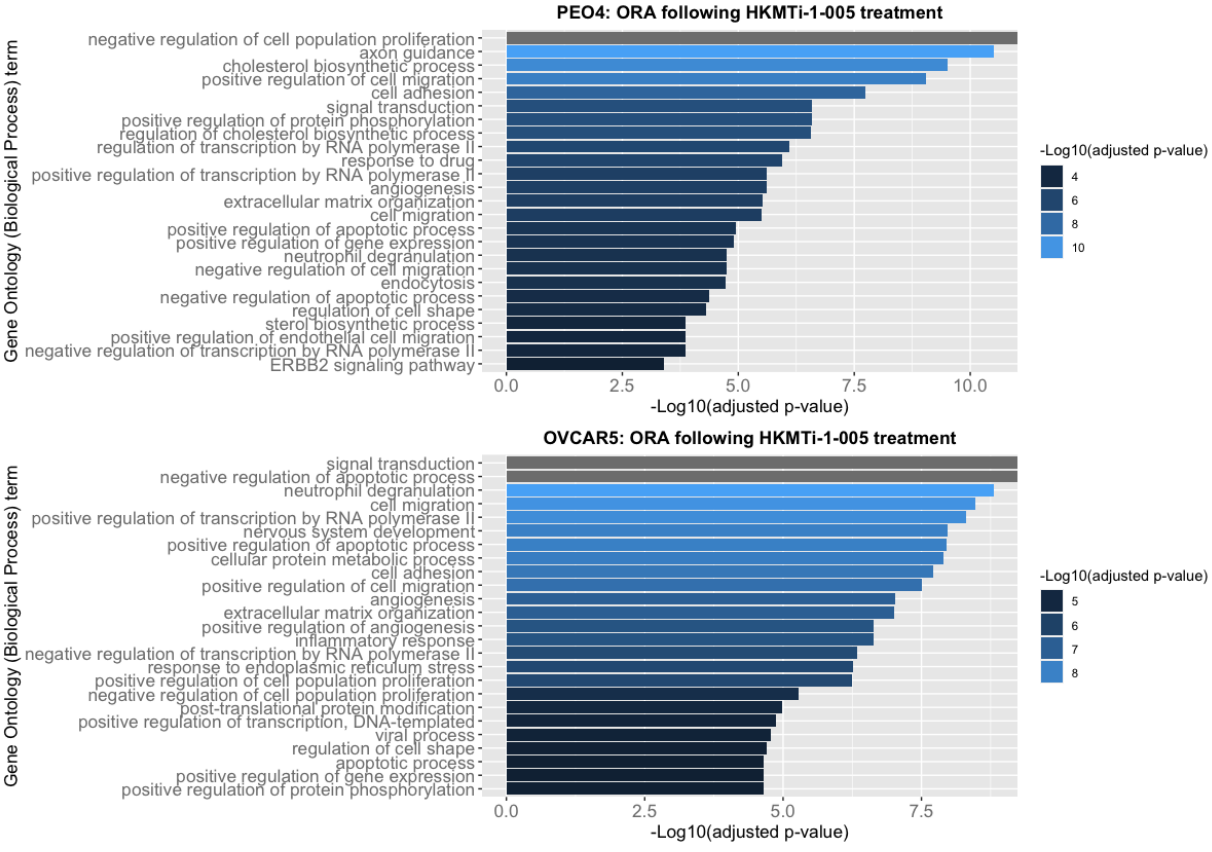
the geneSCF package [398], between 215 and 471 GOBP pathways were over-represented following HKMTi-1-005 treatment in the HGSC cell lines (p-adj < 0.05). Changes in pathways relating to transcription, cell adhesion/migration, apoptosis, immune responses and metabolism consistently appeared in the top 25 across samples (figure 3.7). ORA on the ID8 *in vivo* data was performed using the EnrichR platform [399–401], with 160 pathways enriched including changes in apoptotic, metabolic and immune pathways. However, the *in vivo* data were dominated by immune changes, representing almost all of the top 25 pathways (figure 3.7c). Several other pathways such as 'response to drug' and various signalling cascades were altered significantly across samples.

In contrast to ORA performed only on DEGs, GSEA considers all genes in the RNAseq assay and uses a ranking metric based on differential expression to give a directional enrichment result. This allows analysis of whether certain genesets are more up- or down-regulated than expected versus the background of transcriptomic changes observed. GSEA was performed using WebGestalt [402] for all samples. GSEA GOBP results were slightly less conserved across samples than observed by ORA, but suggested a general downregulation in cell adhesion/migration pathways and several chromatin organisation pathways, whilst immune responses and metabolic processes tended to be upregulated (appendix figure A.3). The ID8 *in vivo* model once again demonstrated a striking upregulation of numerous immune response genesets including adaptive immune pathways (appendix figure A.3).

Hallmark pathways were considered by GSEA alongside GOBP genesets (appendix figure A.4). These provided support for GOBP findings – for example, the upregulation of immune, apoptotic and metabolic genesets with downregulation of epithelial-mesenchymal transition genesets – yet additionally demonstrated strong downregulation in pathways relating to cMYC signalling and to cell cycle progression (appendix figure A.4). Of note, PEO4 was an outlier in terms of immune changes, demonstrating a significant downregulation of interferon alpha and gamma signalling. GSEA results demonstrating the changes in pathways such as immune and cMYC pathways are included in the relevant results chapters (chapters 4 and 5 respectively).

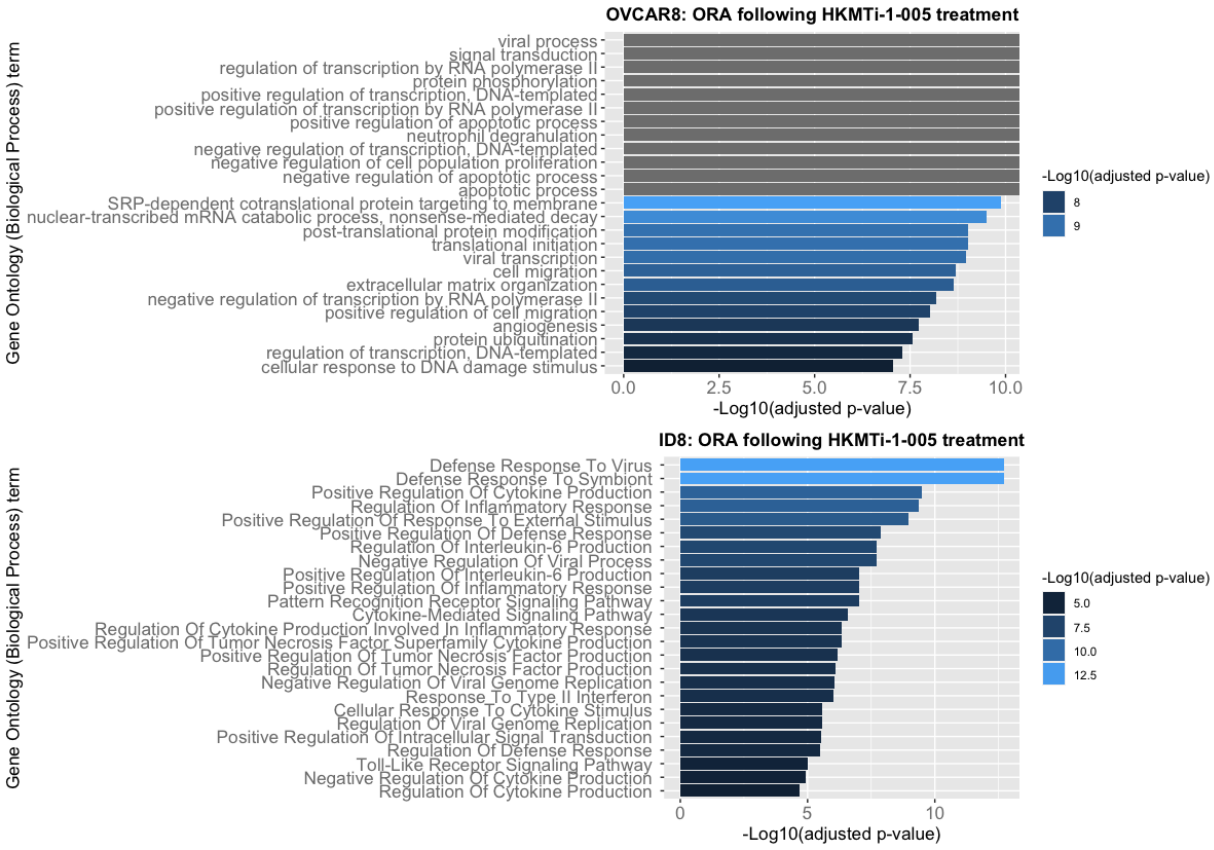


(a)



(b)

Figure 3.7 **RNAseq ORA results, HGSC samples.** Continued overleaf.



(c)

Figure 3.7 **RNAseq ORA results, HGSC samples.** ORA results showing the top 25 enriched Gene Ontology Biological Process (GOBP terms), based on DEGs following HKMTi-1-005 treatment in (a) PEA2 and PEO1 cell lines, (b) PEO4 and OVCAR5 cell lines, (c) OVCAR8 cell line and ID8 *in vivo* HGSC model. Lighter blue bar colour represents smaller p-adj (or greater -log10(p-adj)), see legend. Grey bars represent p-adj of zero, values that could not be log10-transformed. ORA was performed using the geneSCF tool, except for ID8 *in vivo* mouse data where the EnrichR tool was used. DEGs are defined as genes with p-adj < 0.05 and absolute log2FC > 0 following HKMTi-1-005 treatment.

3.2.4 HKMTi-1-005 transcriptional effects at EZH2 and G9a targets

Casciello et al. demonstrated that *IL24* is silenced by dual G9a and EZH2 repression in several solid tumour cell lines, including HGSC lines, with single inhibition of either enzyme insufficient for its induction [107]. They further showed that IL24 was both necessary and sufficient for cell death following dual G9a and EZH2 inhibition in these lines [107]. Hence, *IL24* expression was used as a gene marker for preliminary assessment of G9a and EZH2 inhibitory activity in our HGSC models. As anticipated, a positive fold-change was observed in each HGSC cell line and the *in vivo* sample, supporting G9ai and EZH2i by HKMTi-1-005 (figure 3.8).

Expanding the analysis, I looked to assess changes in genesets reflecting G9a and EZH2 activity. The molecular signatures database (MSigDB) [416] was mined for genesets relating to EZH2 or G9a regulation of gene expression. No G9a genesets in an appropriate model were found; the single geneset available was defined as genes upregulated in the placenta of mice upon knockout of *G9a*. However, three paired genesets describing EZH2-repressed and EZH2-activated genes were identified, defined using knockdown approaches in ovarian cancer cells, prostate cancer cells, or human embryonic fibroblast cells. Additionally, a geneset corresponding to dual EZH2/G9a regulation was described by Casciello et al. [107]. This included genes that were up- and down-regulated across three cell line models consistently (breast, ovarian and melanoma cancer lines) following simultaneous inhibition of both EZH2 and G9a, but not with inhibition of either individually [107].

Whilst GSEA is an excellent tool for a broad-sweep analysis of changes in pathways, it is unable to estimate small p-values accurately [404]. Hence, changes in the EZH2 and dual EZH2/G9a genesets identified were assessed using fast gene set enrichment analysis (fGSEA), which allows estimation of small p-values [404]. These analyses strongly supported EZH2i and G9ai activity of HKMTi-1-005 in the HGSC cell lines and ID8 *in vivo* samples: genes repressed by EZH2 were significantly upregulated

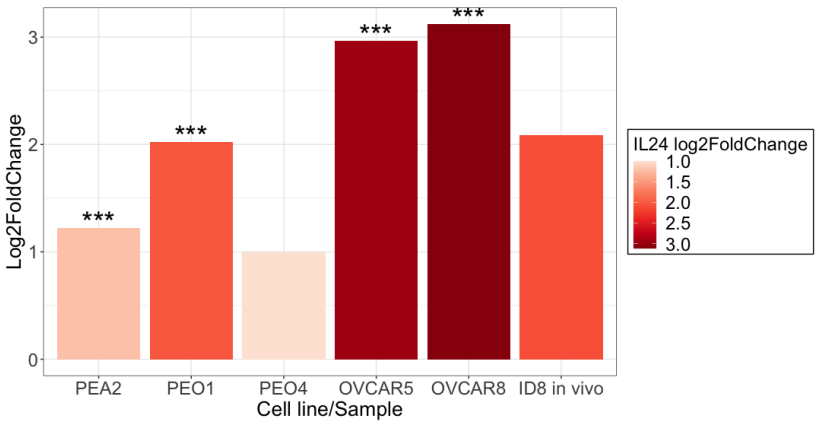


Figure 3.8 ***IL24* differential expression following HKMTi-1-005 treatment.** Change in expression of the *IL24* gene control versus HKMTi-1-005-treated, in the PEA2, PEO1, PEO4, OVCAR5, OVCAR8 cell lines and the ID8 *in vivo* sample. * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001.

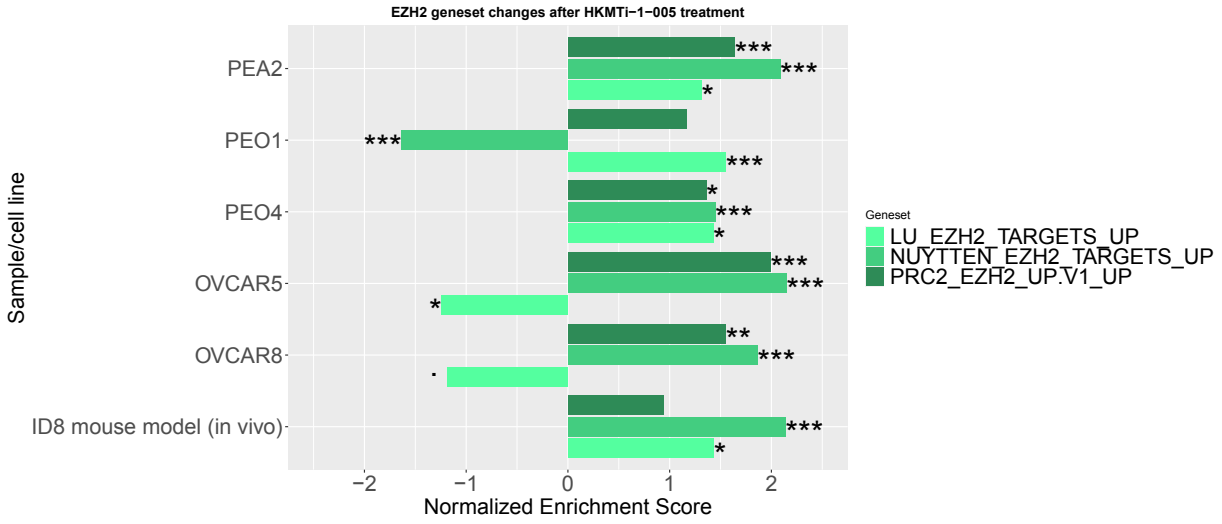
following treatment in most samples (figure 3.9a), those activated by EZH2 were downregulated (figure 3.9b), with significant change in the dual EZH2/G9a geneset in every sample (figure 3.9c).

3.2.5 Preliminary investigation into potential non-enzymatic EZH2i effects of HKMTi-1-005

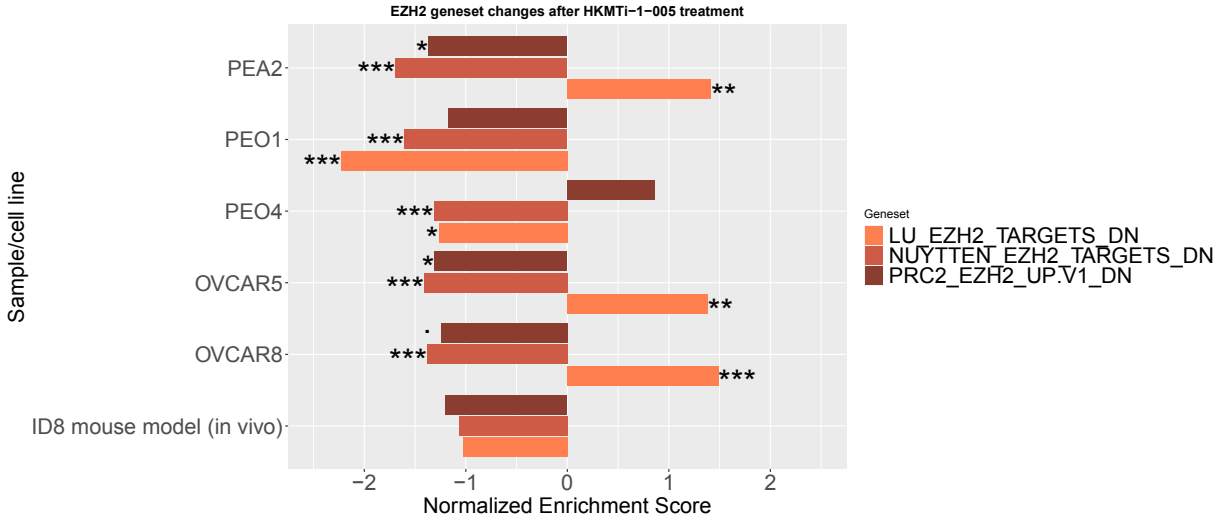
Given that HKMTi-1-005 was developed to inhibit EZH2 methyltransferase activity, it was interesting to note that treatment downregulated EZH2-activated genes that are not necessarily dependent on EZH2 enzymatic activity (see 1.3.3). Additionally, the three paired EZH2 genesets assessed by fGSEA were created based on DEGs following EZH2 knockdown by RNAi, rather than enzymatic inhibitors. These data suggested a potential effect of HKMTi-1-005 on PRC2 or EZH2 more broadly, perhaps at the expression or protein level. To investigate whether HKMTi-1-005 effects could in part be due to depletion of histone methyltransferase machinery, MSigDB genesets relating to H3K27 or H3K9 methylation modifiers were analysed. The genesets consist of very few genes, encoding histone modifier complex components or principal interacting factors. These data indicated that although adjusted p-values did not reach significance in most instances, a strong trend towards downregulation of histone genesets was observed following HKMTi-1-005 treatment (figure 3.10).

I next looked at expression levels of PRC2 components and recruitment factors. EZH2 was not consistently downregulated at the mRNA level according to RNAseq data; the only significant effect was downregulation of EZH2 in the PEO1 line. Broadening to all PRC2 components as defined in [61], the effect of HKMTi-1-005 treatment was assessed. This demonstrated a weak trend towards downregulation of PRC2 components, but with a mixed result across samples (figure 3.11). The *PHF19* gene (encoding the PLC3 PRC2 subunit) was the most consistently downregulated component, with significant downregulation in 4 of the 5 cell lines (p-adj < 0.05), although it was not differentially expressed in the ID8 *in vivo* sample.

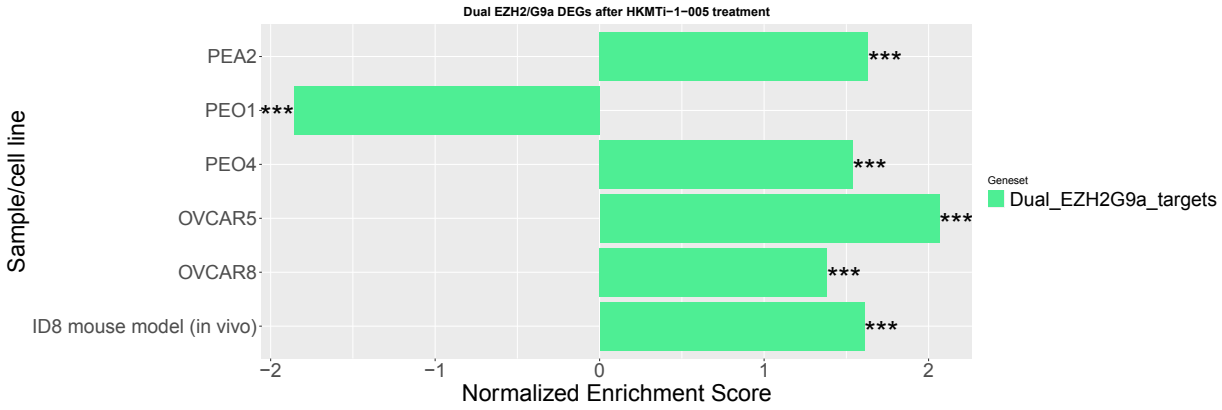
Subsequently, I assessed changes in PRC2 recruitment factors. Defining such factors is complex: PRC2 utilises a broad range of chromatin targeting mechanisms (section 1.3.1), many of which remain under investigation to determine their relevance and some of which may be non-specific RNAs capable of binding and tethering EZH2 or SUZ12 [61]. Others, such as the H3K27me3 mark and alternative histone marks, PRC1 occupancy, DNA conformation and CGI density cannot be captured by transcriptome analysis [61]. Hence, I analysed transcriptional changes in lncRNAs that are proposed to recruit PRC2 [428] and can be measured by RNAseq. There was no consistent reduction in PRC2-recruiting lncRNAs following HKMTi-1-005 treatment (figure 3.12), suggesting EZH2 target gene effects may be independent of changes in PRC2 recruitment, or mediated by changes in alternative recruitment factors or PRC2 components.



(a)



(b)



(c)

Figure 3.9 **fGSEA on RNAseq data: EZH2 and G9a genesets.** fGSEA results for PEA2, PEO1, PEO4, OVCAR5, OVCAR8 cell lines and ID8 *in vivo* sample, comparing RNAseq data from control versus HKMTi-1-005-treated samples. (a) fGSEA results for EZH2 genesets, based on genes that are upregulated when EZH2 is knocked down. (b) fGSEA results for EZH2 genesets, based on genes that are downregulated when EZH2 is knocked down. (c) fGSEA results for a geneset defined by Casciello et al. [107], based on genes that are differentially expressed following dual inhibition of both EZH2 and G9a, but not when either enzyme is inhibited individually. . = p-adj < 0.1, * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001.

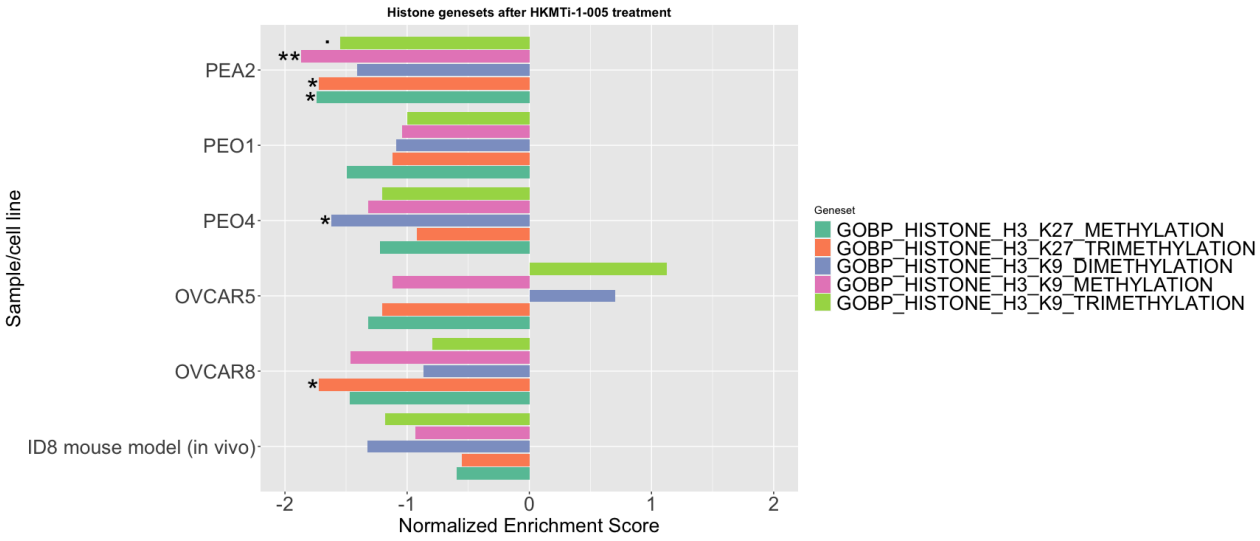


Figure 3.10 **fGSEA on RNAseq data: histone genesets.** fGSEA results for PEA2, PEO1, PEO4, OVCAR5, OVCAR8 cell lines and ID8 *in vivo* sample, comparing control and HKMTi-1-005-treated samples. Genesets include those available at MSigDB [416] relating to histone 3 lysine 9 (H3K9) or histone 3 lysine 27 (H3K27) methylation. . = p-adj < 0.1, * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001.

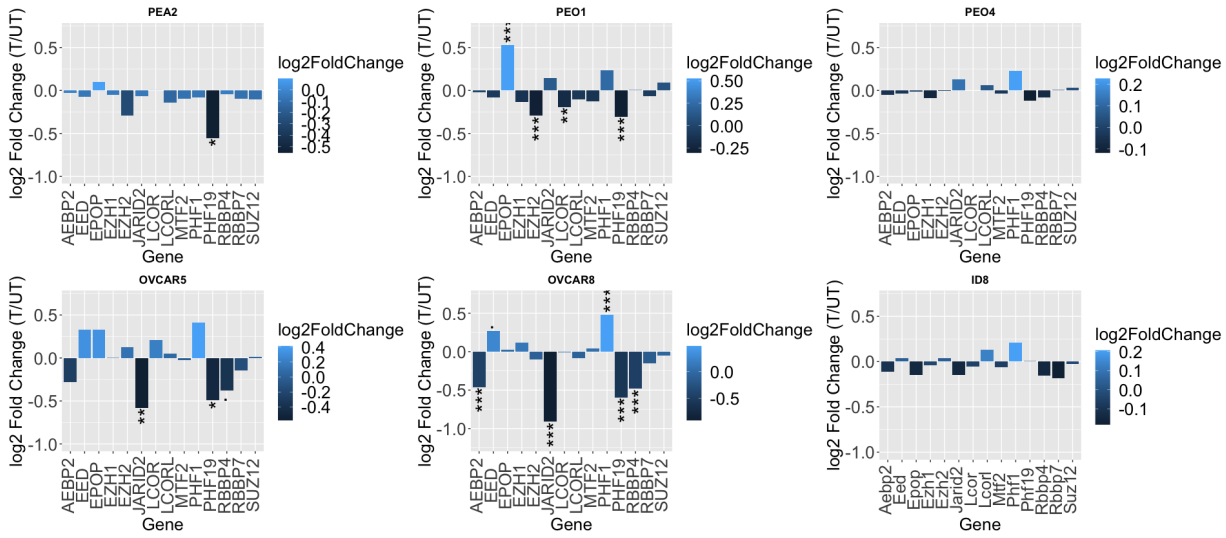


Figure 3.11 **PRC2 components: differential expression.** Transcriptional changes in the PRC2 components as measured by RNAseq, in the PEA2, PEO1, PEO4, OVCAR5, OVCAR8 cell lines and ID8 *in vivo* sample, comparing control and HKMTi-1-005-treated samples. PRC2 components were defined in [61], with mouse orthologues assessed for ID8 *in vivo* sample. EZH1/EZH2 = Enhancer of Zeste Homologue 1/2, SUZ12 = SUZ12 polycomb repressive complex 2 subunit, RBBP4/7 = Retinoblastoma-Binding Protein 4/7, PALI1/2 = LCOR/LCORL = Phylogeny and ALIgnment of homologous protein structures 1/2, PCL1/2/3 = PHF1/MTF2/PHF19 = Polycomblike 1/2/3, JARID2 = Jumonji And AT-Rich Interaction Domain Containing 2, AEBP2 = AE Binding Protein 2. . = p-adj < 0.1, * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001.

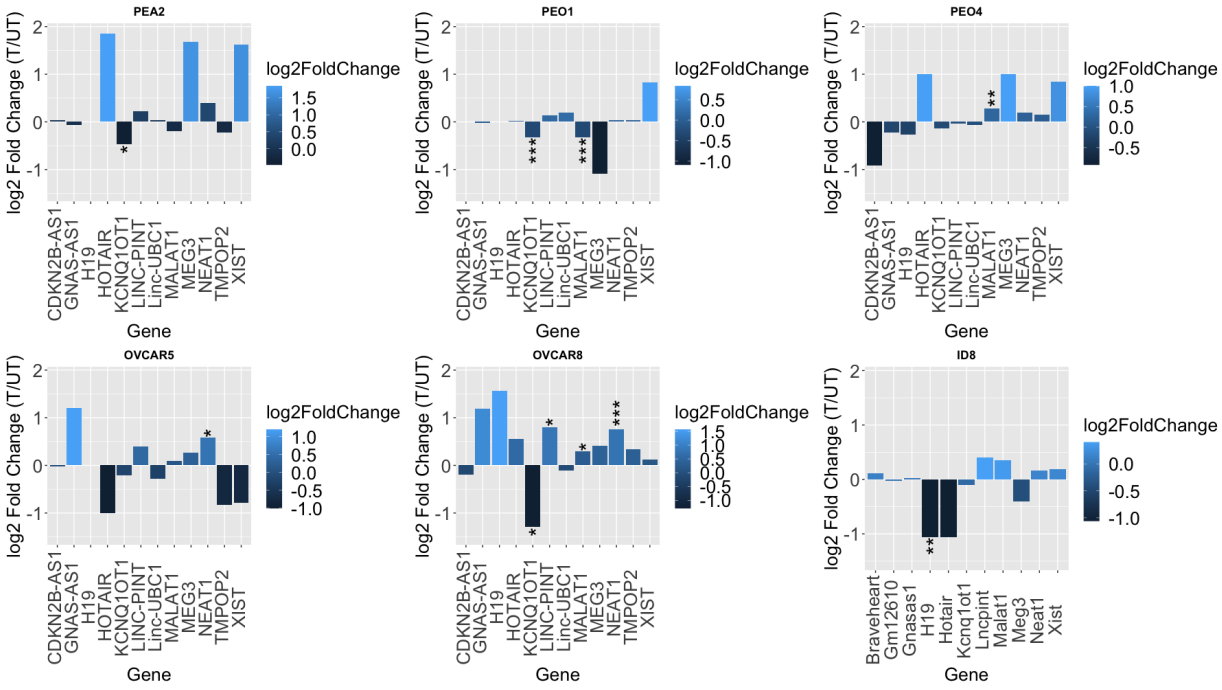


Figure 3.12 **PRC2 recruitment factors: differential expression.** Transcriptional changes in lncRNAs proposed to recruit PRC2 [428]. Based on RNAseq in the PEA2, PEO1, PEO4, OVCAR5, OVCAR8 cell lines and ID8 *in vivo* sample, comparing control and HKMTi-1-005-treated samples. CDKN2B-AS1 = CDKN2B Antisense RNA 1, GNAS-AS1 = GNAS antisense RNA 1, H19 = H19 imprinted maternally expressed transcript, HOTAIR = HOX Transcript Antisense RNA, KCNQ1OT1 = KCNQ1 Opposite Strand/Antisense Transcript 1, LINC-PINT = Long Intergenic Non-Protein Coding RNA, P53 Induced Transcript, Linc-UBC1 = Long Intergenic Noncoding RNA Up-Regulated In Bladder Cancer, MALAT1 = metastasis associated lung adenocarcinoma transcript 1, MEG3 = maternally expressed 3, NEAT1 = Nuclear Paraspeckle Assembly Transcript 1, TMPOP2 = thymopoietin pseudogene 2, XIST = X Inactive Specific Transcript. Where equivalent mouse orthologues existed, these were assessed in the ID8 *in vivo* sample. Additionally, the Braveheart lncRNA that exists in mice but not humans was assessed in the ID8 *in vivo* sample. .= p-adj < 0.1, * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001.

3.3 *In vitro* and *in vivo* chromatin accessibility analysis: HKMTi-1-005 treatment effect on HGSC

3.3.1 ATACseq data processing: HGSC cell lines

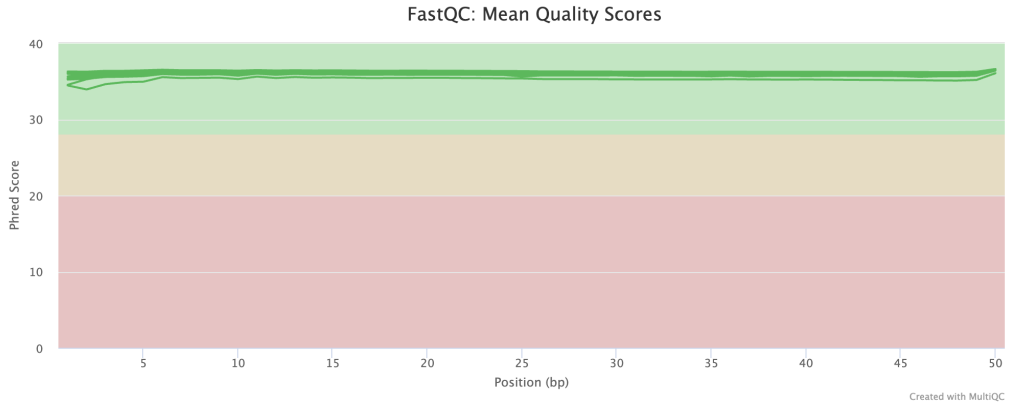
An assay for transposase-accessible chromatin with sequencing (ATACseq) was performed on the paired PEO1 and PEO4 cell lines by Dr Ian Garner. Cells were treated with HKMTi-1-005 in triplicate at their IC_{50} values for 72h, followed by harvesting and processing using the protocol described by Brunton et al. [385]. Samples were sent to the ICR for sequencing ahead of downstream processing.

I first processed raw data using FastQC [371] to assess quality metrics, followed by read trimming with trimmomatic [373]. QC metrics such as Phred score, per-base N content and adapter content were collected using FastQC [371] and were found to be satisfactory (figures 3.13a and 3.13b). Reads were aligned to the GRCh38 genome release 79 using Bowtie2 after building genome indices [386]. MultiQC [372] was used to collate QC metrics and an in-house threshold of minimum 25M reads, at least 90% total alignment, with a maximum 55% multiple alignments was used for sample selection. This resulted in exclusion of the third untreated PEO4 replicate (PEO4 UT3), due to a low total number of reads (figure 3.13c).

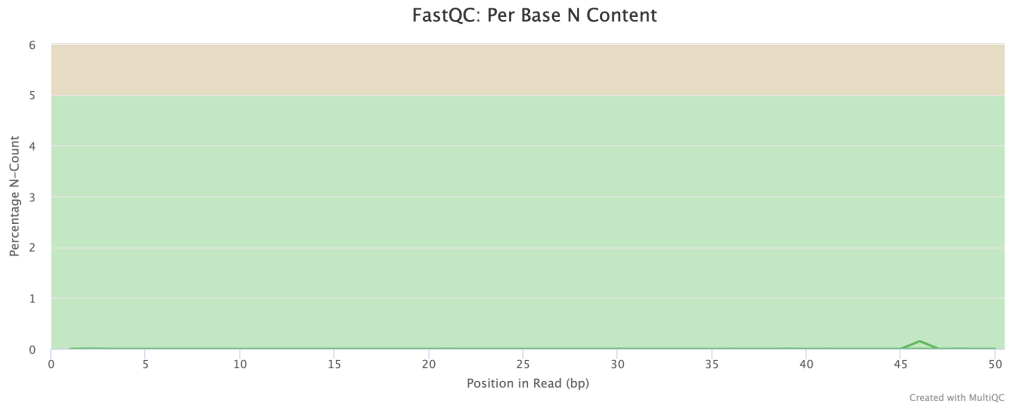
Reads were filtered to remove mitochondrial alignments and a further QC step was performed using the Picard tool CollectInsertSizeMetrics. The distribution of insert sizes was plotted, demonstrating that many short inserts fall in open chromatin, with subsequent peaks at insert lengths reflecting mono- and di-nucleosome-sized fragments as expected (figure 3.14). Final filtering steps were performed to retain only properly paired reads as per the ENCODE guidelines [387], prior to read shifting using alignmentSieve to account for the insertion of the Tn5 transposase enzymes used in ATACseq [388]. Peaks were called using the Macs2 tool [389], followed by filtering to remove blacklisted regions [426].

Subsequently, two approaches were used to assess changes in accessibility induced by HKMTi-1-005. These are summarised in figure 3.15. The first approach (the DAC approach) uses an RNAseq-like method, to determine a fold-change in accessibility with an associated p-value, allowing identification of differentially accessible chromatin regions (DACs). This was performed using DiffBind [390], which allowed plotting of heatmaps and PCA plots to visualise the distance between samples. Both PEO1 and PEO4 demonstrated segregation by treatment status (figures 3.16 and 3.17). The second approach (the subtraction approach, figure 3.15), retains any peaks in the treated (T) condition that do not exist in the untreated (UT) condition. To do so, all peaks in the T condition are merged, followed by subtraction of all merged UT peaks. This method was performed using bedtools

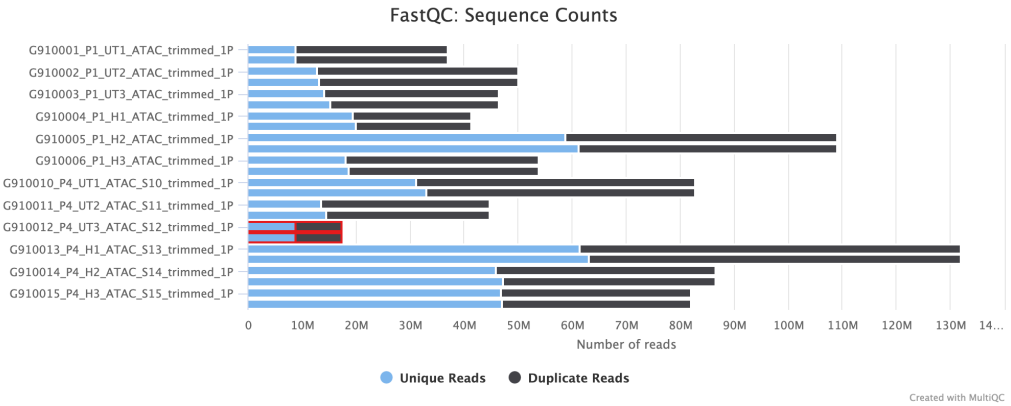
[377] with subsequent annotation of peaks via the HOMER program [392]. Note that the results from each approach overlap highly significantly (appendix table A.4), as calculated by comparison between genes in unique HKMTi-1-005 treatment peaks and genes associated with upregulated DACs.



(a)

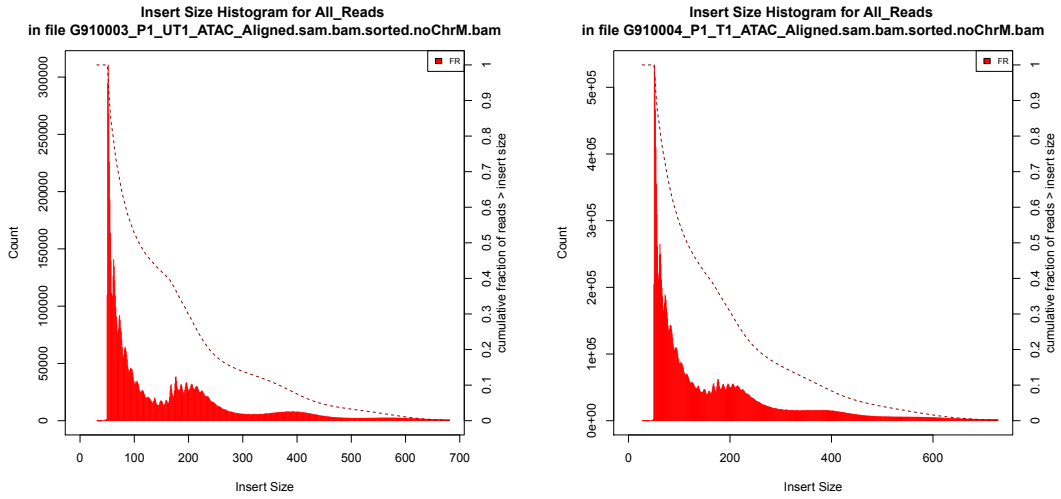


(b)



(c)

Figure 3.13 **QC for *in vitro* HGSC ATACseq data.** QC metrics for ATACseq data derived from PEO1 and PEO4 HGSC cell lines, control and HKMTi-1-005-treated, after read trimming. (a) Phred scores, reflecting a measure of the accuracy of base calling. A Phred score of >30 reflects a 99.9% base call accuracy. (b) Per-base N content, a quality metric based on the percentage of bases with no base call at each read position. (c) FastQC sequence count for each sample, with an estimate of the proportion of duplicate reads. Low read counts were observed for the third PEO4 untreated replicate, indicated in red. P1 = PEO1, P4 = PEO4, both run in triplicate. Treated samples are labelled as H1/H2/H3, whilst untreated samples are labelled UT1/UT2/UT3. All adapter content $< 0.1\%$ for samples listed, hence no adapter contamination graph was included.



(a) (b)

Figure 3.14 **Insert size metrics for *in vitro* HGSC ATACseq data.** Representative plots demonstrating distribution of ATACseq insert sizes, derived from the PEO1 HGSC cell line in (a) one untreated and (b) one treated replicate. Each graph demonstrates a peak of insert sizes < 100bp (open chromatin), ∼200bp (mono-nucleosome) and ∼400bp (di-nucleosome) etc. bp = base pairs.

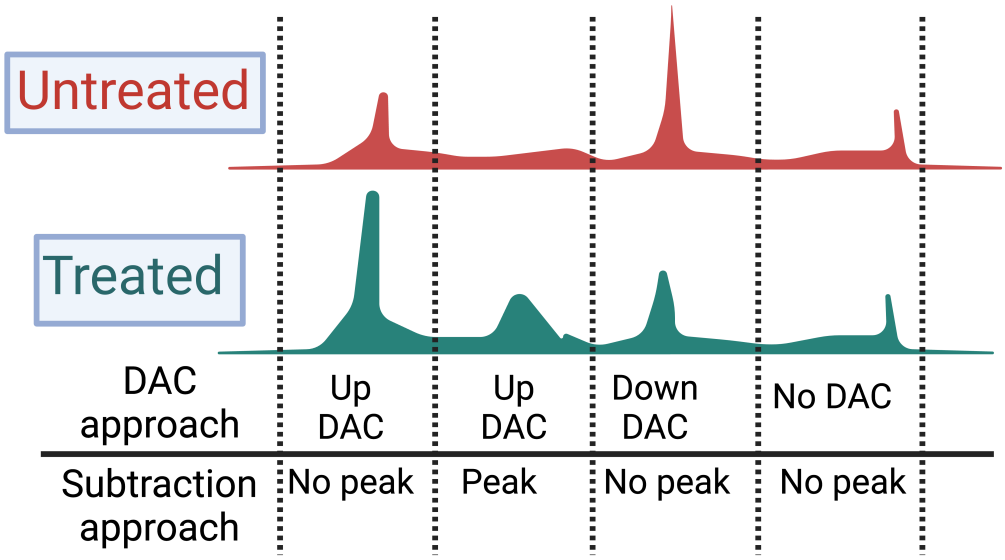
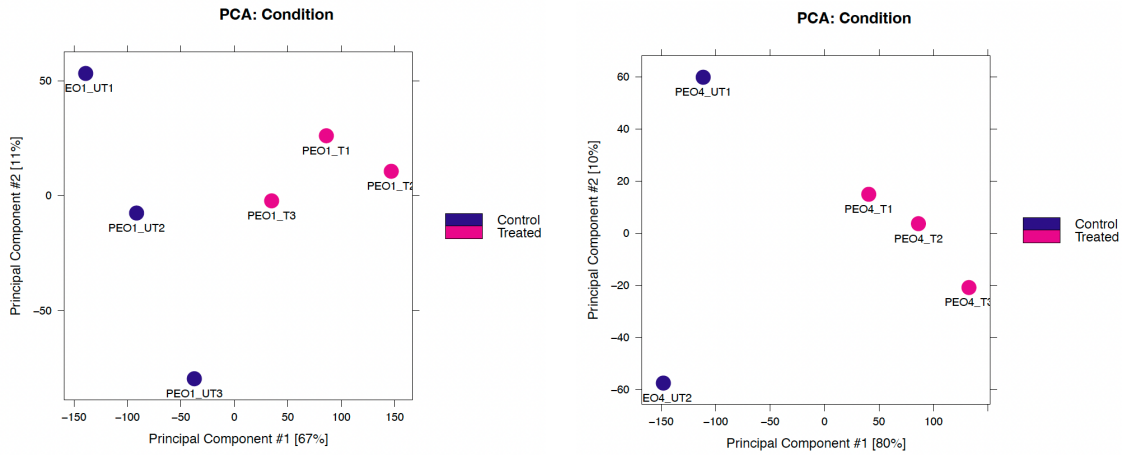
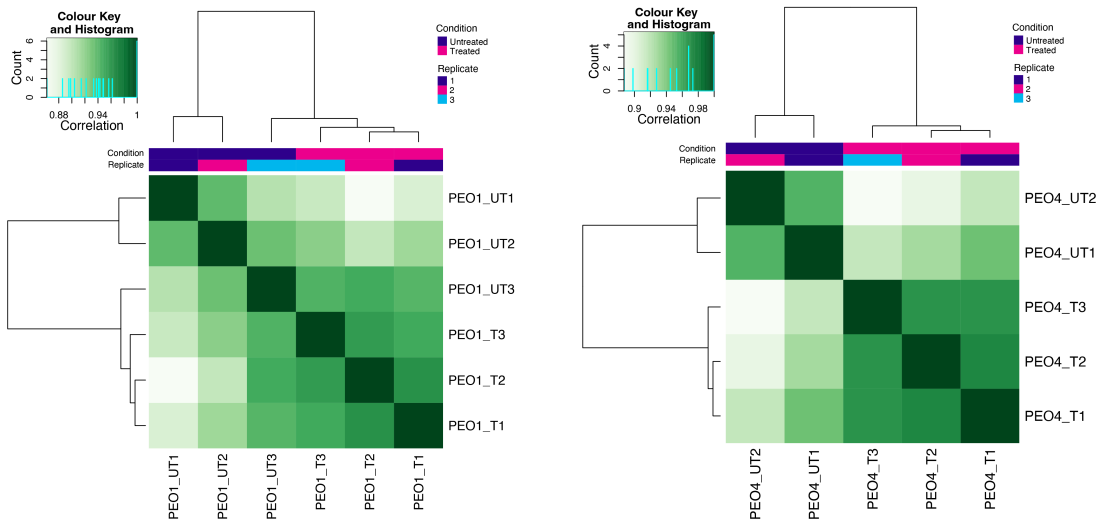


Figure 3.15 **Schematic illustrating the two methods of ATACseq analysis.** ATACseq analysis methods: the DAC approach and the subtraction approach. Green and red peaks represent example accessibility profiles of treated (green) and untreated (red) samples. The DAC approach identifies DACs in either direction, with an associated log2FC value and p-adj, using a pipeline similar to RNAseq. The subtraction approach looks to identify unique HKMTi-1-005 treatment peaks, by subtracting peaks in untreated conditions from those in treated conditions.



(a) (b)

Figure 3.16 **PCA plots for *in vitro* HGSC ATACseq data.** PCA plots after differential accessibility analysis, for ATACseq data derived from (a) PEO1 and (b) PEO4 HGSC cell lines, control and HKMTi-1-005-treated.



(a) (b)

Figure 3.17 **Heatmap plots for *in vitro* HGSC ATACseq data.** Heatmap plots after differential accessibility analysis, for ATACseq data derived from (a) PEO1 and (b) PEO4 HGSC cell lines, control and HKMTi-1-005-treated.

3.3.2 ATACseq data processing: ID8 *in vitro* and *in vivo* murine models of HGSC

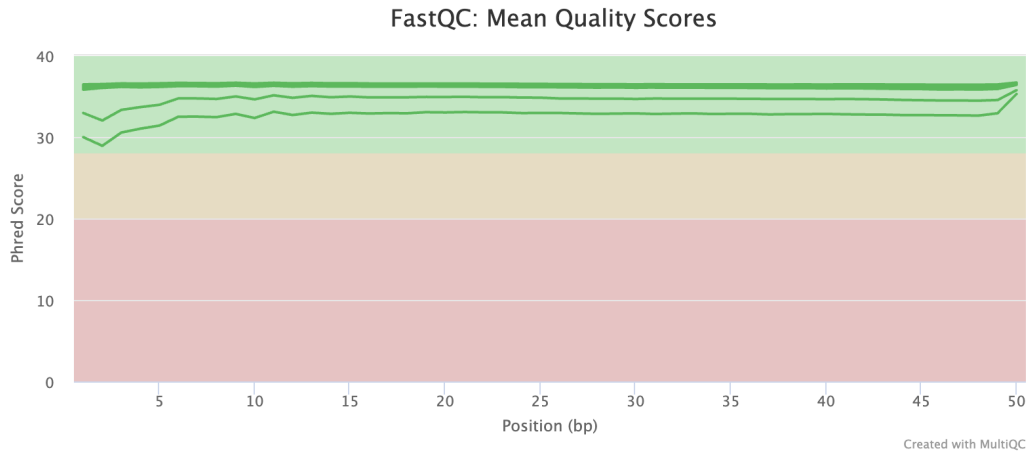
ATACseq data derived from the ID8 *in vivo* model were re-analysed alongside PEO1 and PEO4 cell line data. ATACseq was performed on tumour cells extracted from mice bearing intraperitoneal *Trp53*$^{-/-}$ ID8 tumours treated with HKMTi-1-005 (20 mg/kg intraperitoneally twice a day for 14 days, n=6) or with vehicle (1% Tween/3.4% DMSO in 0.9% NaCl i.p. twice a day for 14 days, n=6). Similar to the re-analysis of RNAseq data, this included an updated genome alignment (mm39 as compared to mm10) and a broader assessment of HKMTi-1-005-mediated effects compared to the initial analyses described [155].

In addition to human HGSC cell lines and the ID8 *in vivo* tumours, several ID8 cell line models of HGSC were treated with HKMTi-1-005 and harvested for ATACseq, to consolidate effects of HKMTi-1-005 on chromatin. This was performed by Dr Ian Garner and included four CRISPR-generated ID8 genotypes in duplicate, treated as per the PEO1/PEO4 protocol [385]: ID8 *Trp53*$^{-/-}$;*Brca1*$^{-/-}$ (referred to as BRCA1) [396], ID8 *Trp53*$^{-/-}$;*Brca2*$^{-/-}$ (referred to as BRCA2) [394], ID8 *Trp53*$^{-/-}$ (referred to as F3) [394] and the ID8 CRISPR control (referred to as C3) [394].

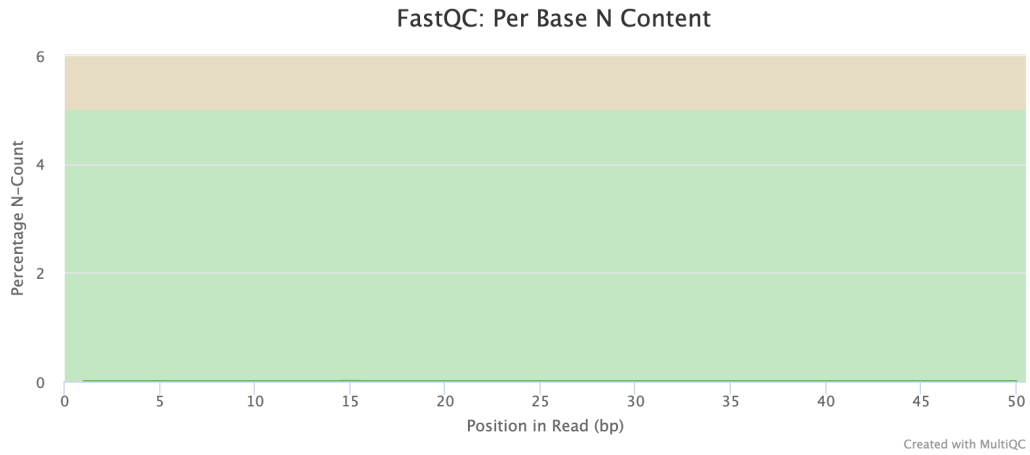
ID8 *in vitro* and *in vivo* data were processed as similarly as possible to the PEO1 and PEO4 HGSC lines. ID8 *in vivo* samples had been demonstrated to have sufficient quality scores following trimming (figure 3.18), excepting one sample that had already been excluded due to a low read count prior to trimming (appendix figure A.5). ID8 *in vitro* data were processed using FastQC [371], trimmomatic [373] and MultiQC [372] for trimming and QC. All samples demonstrated adequate Phred scores, per-base N content and adapter content (figure 3.19). Due to space limitations on the data processing server, the bwa-mem aligner tool was used in place of Bowtie2 for read alignment to GRCm39 for ID8 *in vivo* and *in vitro* samples. Following alignment, one treated ID8 *in vitro* sample (C3 T2) was removed due to low total read count (<25M) according to the in-house ATACseq thresholds described (figure 3.19c).

Filtering, shifting and peak calling were performed according to an identical pipeline to that described for the HGSC cell line data, as were the two downstream approaches to assess changes in accessibility following HKMTi-1-005 treatment (figure 3.15). As ID8 *in vitro* samples were only treated in duplicate, the filtering of C3 T2 left only one treated sample, hence C3 was excluded from further analyses. Insert sizes were assessed across the remaining samples (figures 3.20 and 3.21). Of note, insert size analysis demonstrated that the ID8 *in vivo* data consisted of a high quantity of short insert sizes in every treated sample, but not untreated samples, suggesting a highly accessible chromatin structure induced by HKMTi-1-005 (figure 3.20).

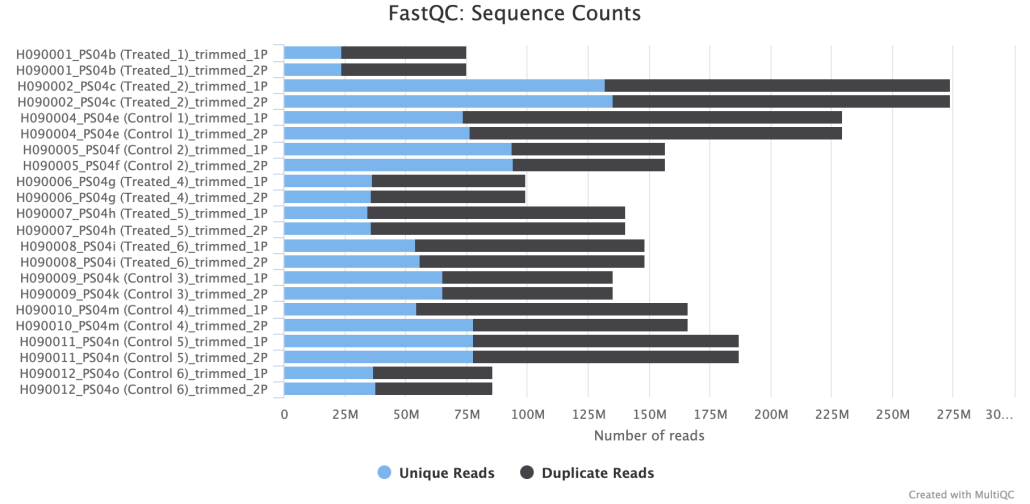
The differential accessibility approach, performed using DiffBind [390], allowed plotting of PCA



(a)



(b)

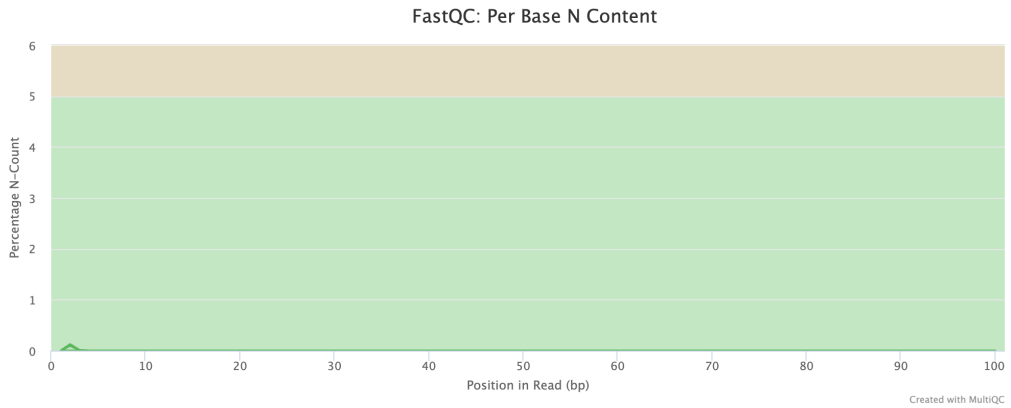


(c)

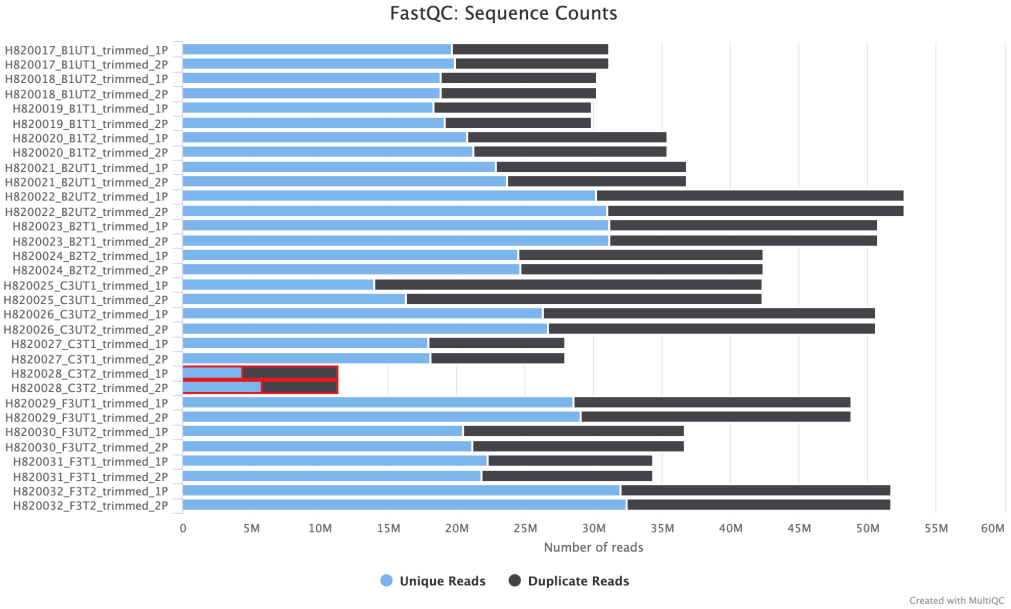
Figure 3.18 **QC metrics for ID8 *in vivo* ATACseq data.** QC metrics for ATACseq data derived from ID8 *in vivo* samples, control and HKMTi-1-005-treated, after read trimming. (a) Phred scores, reflecting a measure of the accuracy of base calling. A Phred score of >30 reflects a 99.9% base call accuracy. (b) Per-base N content, a quality metric based on the percentage of bases with no base call at each read position. (c) FastQC sequence count for each sample, with an estimate of the proportion of duplicate reads. All adapter content < 1% for samples listed, hence no adapter contamination graph was included. N=5 treated (PS04b, PS04c, PS04g, PS04h, PS04i) and n=6 untreated (PS04e, PS04f, PS04k, PS04m, PS04n, PS04o) replicates. N.B. One treated ID8 *in vivo* sample (PS04d, T3) was removed prior to trimming, due to low read count (<25M).



(a)

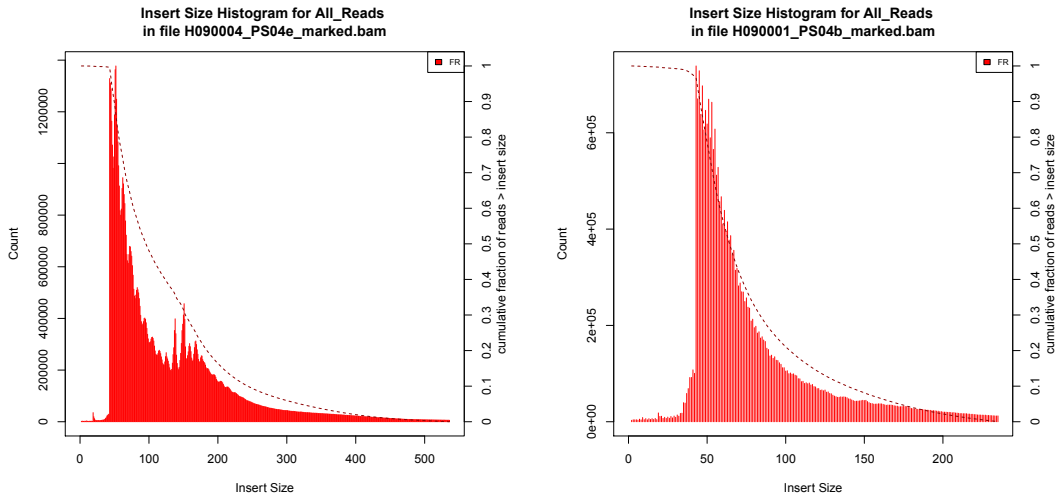


(b)



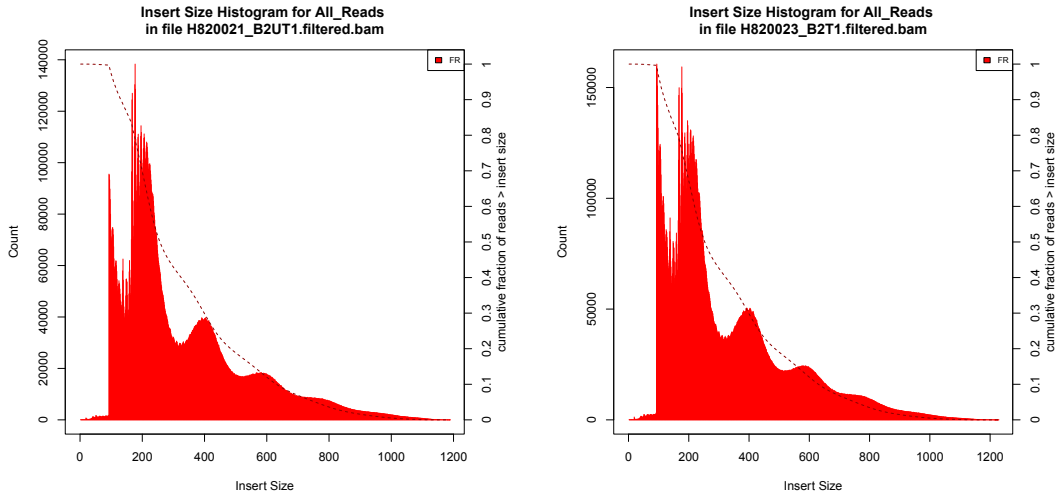
(c)

Figure 3.19 **QC metrics for ID8 *in vitro* ATACseq data.** QC metrics for ATACseq data derived from CRISPR-generated ID8 cell lines (BRCA1, BRCA2, C3, F3), control and HKMTi-1-005-treated, after read trimming. (a) Phred scores, reflecting a measure of the accuracy of base calling. A Phred score of >30 reflects a 99.9% base call accuracy. (b) Per-base N content, a quality metric based on the percentage of bases with no base call at each read position. (c) FastQC sequence count for each sample, with an estimate of the proportion of duplicate reads. Low read counts were observed for the second C3 treated replicate, indicated in red. B1 = BRCA1, B2 = BRCA2. Treated samples are labelled as T1/T2/T3, whilst untreated samples are labelled UT1/UT2/UT3. All adapter content $< 0.1\%$ for samples listed, hence no adapter contamination graph was included.



(a) (b)

Figure 3.20 **Insert size metrics for ID8 *in vivo* ATACseq data.** Representative plots demonstrating distribution of ATACseq insert sizes, derived from the ID8 *in vivo* mouse model of HGSC in (a) one untreated and (b) one treated replicate. The untreated sample demonstrates a peak of insert sizes < 100bp (open chromatin), ∼200bp (mono-nucleosome) and ∼400bp (di-nucleosome) etc., whilst the treated sample shows a highly open chromatin structure with many short insert sizes. bp = base pairs.



(a) (b)

Figure 3.21 **Insert size metrics for BRCA1 ID8 *in vitro* ATACseq data.** Representative plots demonstrating distribution of ATACseq insert sizes, derived from the BRCA1 ID8 *in vitro* mouse model of HGSC in (a) one untreated and (b) one treated replicate. Each graph demonstrates a peak of insert sizes < 100bp (open chromatin), ∼200bp (mono-nucleosome) and ∼400bp (di-nucleosome) etc. bp = base pairs.

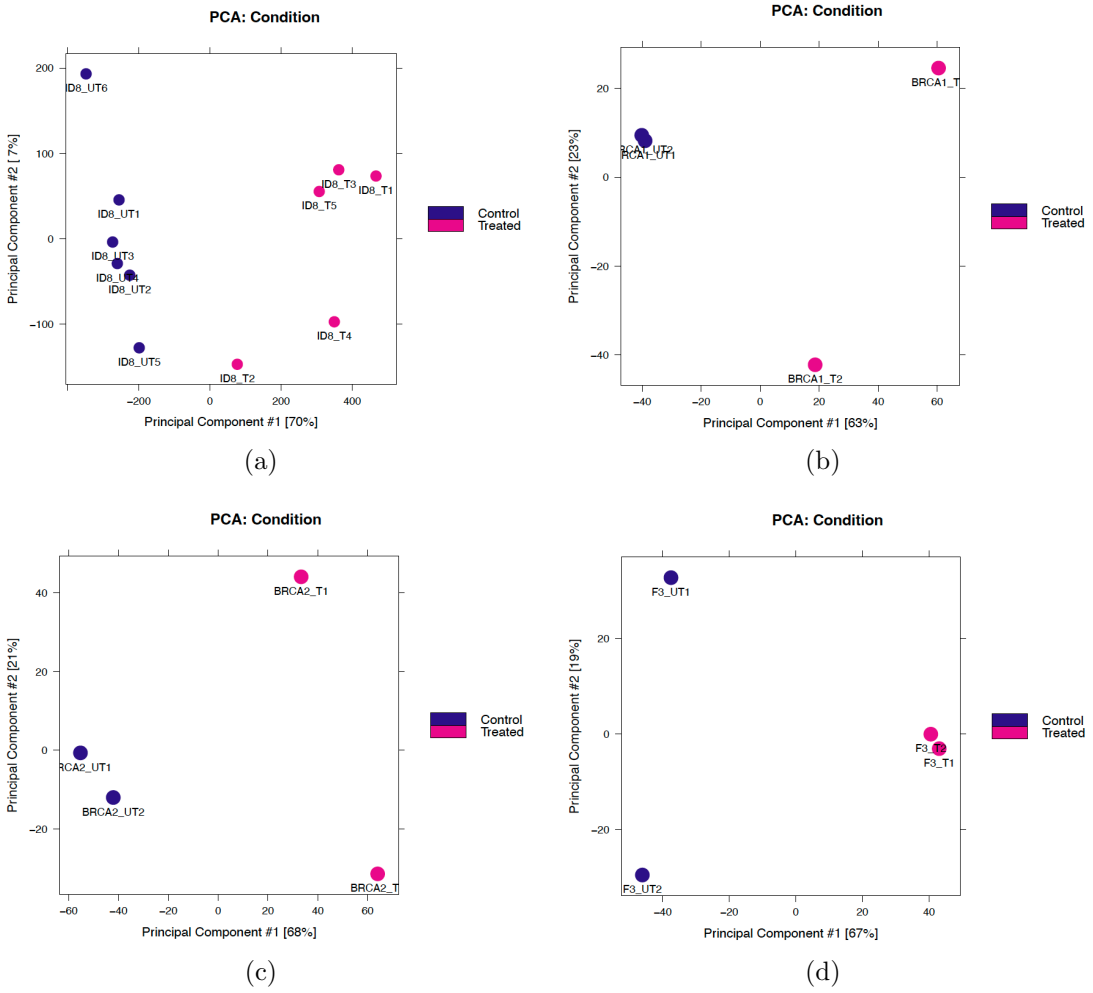


Figure 3.22 **PCA plots for ID8 *in vivo* and *in vitro* ATACseq data.** PCA plots after differential accessibility analysis, for ATACseq data derived from the (a) ID8 *in vivo* mouse HGSC model, (b) BRCA1 ID8 cell line, (c) BRCA2 ID8 cell line and (d) F3 ID8 cell line, following HKMTi-1-005 treatment.

graphs. The choice of normalisation parameters for DiffBind analyses in the *in vitro* samples differed from that used *in vivo*, as discussed in the Methods section (see 2). Nonetheless, both ID8 *in vitro* and *in vivo* samples demonstrated a clear treatment effect, with clustering of treated samples separately from untreated samples (figure 3.22).

3.3.3 Overview of chromatin accessibility effects

To address the most basic question of whether chromatin showed the expected increase in accessibility following HKMTi-1-005 treatment, I plotted the number of peaks of accessibility for each sample in control versus HKMTi-1-005-treated conditions. To do so, the mean number of peaks across replicates for each condition and each sample was calculated. This indicated that in all samples except the ID8 BRCA1 cell line, there was an increase in the number of peaks of accessibility following HKMTi-1-005 treatment (figure 3.23). This effect was greatest in the ID8 *in vivo* sample, followed by the HGSC cell lines.

An overview of the extent of differentially accessible chromatin and number of sites can be visualised using a volcano plot (figure 3.24). These indicated that there was a very strong skew towards upregulated DACs in the ID8 *in vivo* sample, with a more mixed effect across the two *in vitro* sample sets. DACs were defined as p-adj < 0.05 with any log2FC, except for the ID8 *in vivo* where a minimum absolute log2FC of 1 was set to increase the stringency, given the extensive chromatin accessibility changes observed. Across samples there was a large effect on chromatin accessibility, with between 3,462 (PEO1) and 53,824 (ID8 *in vivo*) DACs after treatment with HKMTi-1-005.

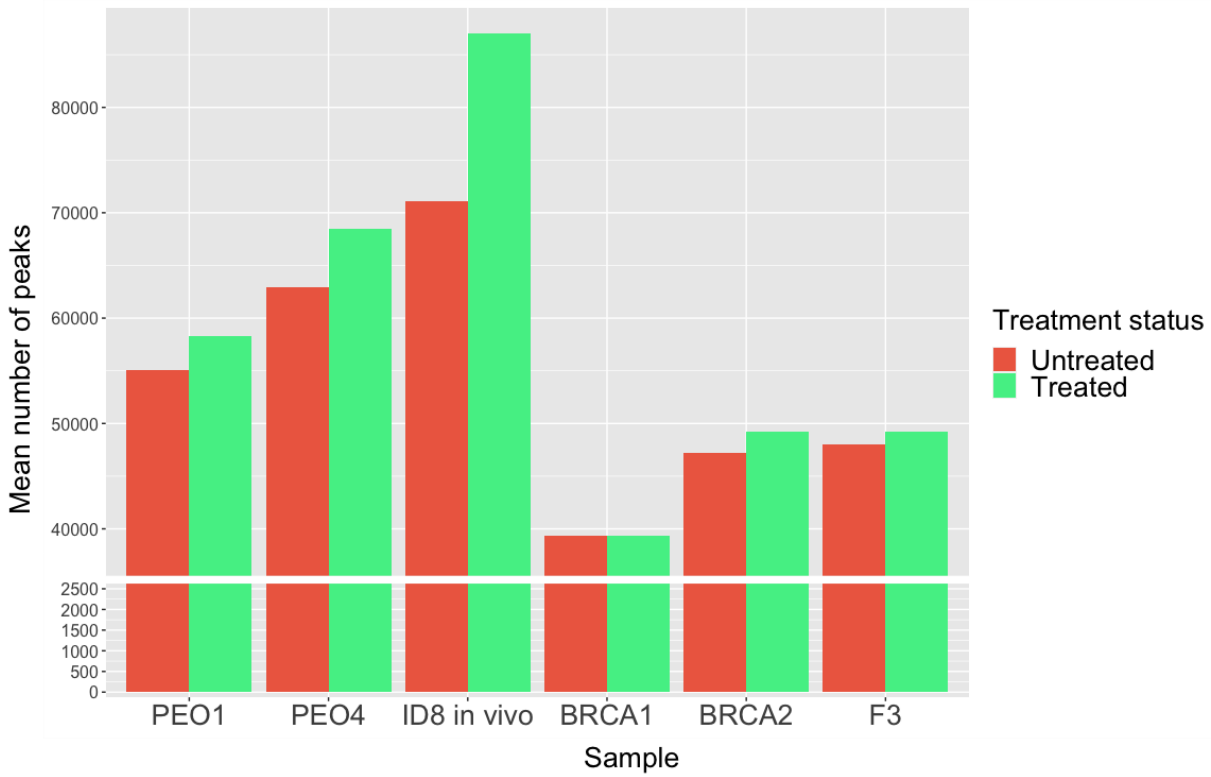
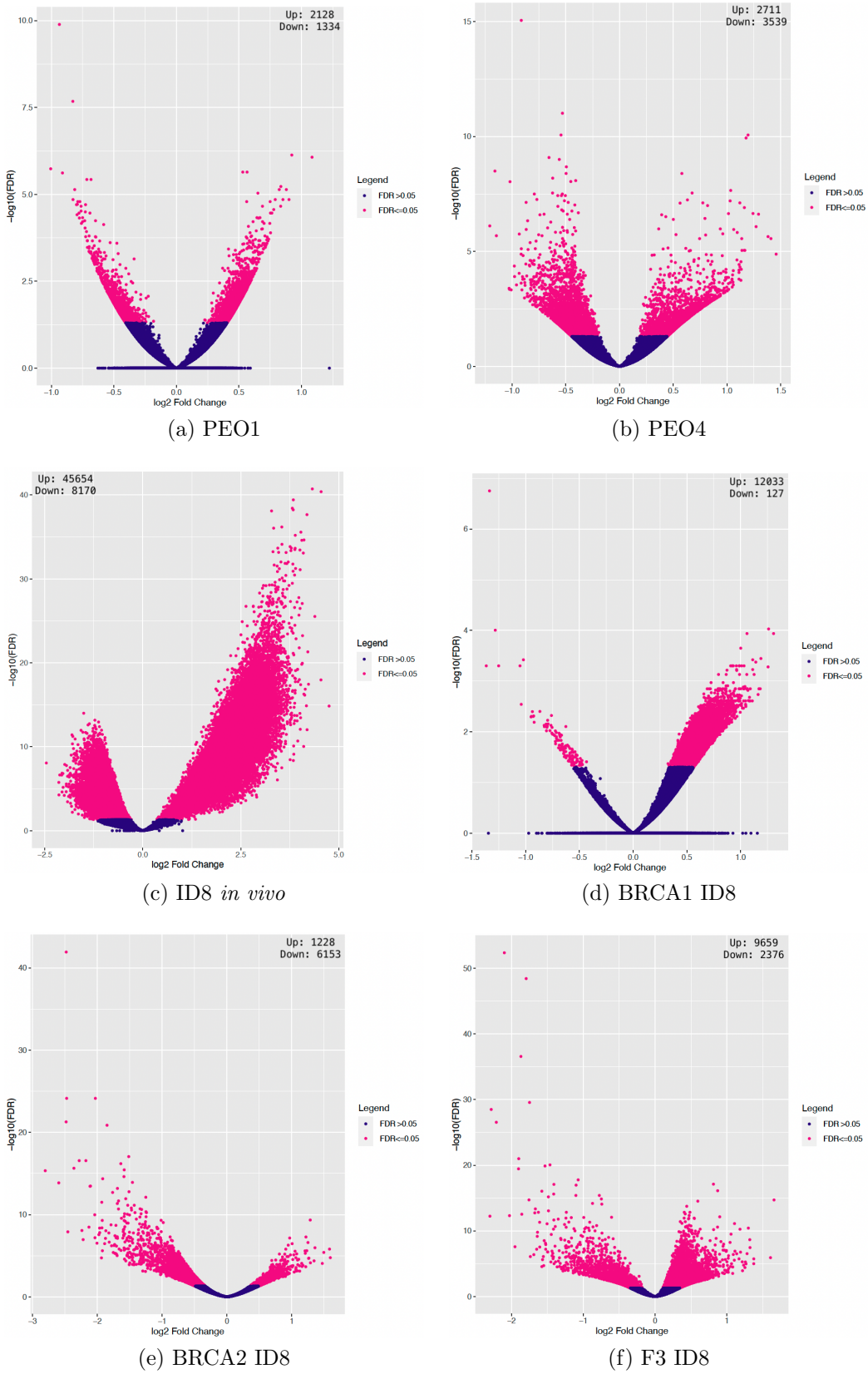


Figure 3.23 **ATACseq peaks of accessibility, in control versus HKMTi-1-005-treated conditions.** Control versus HKMTi-1-005-treated peaks of accessibility, as measured by ATACseq. Samples include the PEO1 cell line, the PEO4 cell line, the ID8 *in vivo* mouse model, the BRCA2 ID8 cell line, the BRCA2 ID8 cell line and the F3 ID8 cell line. The mean number of peaks per treatment condition for each sample is plotted.



(a) PEO1 (b) PEO4

(c) ID8 *in vivo* (d) BRCA1 ID8

(e) BRCA2 ID8 (f) F3 ID8

Figure 3.24 **ATACseq volcano plots.** Volcano plots generated using DiffBind [390], with visualisation of differentially accessible chromatin regions (DACs) in the (a) PEO1 cell line, (b) PEO4 cell line, (c) ID8 *in vivo* sample, (d) BRCA1 ID8 cell line, (e) BRCA2 ID8 cell line and (f) F3 ID8 cell line, control vs HKMTi-1-005-treated. Sites labelled by DiffBind are categorised by colour, based on significance: purple = not significant, pink = FDR < 0.05. N.B. For further analyses, ID8 *in vivo* sample DACs were also required to reach a threshold of absolute log2FC > 1. This threshold was also used to count total up- versus down-regulated DACs.

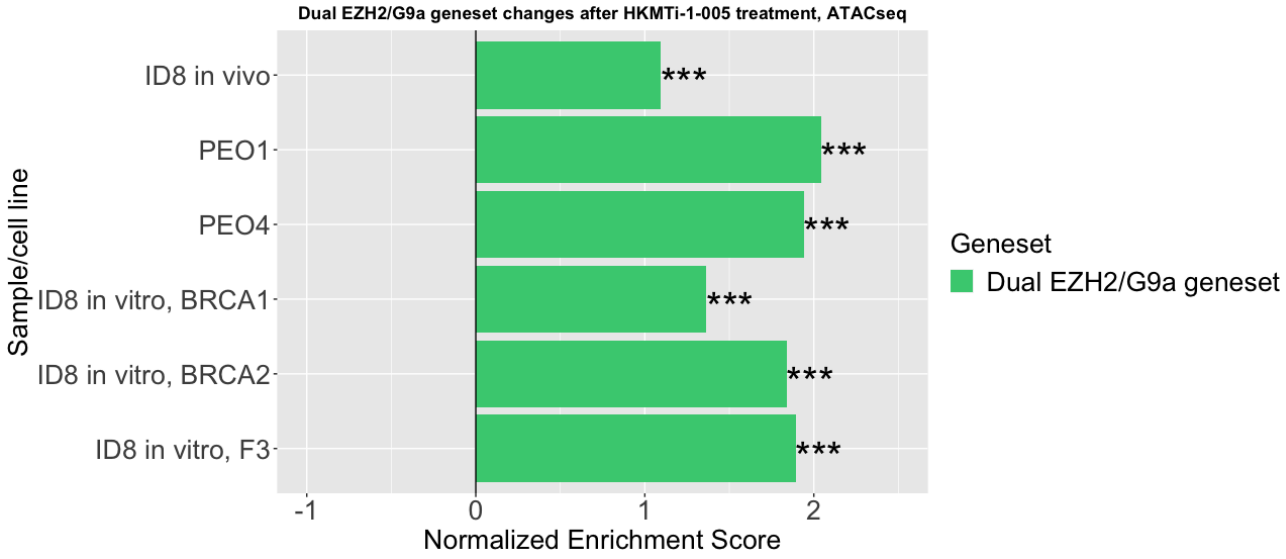
3.3.4 HKMTi-1-005 chromatin effects at EZH2 and G9a targets

Next, I assessed the effect of HKMTi-1-005 on chromatin accessibility more specifically at EZH2 and G9a target sites, beginning with the *IL24* locus. Using the DAC approach, *IL24* was only significantly increased in accessibility in the ID8 *in vivo* sample. However, the subtraction approach demonstrated that IL24 was in unique HKMTi-1-005 treatment peaks in every sample, suggesting increased accessibility at *IL24* may underlie its transcriptional upregulation after HKMTi-1-005 treatment, but that these changes may be subtle.

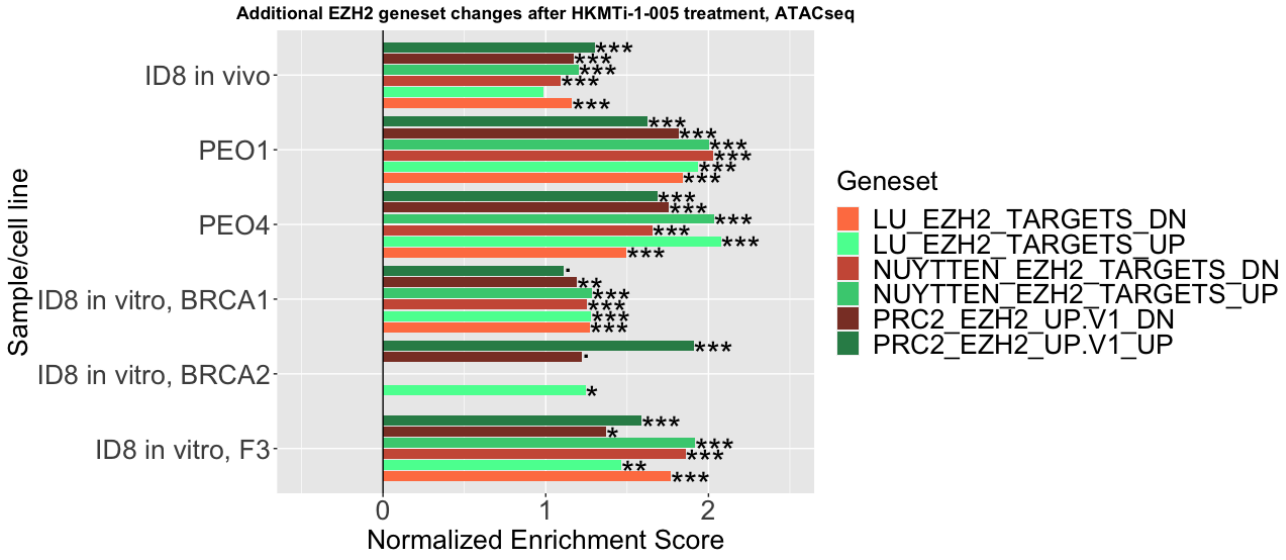
I next investigated accessibility changes in the EZH2- and G9a-regulated genesets defined in sections 3.2.4 and 3.2.5, using fGSEA [404]. These data demonstrated a strong increase in accessibility in both EZH2-only and dual EZH2/G9a genesets across every sample, supporting HKMTi-1-005 inhibitory activity at EZH2 and G9a targets (figure 3.25). It was interesting to note that despite transcriptional downregulation of EZH2-activated genesets, these genesets were more accessible at the chromatin level following HKMTi-1-005 treatment. This supports the hypothesis that HKMTi-1-005 may impact PRC2 complex composition in a manner that prevents its non-canonical transcriptional activation effects, alongside its canonical repressive effects.

Given the lack of G9a genesets, and to determine effects of HKMTi-1-005 at intergenic regions as well as over gene bodies, I examined chromatin accessibility associated with EZH2/G9a more broadly. This involved visualisation of accessibility changes at genomic coordinates associated with H3K27me3 and H3K9me3 methylation marks. Loci were obtained from ENCODE, based on ChIPseq performed in H1 ESCs and mouse neural tube cells (12.5-day embryo), aligned to the hg19 and mm10 genomes respectively. Following liftover to the hg38 and mm39 genomes, accessibility across the HGSC samples was plotted over sites labelled as H3K27me3- or H3K9me3-associated. To determine the specificity of HKMTi-1-005 effects at H3K27me3/H3K9me3 loci, accessibility at loci associated with other histone marks was also visualised, including H3K27Ac, H3K9Ac, H3K36me3, H3K4me1 and H3K4me3. No H3K9me1/2 tracks were available for mouse or human datasets.

These plots demonstrated a clear increase in accessibility at chromatin associated with H3K27me3 and H3K9me3 across all samples assessed (figures 3.26 and 3.27). Importantly, the results suggested that although accessibility at most chromatin domains increased, there was some specificity in the effects of HKMTi-1-005: accessibility at chromatin associated with alternative histone marks was similar in control and HKMTi-1-005-treated samples, in some instances even decreasing in accessibility after treatment (figure 3.28, appendix figures A.6 and A.7). The exception to this were H3K4me1- and H3K36me3-associated chromatin regions, which demonstrated large accessibility increases in most samples (figure 3.28, appendix figures A.8 and A.9). Changes at these chromatin domains are investigated in further downstream analysis (section 3.3.6).



(a)



(b)

Figure 3.25 **fGSEA on ATACseq data.** fGSEA results for the ID8 *in vivo* sample alongside the PEO1, PEO4, BRCA1 ID8, BRCA2 ID8 and F3 ID8 cell lines, comparing ATACseq data in control versus HKMTi-1-005-treated conditions. (a) fGSEA results for a geneset defined by Casciello et al. [107], based on genes that are differentially expressed following dual inhibition of both EZH2 and G9a, but not when either enzyme is inhibited individually. (b) fGSEA results for EZH2 genesets based on genes that are up- (green) or downregulated (red) when EZH2 is knocked down. . = p-adj < 0.1, * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001.

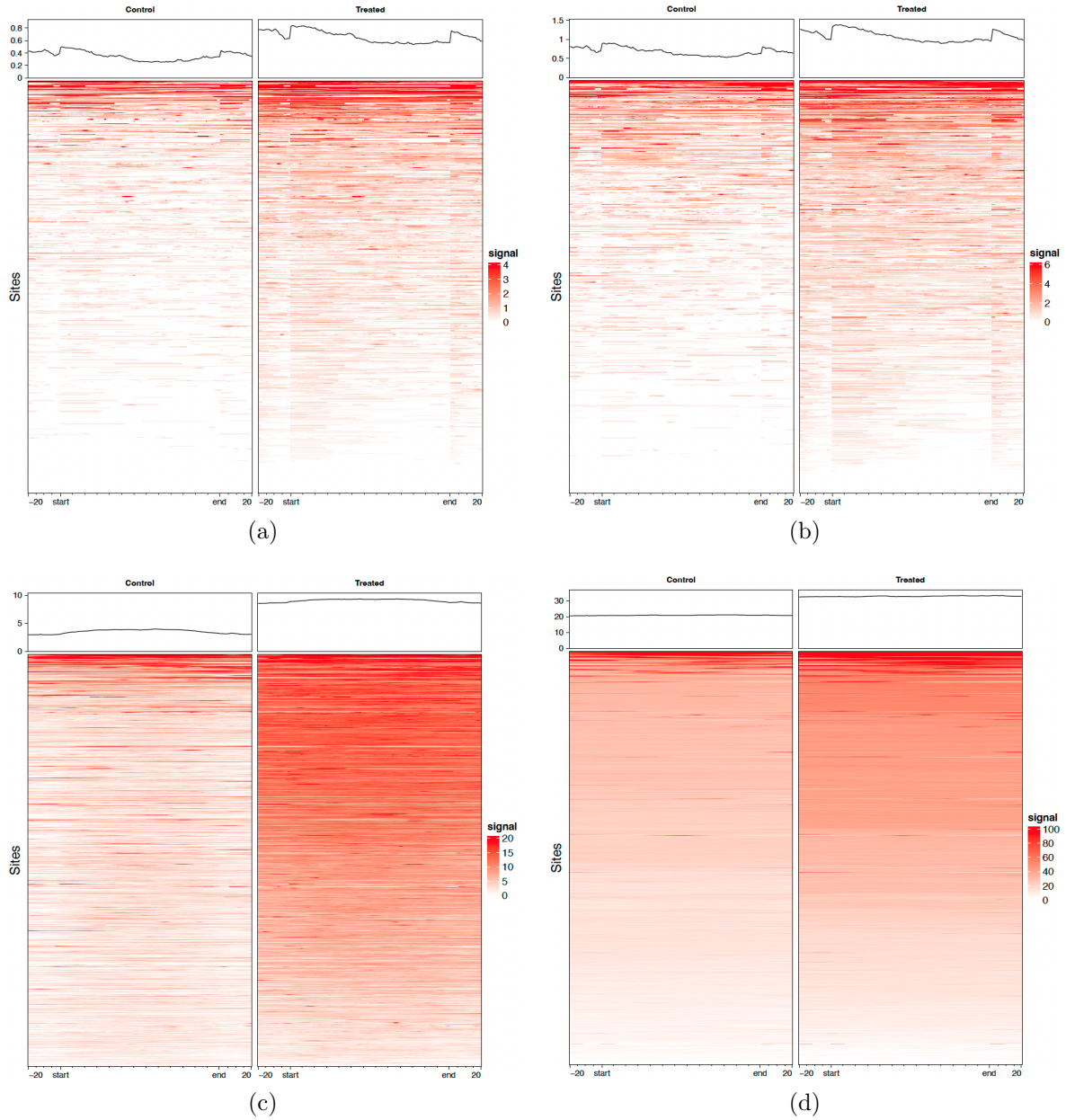
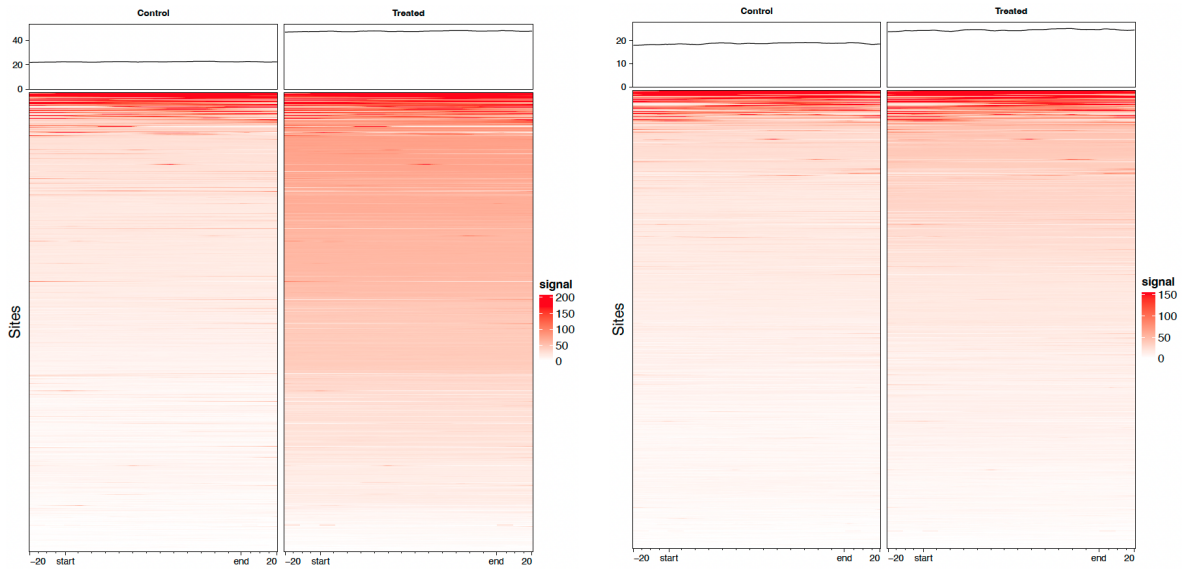
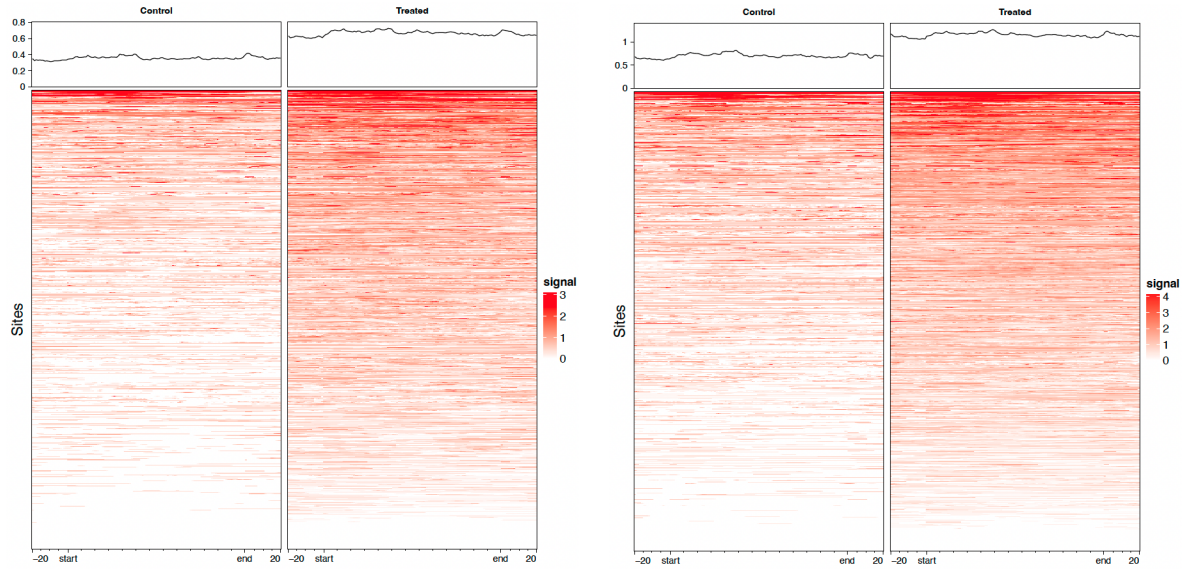


Figure 3.26 **H3K27me3 accessibility heatmaps.** Continued overleaf.



(e) (f)

Figure 3.26 **H3K27me3 accessibility heatmaps.** Heatmaps generated using DiffBind [390], displaying control versus HKMTi-1-005-treated chromatin accessibility at DNA associated with H3K27me3 histone marks. Samples include the (a) PEO1 cell line, (b) PEO4 cell line, (c) ID8 *in vivo* sample, (d) BRCA1 ID8 cell line, (e) BRCA2 ID8 cell line and (f) F3 ID8 cell line. Sites shown on the y-axis represent distinct H3K27me3-associated regions. The x-axis labels the start and end point of each region, divided into 100 bins and with 20% up- and down-stream of the region plotted. Depth of colour indicates the normalised mean read counts in each condition at H3K27me3-associated regions (trace shown at the top of the plot represents the normalised mean read count across all sites combined).



(a) (b)

Figure 3.27 **H3K9me3 accessibility heatmaps.** Continued overleaf.

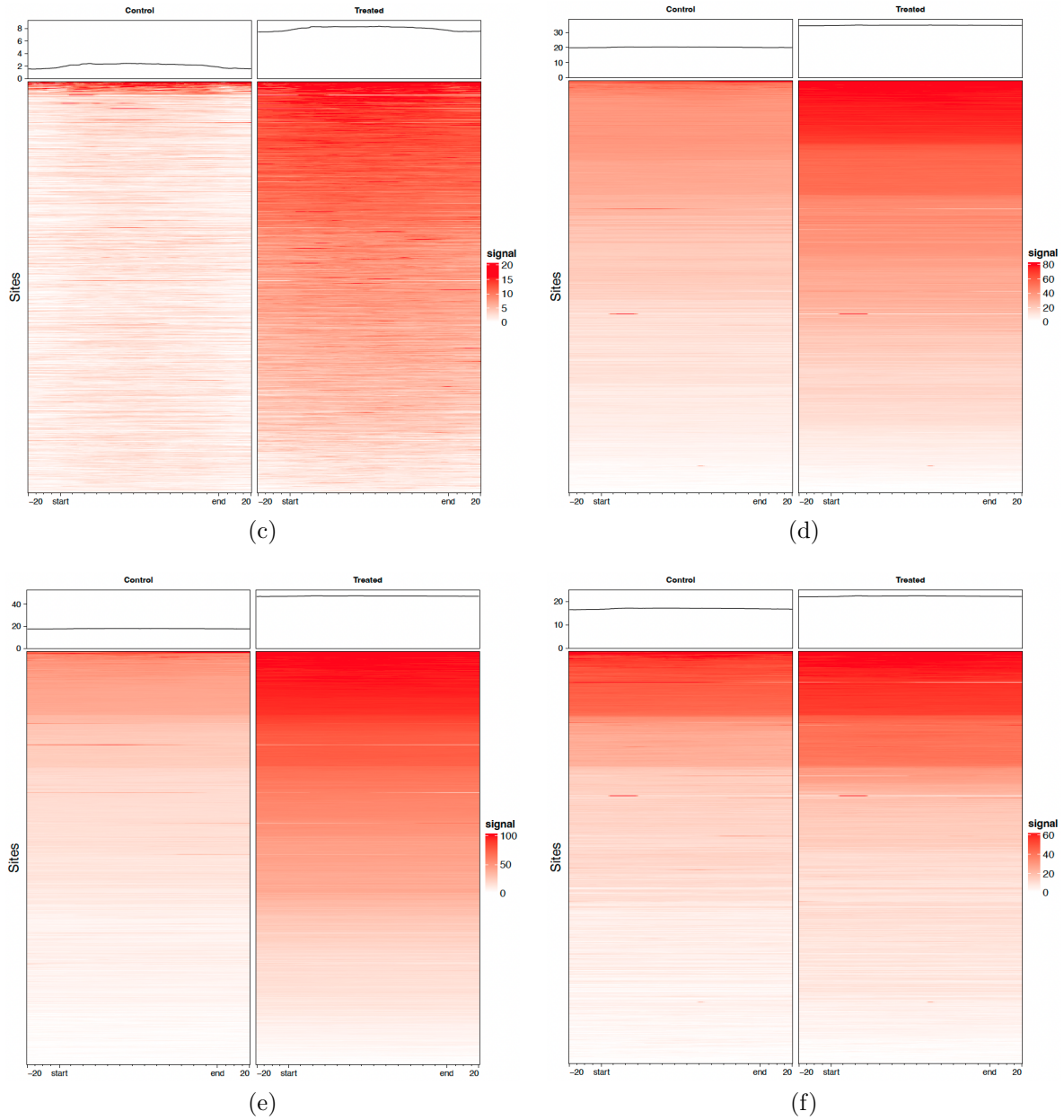


Figure 3.27 **H3K9me3 accessibility heatmaps.** Heatmaps generated using DiffBind [390], displaying control versus HKMTi-1-005-treated chromatin accessibility at DNA associated with H3K9me3 histone marks. Samples include the (a) PEO1 cell line, (b) PEO4 cell line, (c) ID8 *in vivo* sample, (d) BRCA1 ID8 cell line, (e) BRCA2 ID8 cell line and (f) F3 ID8 cell line.

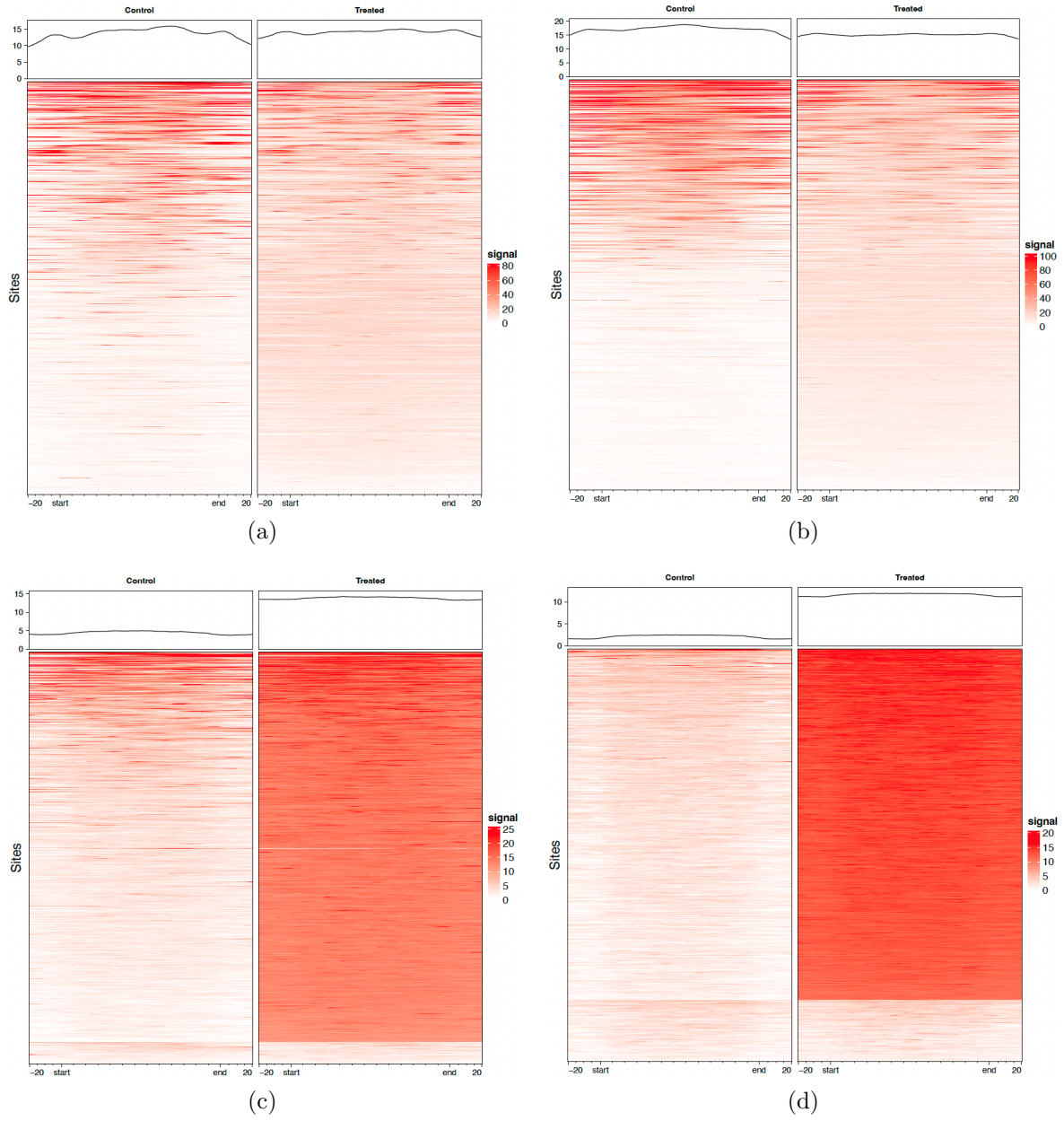


Figure 3.28 **H3K9Ac, H3K27Ac, H3K4me1 and H3K36me3 accessibility heatmaps.** Heatmaps generated using DiffBind [390], displaying control versus HKMTi-1-005-treated chromatin accessibility in the ID8 *in vivo* sample at DNA associated with (a) H3K9ac histone marks, (b) H3K27ac histone marks, (c) H3K4me1 histone marks and (d) H3K36me3 histone marks.

3.3.5 Characterisation of differentially accessible chromatin

Having established EZH2i- and G9ai-like effects on chromatin, I looked to characterise the genomic location of DACs more broadly. DACs were annotated using the ChIPseeker package [391], both in terms of genomic feature labels (e.g. 5'UTR, 3'UTR, intron, exon) and distance to the TSS for genes (e.g. $<$ 1kb, 1-3kb, 3-5kb). In most samples, there were a greater number of DACs near promoter regions (figures 3.29 and 3.30) and TSSs (appendix figures A.10 and A.11) than observed for background chromatin peaks. This was least evident for the ID8 *in vivo* and the BRCA2 ID8 *in vitro* sample, where the proportion of DAC peaks at promoters/TSSs was similar to or less than in background peaks. For these two samples there was instead a large proportion of DACs found >10kb or >100kb from TSSs, and in distal intergenic regions. Indeed, all samples demonstrated more than 15% of DACs in distal intergenic regions, indicating that HKMTi-1-005 treatment effects are not constrained to promoter regions.

To accompany annotation of DAC features, I performed an enrichment analysis of changes in accessibility at transcription factor motifs across the genome. The HOMER findMotifsGenome.pl tool [392] was used to achieve this, with analysis of chromatin changes obtained via both ATACseq methods (DAC and subtraction approaches, see figure 3.15). The former approach includes only accessibility changes below the threshold of p-adj $<$ 0.05 (with an additional absolute log2FC $>$ 1 threshold for the ID8 *in vivo* sample), whilst the latter selects unique HKMTi-1-005 treatment peaks but allows changes of any magnitude.

The results of both analyses suggested changes in many AP1 TF family motifs, including those of the Fos, Atf and Jun sub-families (figure 3.31, appendix figure A.12). AP1 family TFs are basic leucine zipper TFs that form homo- and heterodimers and most often promote activation of transcription, with roles in regulating apoptosis, immune responses and proliferation amongst other processes

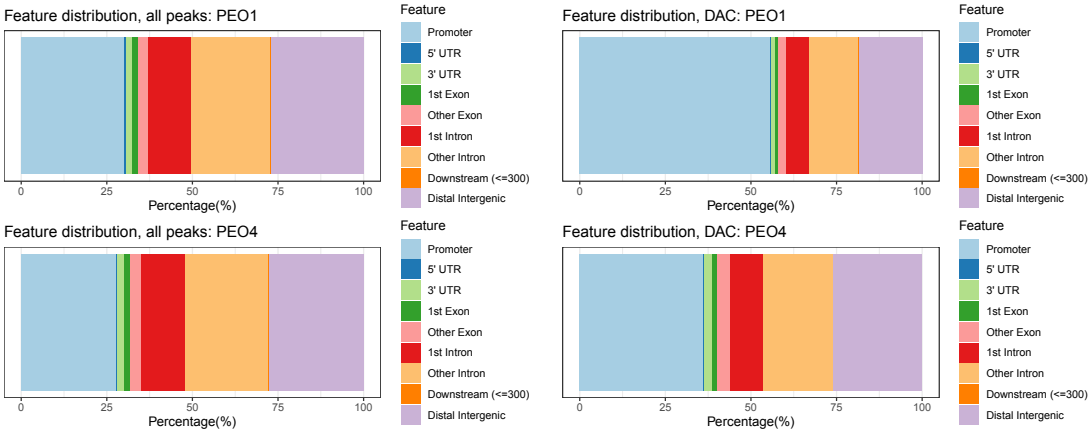


Figure 3.29 **Accessibility changes at genomic features, HGSC cell lines.** Plot visualising the distribution of all peaks of accessibility (left), or differentially accessible chromatin regions (DACs) (right), compared to labelled genomic features. ATACseq data derived from PEO1 cell line (top) and PEO4 cell lines (bottom).

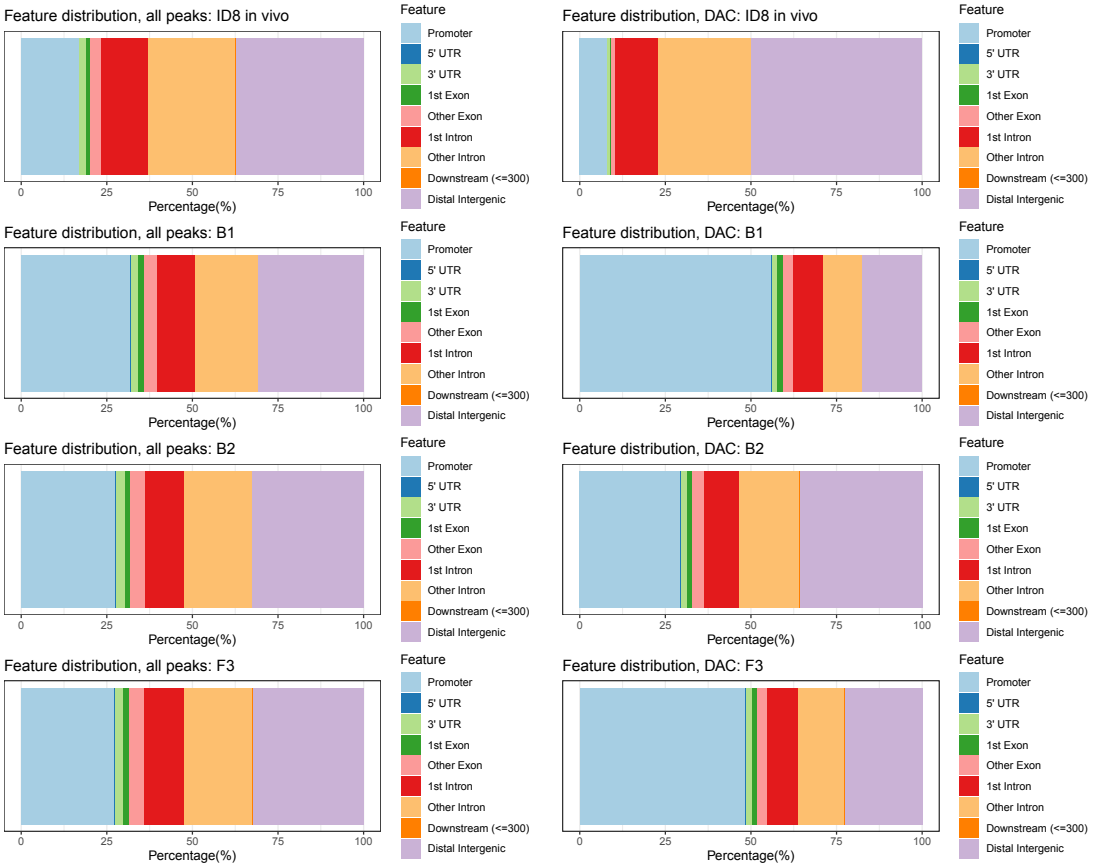


Figure 3.30 **Accessibility changes at genomic features, ID8 *in vivo* and *in vitro* samples.** Plot visualising the distribution of all peaks of accessibility (left), or differentially accessible chromatin regions (DACs) (right), compared to labelled genomic features. ATACseq data derived from ID8 *in vivo* sample (top), BRCA1 (B1, middle), BRCA2 (B2, middle) and F3 ID8 (bottom) cell line models of HGSC.

[429]. Alongside AP1 family TFs, a consistent change in motif accessibility for the architectural protein CTCF was observed across samples. There was considerable overlap in the motifs with altered accessibility across all six samples, with 24 and 47 motifs consistently altered as determined by the DAC and subtraction approaches respectively (appendix table A.5). The data suggest potential TFs that may contribute to the consistent transcriptional effects downstream of HKMTi-1-005 treatment, including those such as apoptotic and immune pathway changes, and support concordant changes in accessibility across different HGSC *in vitro* and *in vivo* models.

3.3.6 Enhancers and bivalent chromatin

Both H3K27me3 and H3K9me3 are found at bivalently marked chromatin, in combination with H3K4me1/H3K4me3 and H3K4me3/H3K36me3 respectively (see 1.4.2) [95–97, 102]. Bivalent chromatin is often found at "poised" enhancers and promoters [95–97, 102]. Additionally, H3K27me3 or H3K9me3 can be found at repressed enhancers [430]. As described, a large proportion of DACs occur at non-promoter regions, including at regions >100kb from TSSs. Furthermore, accessibility increases were observed at H3K4me1- and H3K36me3-marked regions, alongside the expected



Figure 3.31 **Motif enrichment analysis.** Motif analysis performed using HOMER [392], assessing enrichment of transcription factor motifs in DACs following treatment with HKMTi-1-005. Graphs show top 15 motifs enriched in the following samples from left to right, top to bottom: PEO1 cell line, PEO4 cell line, ID8 *in vivo* sample, BRCA1 (B1) ID8, BRCA2 (B2) and ID8 and F3 ID8 cell lines.

H3K27me3/H3K9me3 changes. Hence, I hypothesised that these accessibility changes might be explained by treatment effects at bivalently marked chromatin and enhancer elements.

I first investigated changes at enhancer loci. The Fantom5 database [406] was used to obtain loci labelled as enhancers in hg19/mm10 assemblies, with subsequent liftover to hg38/mm39 coordinates. Accessibility over enhancers in control versus HKMTi-1-005-treated samples was plotted using DiffBind [390]. Enhancers were more accessible in all samples, although to a variable degree, supporting the hypothesis that HKMTi-1-005 may exert some of its chromatin effects at these regulatory loci (figure 3.32). The motif analyses described further support this hypothesis, with the consistent enrichment of motifs for TFs such as CTCF, Fra1 and Fra2 (figure 3.31). CTCF is a crucial protein for establishing and regulating enhancer-promoter contacts [431], whilst Fra1/Fra2 have been shown to regulate broad transcriptional changes via enhancer binding and creation of "chromatin hubs" [432].

Bivalent chromatin was also examined. H3K9me3-H3K36me3 sites showed a marked increase in accessibility across all samples (figure 3.33, appendix figure A.14), whilst H3K9me3-H3K4me3 regions were less consistently more accessible and were represented by a low number of sites (appendix figure A.15). Increased accessibility was also observed across all H3K27me3-H3K4me1 bivalent regions (figure 3.33, appendix figure A.14), with a trend towards increased accessibility also observed at H3K27me3-H3K4me3 chromatin (figure 3.33, appendix figure A.14). Taken together, the results suggest that at least a subset of intergenic and H3K4me1/H3K36me3-associated DACs following HKMTi-1-005 treatment occur at enhancer regions and bivalently marked chromatin.

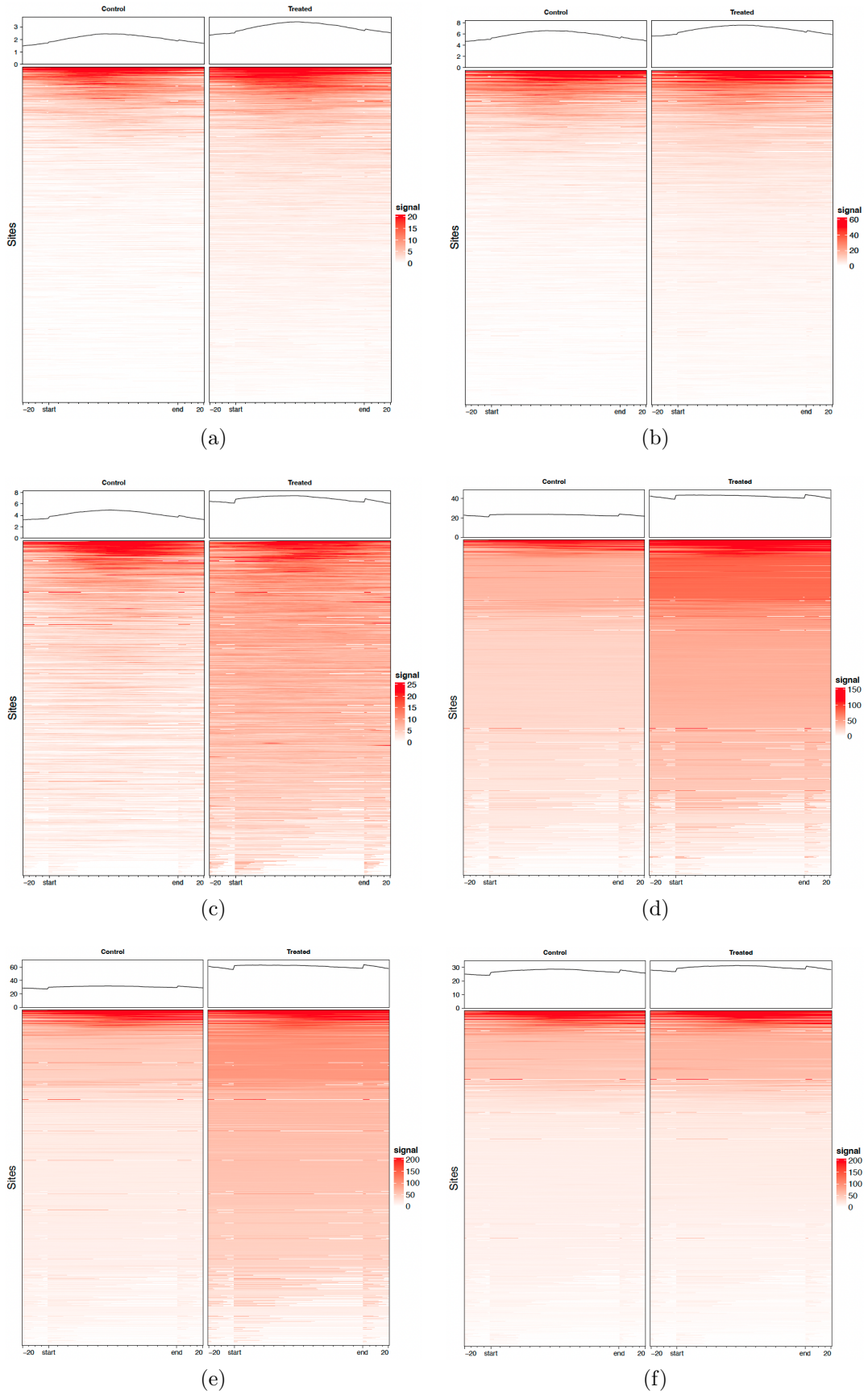


Figure 3.32 **Enhancer accessibility heatmaps.** Heatmaps generated using DiffBind [390], displaying control versus HKMTi-1-005-treated chromatin accessibility at DNA associated with enhancer regulatory elements. Samples include the (a) PEO1 cell line, (b) PEO4 cell line, (c) ID8 *in vivo* sample, (d) BRCA1 ID8 cell line, (e) BRCA2 ID8 cell line and (f) F3 ID8 cell line. Enhancers defined using the Fantom5 database [406].

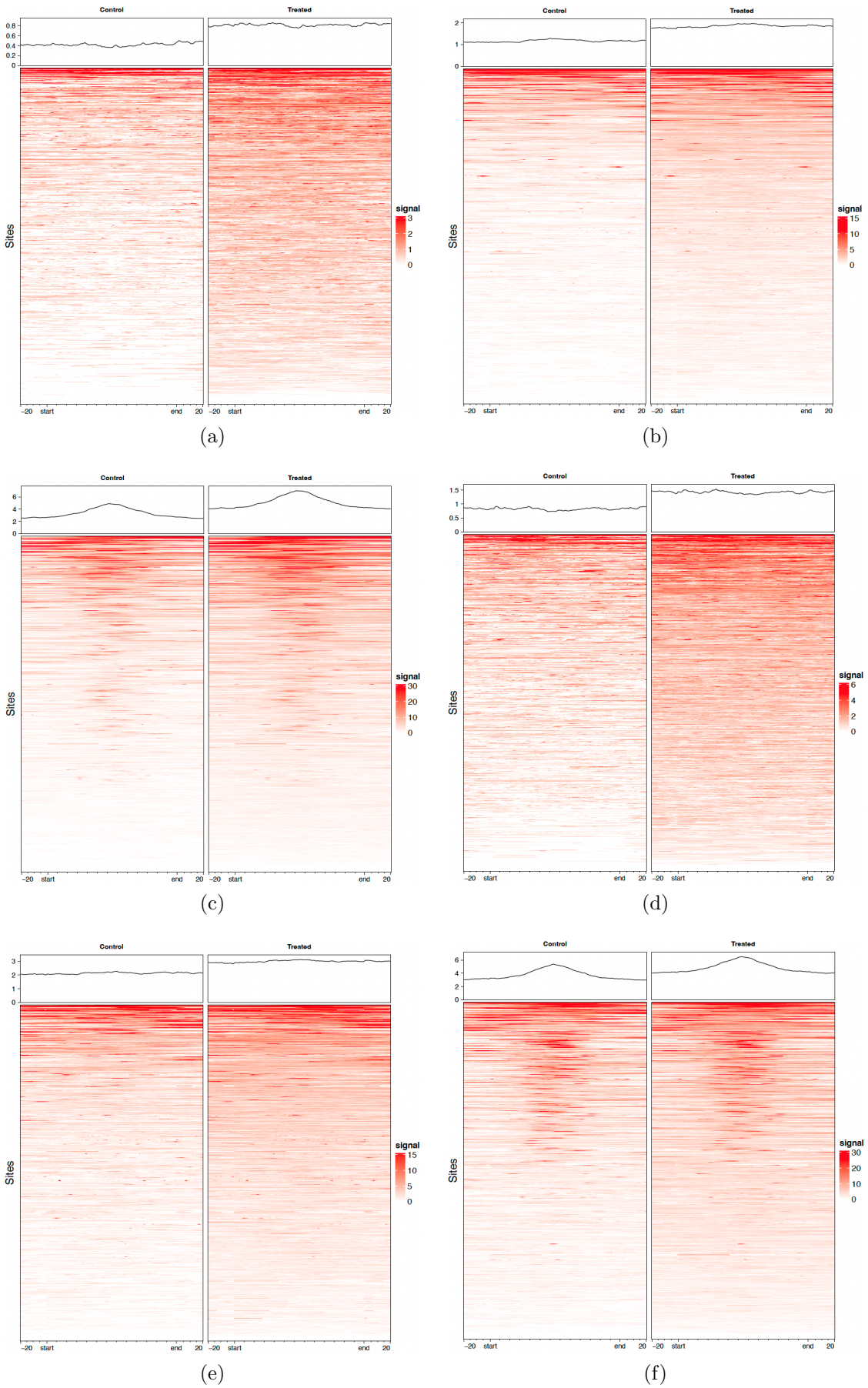


Figure 3.33 **Bivalent region accessibility heatmaps.** Continued overleaf.

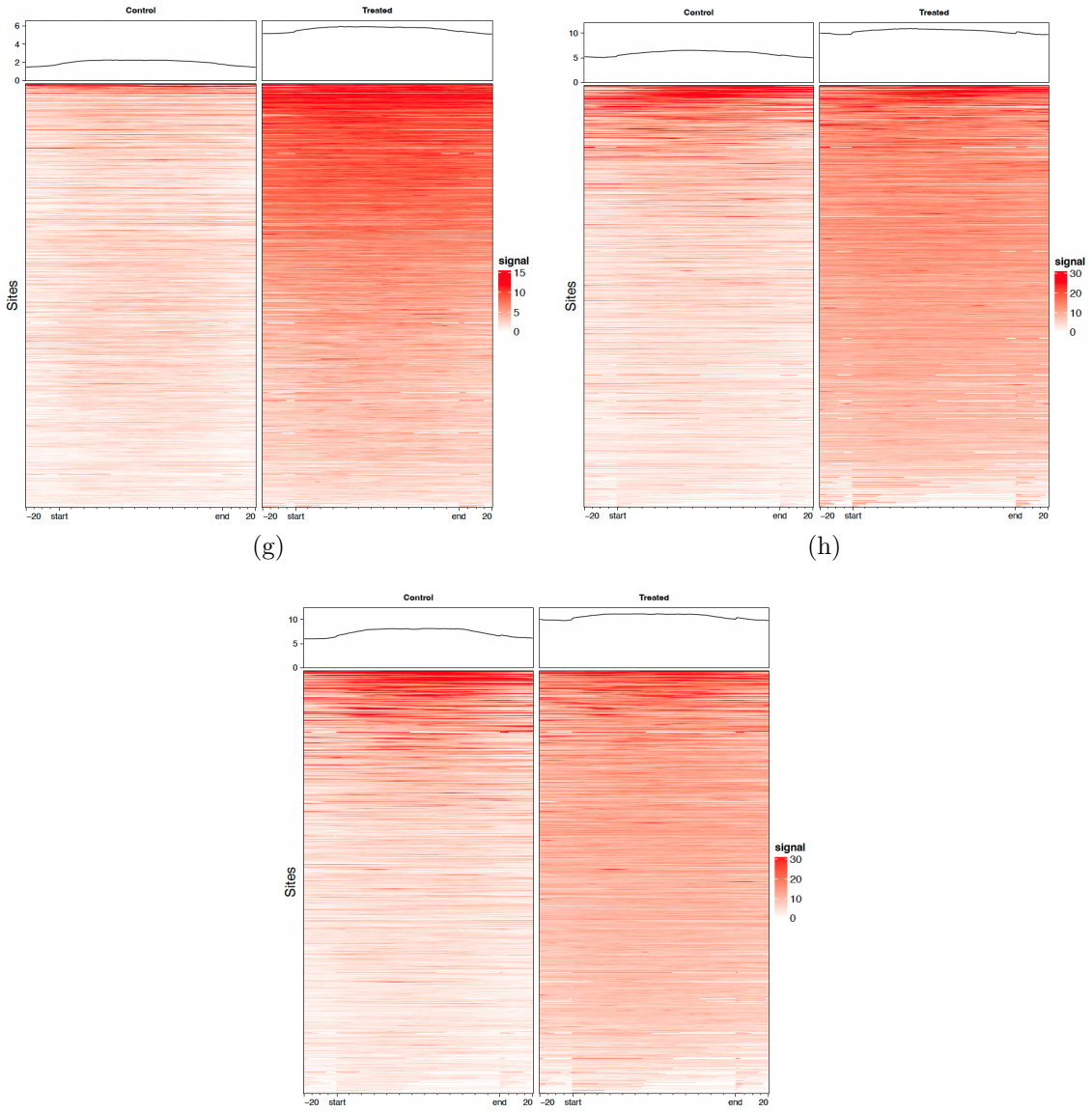


Figure 3.33 **Bivalent region accessibility heatmaps.** Heatmaps generated using DiffBind [390], displaying control versus HKMTi-1-005-treated chromatin accessibility at DNA associated with bivalent H3K9me3-H3K36me3 marks, bivalent H3K27me3-H3K4me1 marks and bivalent H3K27me3-H3K4me3 marks in the following samples (left to right, top to bottom): the PEO1 line (a-c) and the PEO4 line (d-f) and the ID8 *in vivo* model (g-i). Bivalent regions identified using overlap of coordinates defined for individual histone marks.

3.3.7 Relating changes in gene expression to altered chromatin accessibility following HKMTi-1-005 treatment

Increased or decreased chromatin accessibility should correlate with up- or down-regulation of gene expression respectively. An intersection test was performed to assess these relationships, comparing genes that were transcriptionally altered to genes that were in HKMTi-1-005-treated DACs or in unique HKMTi-1-005 treatment peaks in the ID8 *in vivo* and PEO1 and PEO4 *in vitro* samples. ID8 *in vitro* ATACseq data were not included due to the lack of a corresponding RNAseq dataset. The analysis confirmed a significant association between gene expression and chromatin accessibility changes via both ATACseq approaches in all samples (tables 3.1 and 3.2).

To assess similarities in ATACseq and RNAseq data further, changes at the geneset level were analysed. fGSEA using the GOBP and Hallmark databases demonstrated consistent changes in accessibility of genes involved in metabolic, immune and apoptotic pathways (figure 3.34 and appendix figure A.13). These results must be interpreted with care due to multiple DACs existing at the same gene locus, yet support increased accessibility across similar pathways to those observed in RNAseq data. Hence, the overall effects of HKMTi-1-005 at EZH2 and G9a target genes, genomic elements and chromatin domains support EZH2i- and G9ai-like activity of the compound.

Intersections	Degree	Observed.Overlap	Expected.Overlap	Fold enrichment	P.value
P1_upregulated_DEGs & P1_SUBTRACT	2	1654	661.2572588	2.50129579	0
P4_upregulated_DEGs & P4_SUBTRACT	2	409	160.2195065	2.55274785	5.357E-135
P1_upregulated_DEGs & P1_DAC_UP	2	288	61.03865755	4.71832133	1.527E-110
P4_upregulated_DEGs & P4_DAC_UP	2	80	15.21204654	5.25898996	3.5302E-34
ID8_upregulated_DEGs & ID8_DAC_UP	2	461	301.77224	1.52764217	3.6083E-26
ID8_upregulated_DEGs & ID8_SUBTRACT	2	868	746.2121752	1.16320804	1.2765E-18

Table 3.1 **ATACseq versus RNAseq comparison: upregulation.** Statistical testing of the overlap between genes with increased accessibility and those with upregulated expression in the PEO1 (P1) cell line, the PEO4 (P4) cell line and the ID8 *in vivo* model. ATACseq results include genes associated unique HKMTi-1-005 treatment peaks of accessibility (labelled "SUBTRACT", subtraction method) and genes associated with upregulated DACs (labelled "DAC_UP", DAC method). RNAseq results include transcriptionally upregulated genes with p-adj < 0.05 and log2FC > 0 (labelled "upregulated_DEGs"). Test performed using the SuperExactTest package.

Intersections	Degree	Observed.Overlap	Expected.Overlap	Fold enrichment	P.value
ID8_downregulated_DEGs & ID8_DAC_DN	2	536	283.6790475	1.88945925	8.5078E-97
P1_downregulated_DEGs & P1_DAC_DN	2	116	32.04616777	3.61977759	9.2634E-33
P4_downregulated_DEGs & P4_DAC_DN	2	39	15.93757125	2.44704788	3.2295E-07

Table 3.2 **ATACseq versus RNAseq comparison: downregulation.** Statistical testing of the overlap between genes with decreased accessibility and those with downregulated expression in the PEO1 (P1) cell line, the PEO4 (P4) cell line and the ID8 *in vivo* model. ATACseq results include genes associated downregulated DACs (labelled "DAC_DN", DAC method). RNAseq results include transcriptionally downregulated genes with p-adj < 0.05 and log2FC < 0 (labelled "downregulated_DEGs"). Test performed using the SuperExactTest package.

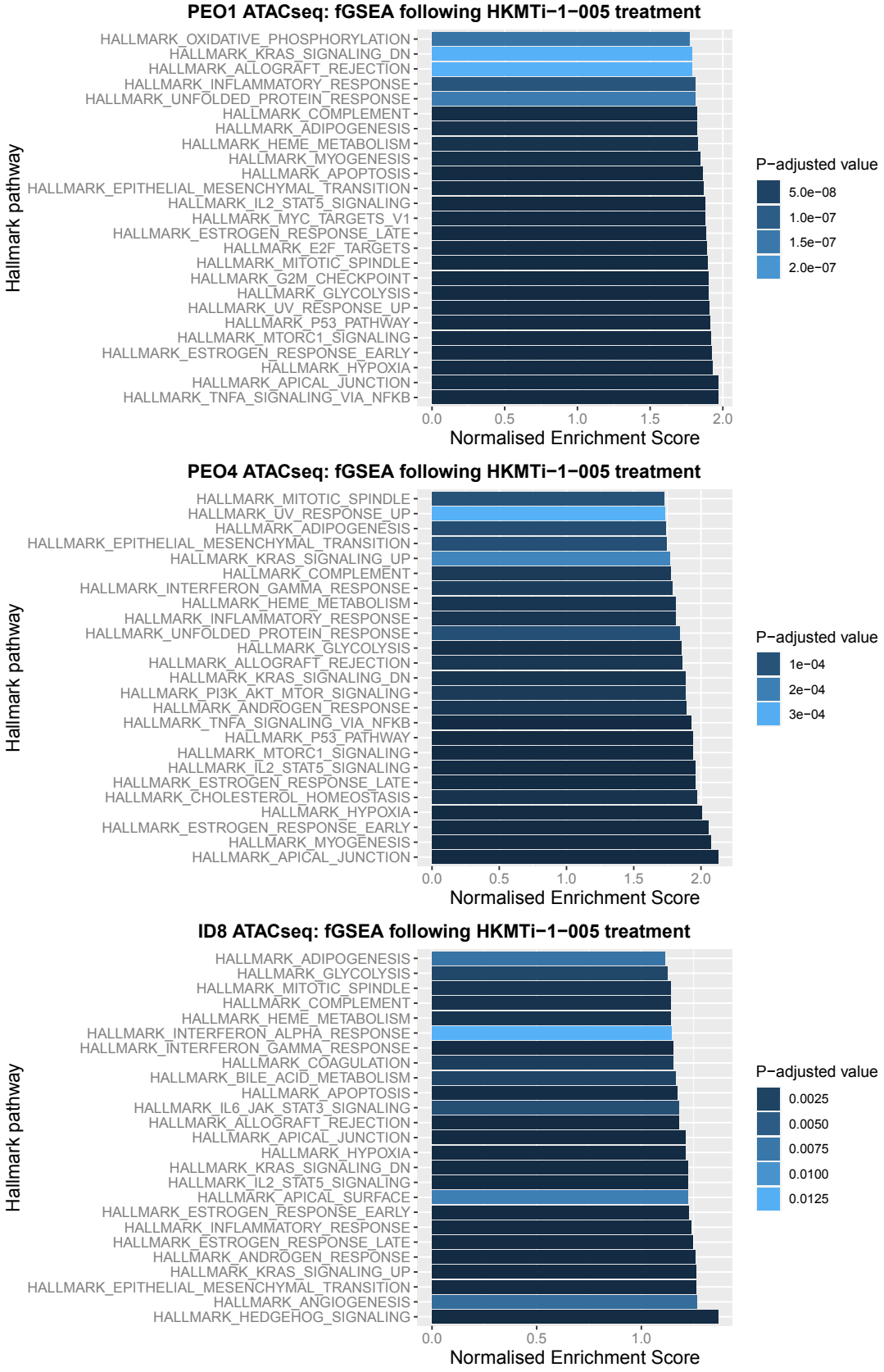


Figure 3.34 **ATACseq fGSEA.** Continued overleaf.

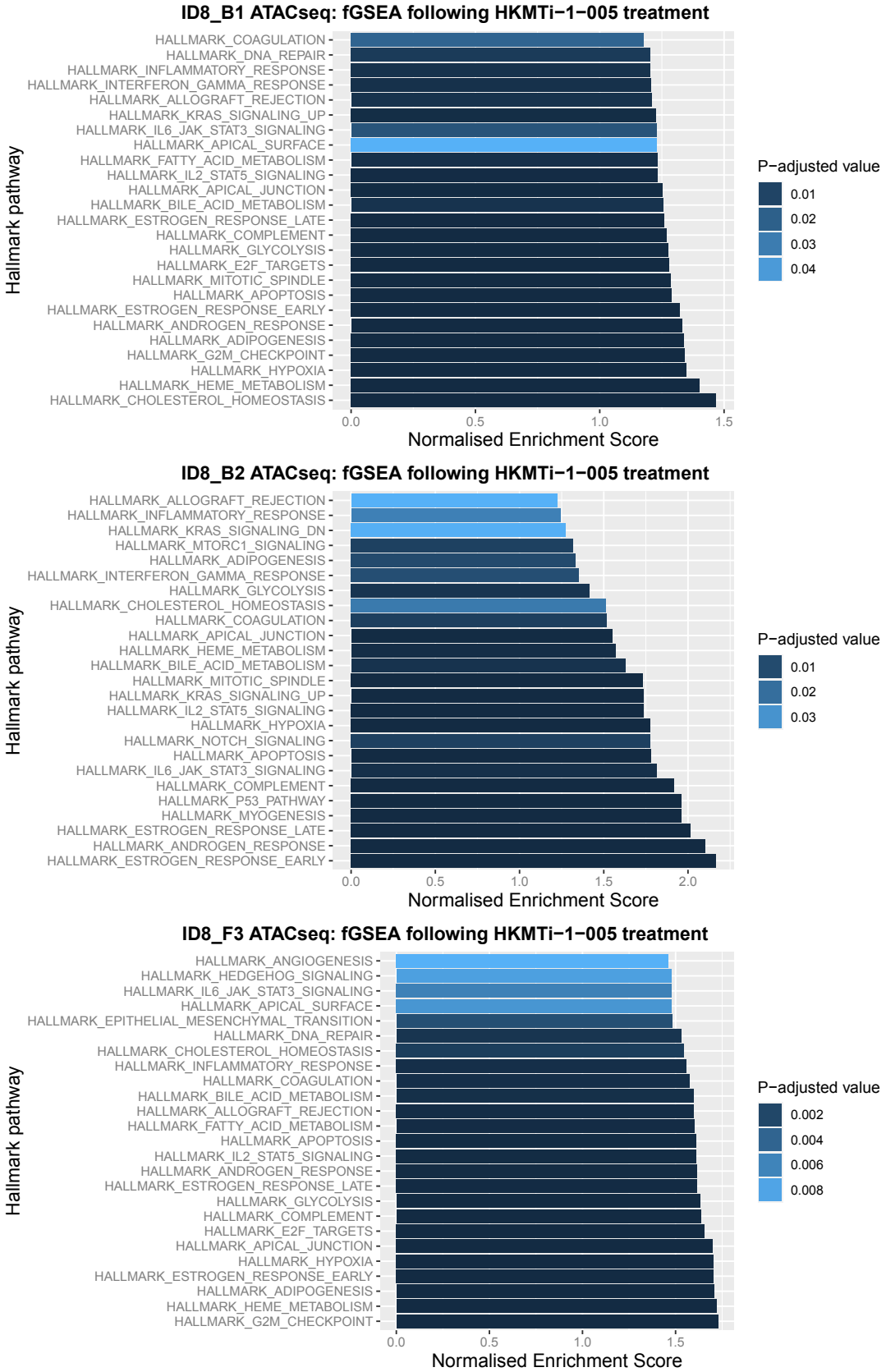


Figure 3.34 **ATACseq fGSEA.** fGSEA results showing the top 25 enriched Hallmark pathways, based on DACs following HKMTi-1-005 treatment in the (a) PEO1 cell line, PEO4 cell line and ID8 *in vivo*, and (b) *in vitro* BRCA1 (B1) ID8, *in vitro* BRCA2 (B2) ID8 and *in vitro* F3 ID8 cell lines. Lighter bar colour represents smaller p-adj (or greater -log10(p-adj)). DACs defined as p-adj < 0.05, absolute log2FC > 0, with an additional threshold of absolute log2FC of 1 for the ID8 *in vivo* sample.

3.4 Investigating HKMTi-1-005 in comparison to single EZH2 and G9a inhibitors

3.4.1 HKMTi-1-005 displays a similar sensitivity profile to single EZH2 or G9a inhibitors

The Genomics of Drug Sensitivity in Cancer 2 (GDSC2) is a publicly available dataset describing the sensitivity of over 1000 pan-cancer cell lines to over 250 compounds, with the aim of identifying genomic markers of sensitivity [425]. HKMTi-1-005 was submitted to the GDSC2 for screening against the cell line panel, and for comparison to single G9a and EZH2 inhibitors as well as classic chemotherapies.

Raw data describing sensitivity of 969 cell lines to 295 compounds were downloaded, as measured by IC_{50} values. Data included sensitivity to HKMTi-1-005, sensitivity to two single G9ai compounds (UNC0638 and A-366) and two screens describing sensitivity to a single EZH2i (GSK343 in two separate assays). The UNC0638 compound was selected over A-366 as the single G9ai for comparison to HKMTi-1-005, as this inhibitor is based on the same BIX-01294 compound from which our compound was derived, with a very similar chemical structure [148]. The GSK343 dataset with the greatest overlap with HKMTi-1-005 data in the cell lines tested was included for an EZH2i comparison (DRUG_ID 2037). Comparison to the DNA damage agent cisplatin was included in addition, as a first-line chemotherapy used in HGSC.

Sensitivity to the four compounds of interest – HKMTi-1-005, GSK343, UNC0638 and cisplatin – was first compared across all matched cell lines in the panel (figure 3.35). These data indicated similar overall sensitivity profiles, with median natural log IC_{50} ($\ln(IC_{50})$) values of between 3 and 4.5 µM across compounds. The median $\ln(IC_{50})$ for HKMTi-1-005 was lower than for single EZH2i or G9ai compounds, suggesting greater growth inhibitory effect was achieved with the dual targeting approach (HKMTi-1-005 95% confidence interval (CI) [3.50,3.88], GSK343 95% CI [4.35,4.64], UNC0638 95% CI [4.29,4.54]). Additionally, the correlation between sensitivity to each compound was assessed (figure 3.36). These data demonstrated a significantly stronger correlation between HKMTi-1-005 and each of the single EZH2i and G9ai compounds as compared to cisplatin, supporting its EZH2i- and G9ai-like behaviour (HKMTi-1-005-GSK343 95% CI [0.73,0.79], HKMTi-1-005-UNC0638 95% CI [0.63,0.72], HKMTi-1-005-cisplatin 95% CI [0.48,0.59]). Indeed, including all epigenetic drugs in the GDSC2 screen, as well as all PARP inhibitors and a panel of chemotherapeutic drugs, sensitivity to HKMTi-1-005 had the highest correlation with a single EZH2i (GSK343) and a single G9ai (UNC0638) (figure 3.37). Amongst all compounds in the GDSC2 screen of any modality, GSK343 and UNC0638 were in the top three correlates (figure 3.38).

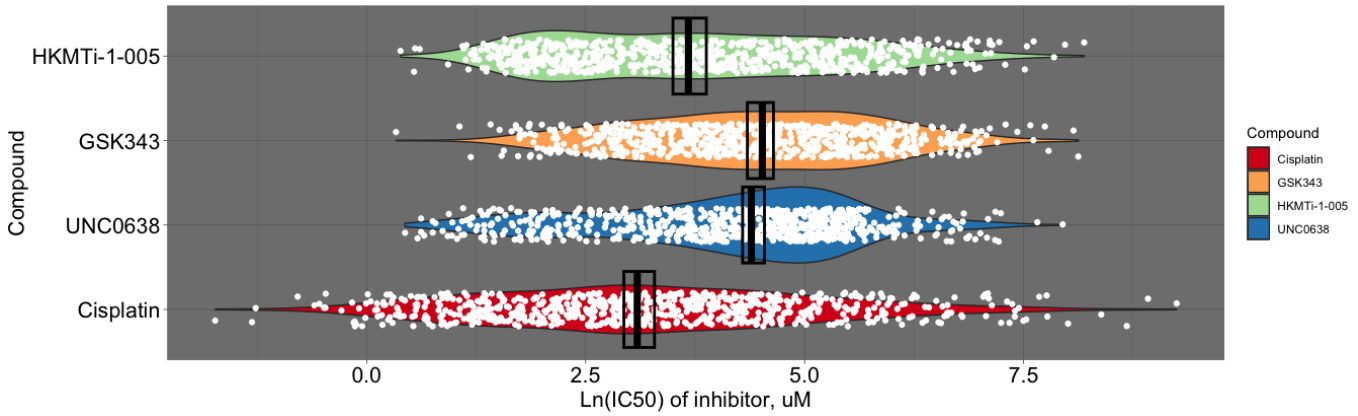


Figure 3.35 **Violin plots illustrating compound sensitivity across the GDSC2 cohort.** Compounds include HKMTi-1-005, GSK343 (EZH2i), UNC0638 (G9ai) and cisplatin (top to bottom), based on cell lines with sensitivity data for all four compounds and with a tumour type annotation (n=696). Black vertical lines represent the median IC_{50} value and 95% CIs for each compound across all cell lines. White dots represent IC_{50} values for each individual cell line. Ln = natural log, IC_{50} = half-maximal inhibitory concentration, uM = µM.

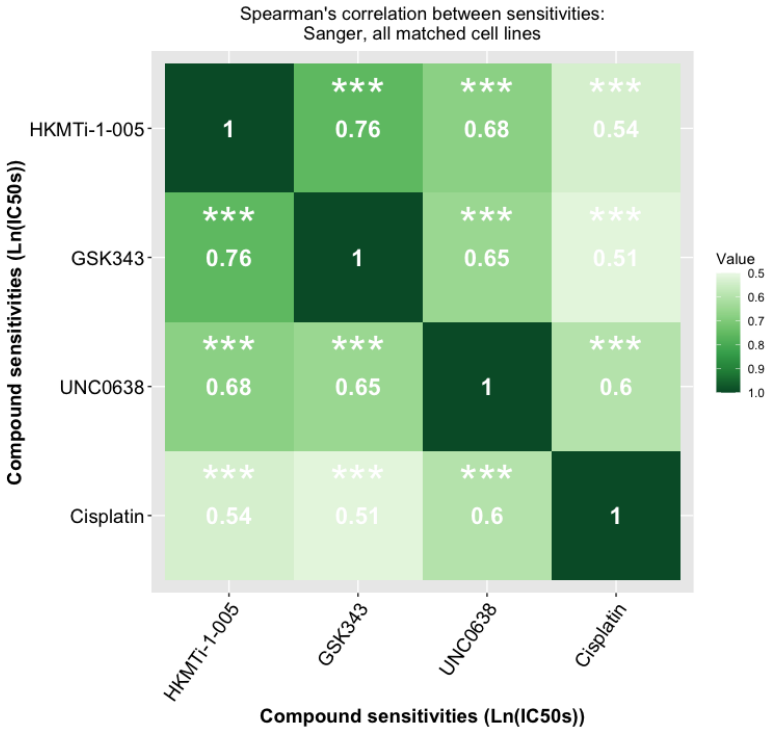


Figure 3.36 **Spearman's correlation between sensitivity to HKMTi-1-005, GSK343 (EZH2i), UNC0638 (G9ai) and cisplatin.** Correlation across 696 cell lines in the GDSC2 cell line panel with sensitivity data for each of the four compounds were included, of any labelled tumour type. Ln(IC_{50}) = natural log of half-maximal inhibitory concentration.

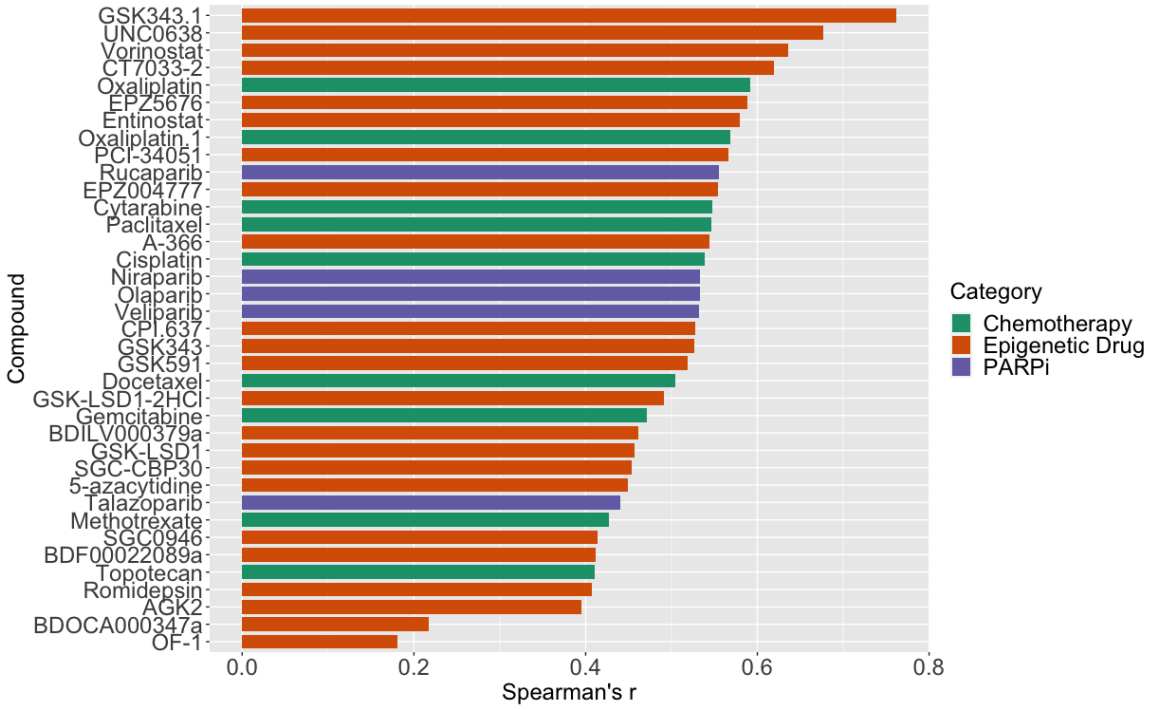


Figure 3.37 **Correlation between sensivity to HKMTi-1-005 and to epigenetic drugs.** Barplot showing the correlation between HKMTi-1-005 and all epigenetic drugs in the GDSC2 screen, as well as all PARPis and a panel of commonly used chemotherapeutic drugs. All cell lines in the GDSC2 panel of any labelled tumour type included in analysis, not just those with sensitivity data for HKMTi-1-005, GSK343, UNC0638 and cisplatin (n = 969 cell lines). Epigenetic drugs defined as those described to affect chromatin pathways. Drugs missing sensitivity values for > half the cell lines were excluded. Where duplicate drug screens existed, ".1" was used to distinguish screens. All p-values are < 0.001.

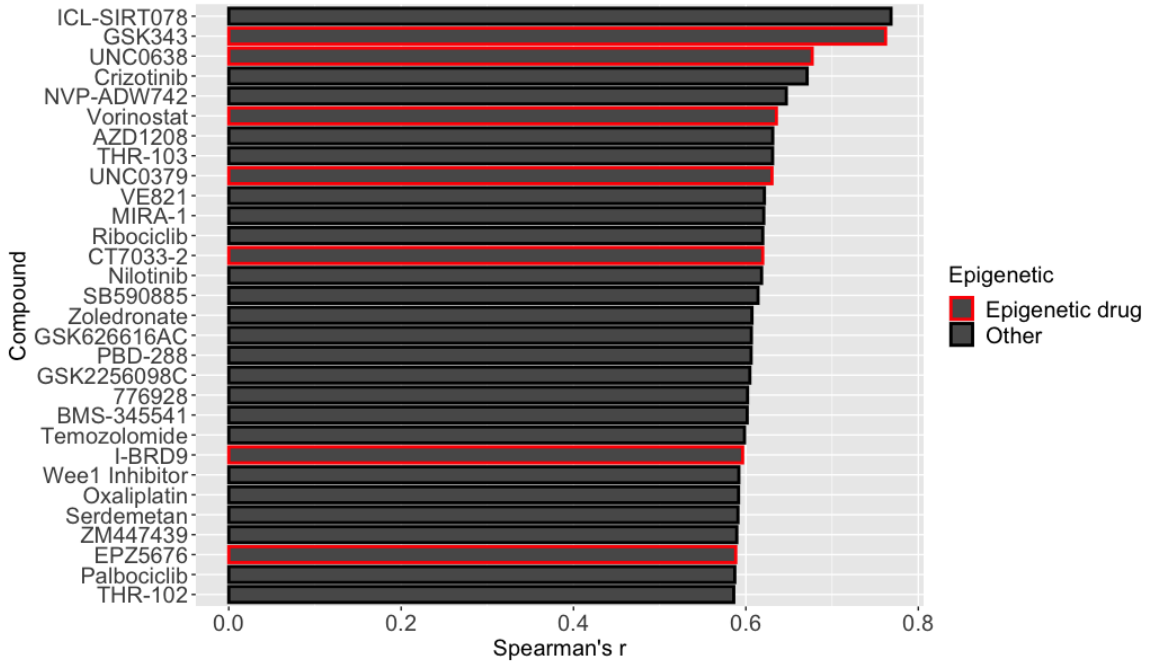


Figure 3.38 **Correlation between sensivity to HKMTi-1-005 and to alternative drugs.** Barplot showing the correlation between HKMTi-1-005 and the top 30 correlates from all drugs in the GDSC2 screen. All cell lines in the GDSC2 panel of any labelled tumour type included in analysis, not just those with sensitivity data for HKMTi-1-005, GSK343, UNC0638 and cisplatin (n = 969 cell lines). Drugs missing sensitivity values for > half the cell lines were excluded. Bars representing epigenetic drugs are labelled with a red outline. All p-values are < 0.001.

3.4.2 Haematological, sarcoma, melanoma and SCLC cell lines display heightened sensitivity to HKMTi-1-005

EZH2 inhibition is already approved for use in follicular lymphoma and epithelioid sarcoma [143, 144]. Additionally, haematological malignancies are proposed to be particularly sensitive to EZH2i [433] and research suggests the same may be true of G9ai in haematological malignancies such as myeloma [139] and leukaemia [434]. Consistent with this, the three haematological tumour types included in the screen (myeloma, leukaemia and lymphoma) were the most sensitive to each of the HKMTi-1-005, GSK343 and UNC0638 compounds (figure 3.39). Whilst haematological lines were still amongst the more sensitive lines to cisplatin, the effect was less marked in the cisplatin dataset as compared to the epigenetic inhibitors.

Other cell lines with heightened sensitivity to HKMTi-1-005 included those derived from Ewing's sarcoma, melanoma and lung SCLC. Considering Ewing's sarcoma lines, a heightened sensitivity was also observed for each of the other compounds, suggesting a more general sensitivity effect. However, melanoma lines were more sensitive to GSK343 than median sensitivity, but not to UNC0638 or cisplatin, suggesting a potential EZH2i-specific effect. In contrast, SCLC lines pointed to a potential G9ai-mediated sensitivity effect, with a skew towards increased sensitivity only to the UNC0638 compound but near median sensitivity to cisplatin and GSK343 sensitivity. Interestingly, each of these three tumour types displayed a markedly skewed distribution in their HKMTi-1-005 sensitivity, with a subgroup of highly sensitive cells and a tail of insensitive ones, that appeared distinct from sensitivity profiles to the other compounds (figure 3.40). This was particularly true of the SCLC cohort, suggesting that investigation into these lines may provide insight into predictors of HKMTi-1-005 sensitivity.

3.4.3 Aberrant *EZH2* or *G9a* is associated with increased sensitivity to HKMTi-1-005

Alterations in *EZH2* or *G9a* (*EHMT2*) have been demonstrated to contribute to hypersensitivity to EZH2i and G9ai respectively [64, 435]. Hence, the relationship between mutation, amplification or overexpression (OE) of *EZH2* or *EHMT2* and compound sensitivities was investigated in the cell line panel. Considering all cell lines, a mutation in *EZH2* was associated with increased sensitivity to each of the four compounds (p-value < 0.05), whilst a mutation in *EHMT2* related to increased sensitivity to HKMTi-1-005 and UNC0638 only (figure 3.41). No statistics were performed on cell lines with amplifications of *EZH2* or *EHMT2*, as these were very few; only three matched cell lines had an *EZH2* amplification and only one had an *EHMT2* amplification. Instead, overexpression of *EZH2* or *EHMT2* was considered, defined using z-score >1. Both *EZH2* OE and *EHMT2* OE were

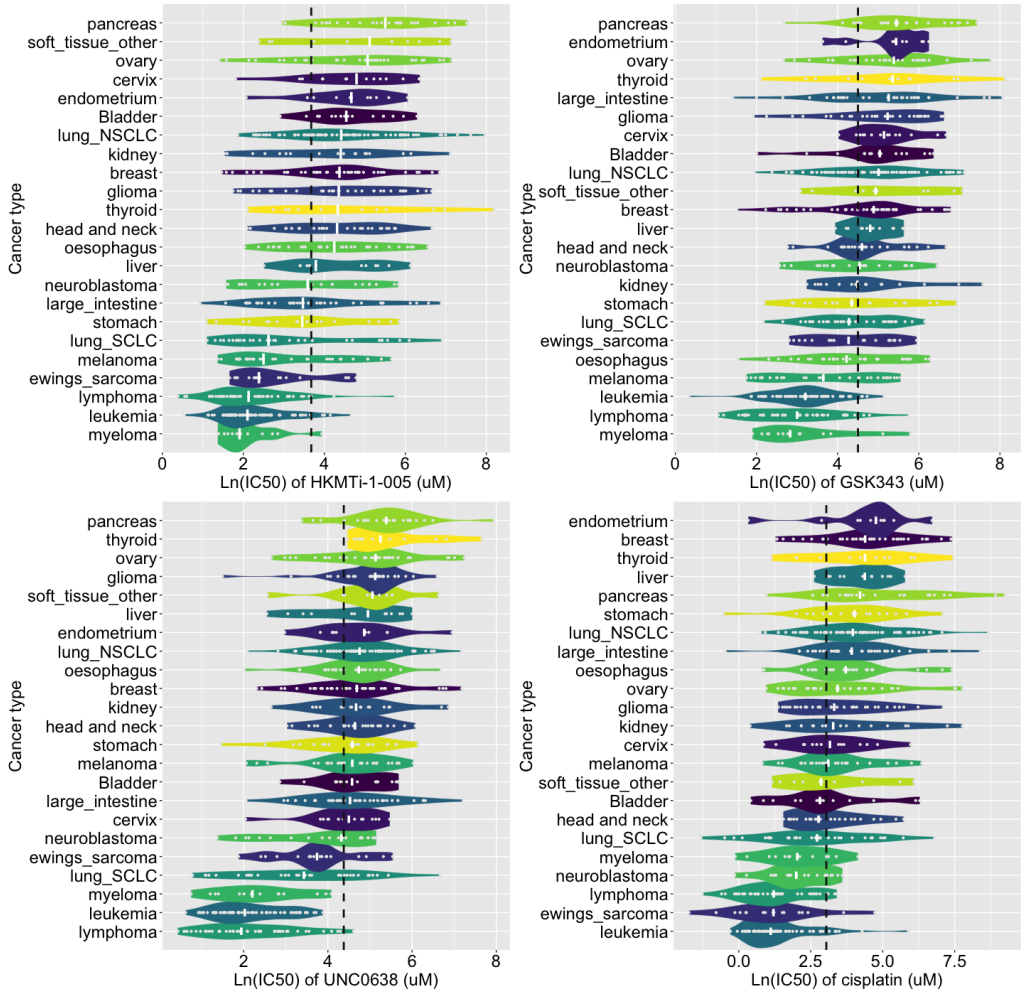


Figure 3.39 **Sensitivity to HKMTi-1-005, GSK343, UNC0638 or cisplatin: visualisation by tumour type.** Violin plot visualisation of sensitivity to HKMTi-1-005, GSK343 (EZH2i), UNC0638 (G9ai) and cisplatin (left to right, top to bottom) in the GDSC2 cohort by tumour type, based on cell lines with sensitivity data for all four compounds. Tumour types represented by fewer than 10 cell lines were not included in this plot to simplify visualisation. Vertical black dashed lines represent the median IC_{50} value for each compound across the cohort, whilst the white vertical lines represent median IC_{50} values by tumour type. White dots reflect IC_{50} values for individual cell lines. Ln(IC_{50}) = natural log of half-maximal inhibitory concentration, uM = µM.

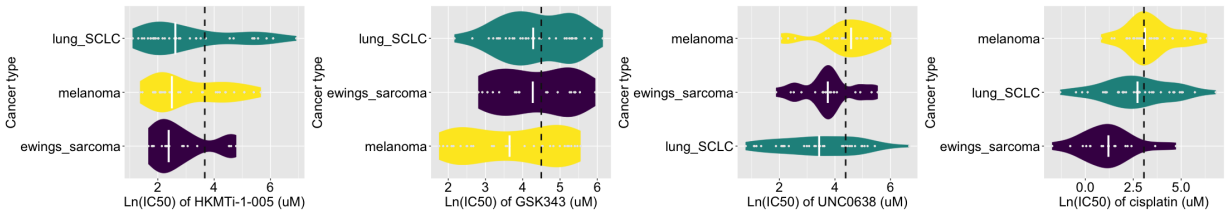
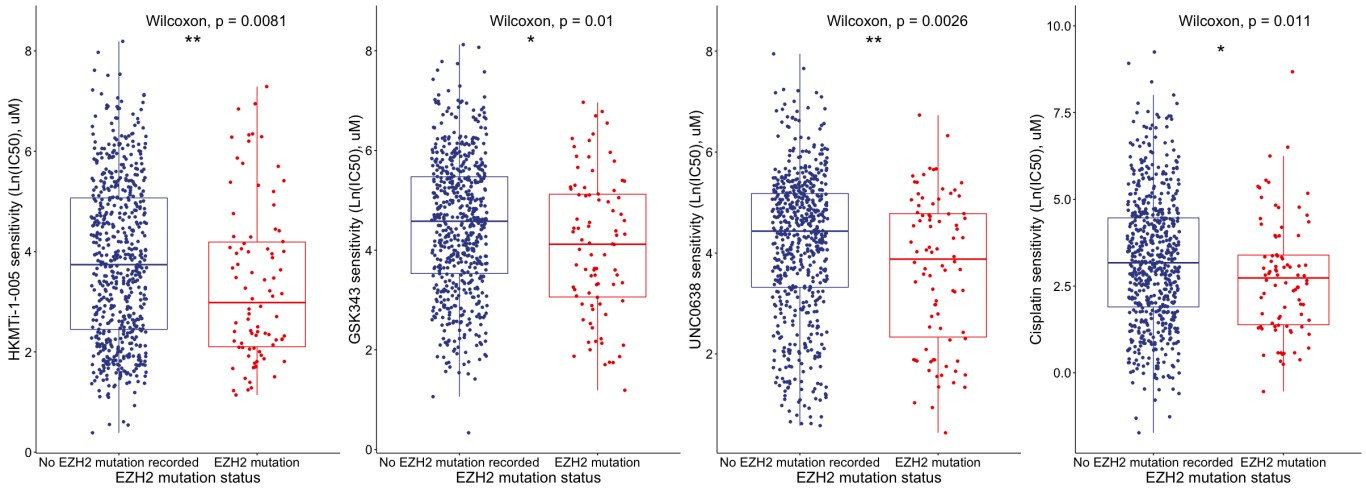
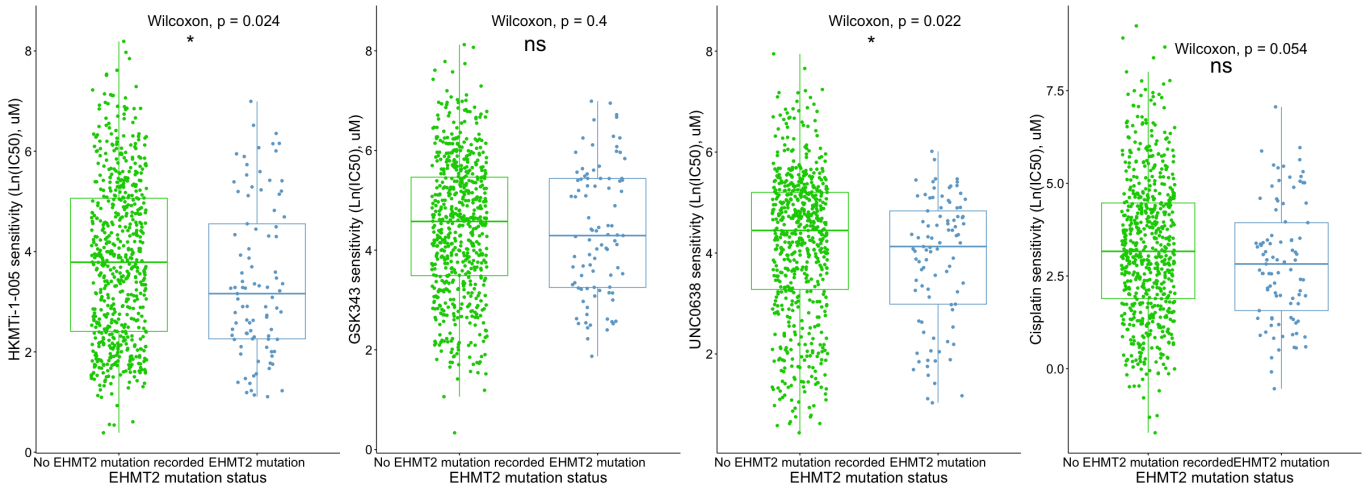


Figure 3.40 **Sensitivity to HKMTi-1-005, GSK343, UNC0638 or cisplatin: lung SCLC, melanoma and Ewing's sarcoma cell lines.** Violin plot visualisation of sensitivity to HKMTi-1-005, GSK343 (EZH2i), UNC0638 (G9ai) and cisplatin (left to right) in the GDSC2 cohort, considering only the lung SCLC, melanoma and Ewing's sarcoma cell lines. Vertical black dashed lines represent the median IC_{50} value for each compound across the whole cell line cohort including all tumour types, whilst the white vertical lines represent median IC_{50} values by tumour type. White dots reflect IC_{50} values for individual cell lines. Ln(IC_{50}) = natural log of half-maximal inhibitory concentration, uM = µM.



(a)



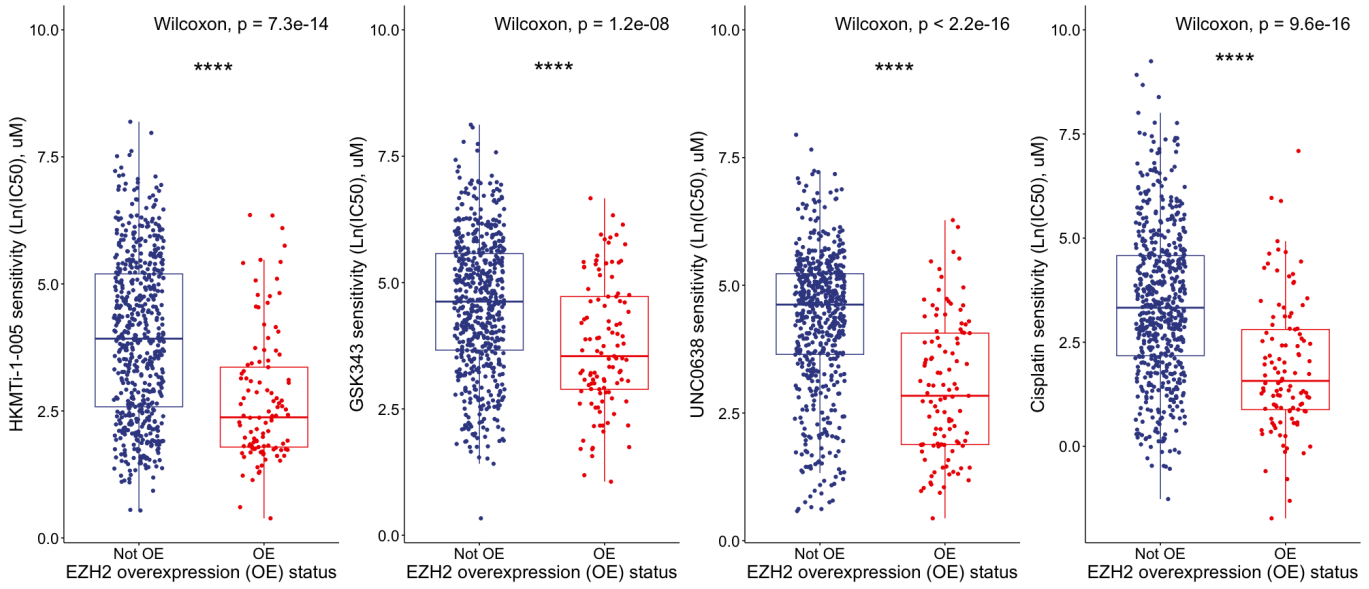
(b)

Figure 3.41 ***EZH2* or *G9a* mutations versus sensitivity.** Relationship between mutations in (a) *EZH2* or (b) *EHMT2* (encoding G9a) and sensitivity to HKMTi-1-005, GSK343 (EZH2i), UNC0638 (G9ai) and cisplatin. Cell lines included are those from the GDSC2 cell line panel with sensitivity data for each of the four compounds, derived from any tumour type (n=696). Mutation data were downloaded from COSMIC [407].

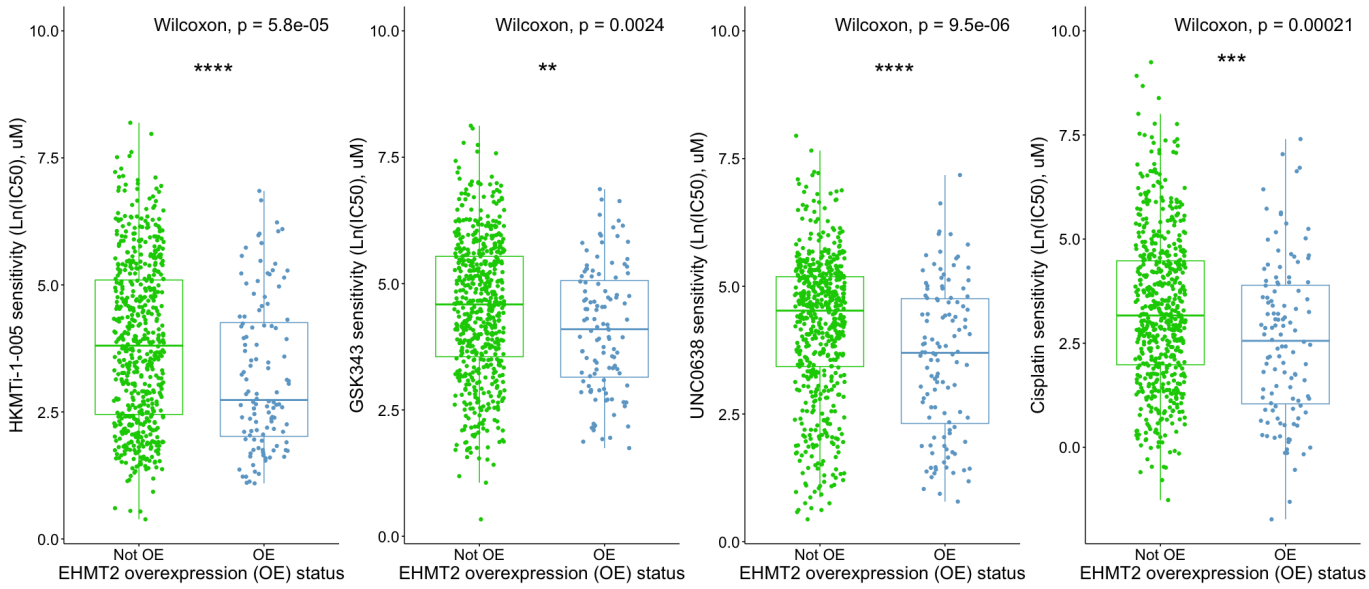
associated with greater sensitivity to every compound (figure 3.42).

Haematological lines and solid cell line cohorts were next analysed separately, given their disparate sensitivity profiles and biology. In the haematological cohort, there was no association between *EZH2/EHMT2* mutation nor OE status and sensitivity to any of the compounds (appendix figure A.16). Indeed, there was a trend towards the opposite relationship for OE of *EZH2* or *EHMT2*, which appeared to relate to lower sensitivity to HKMTi-1-005 ($p <= 0.1$). This is perhaps a reflection of the comparatively low number of haematological lines (n=129) reducing the statistical power of these analyses, although genuine differences in the sensitivity profiles of the OE lines cannot be ruled out, potentially due to alternative genetic or epigenetic drivers.

Further analyses focused on solid tumour cell lines only (n=567). This allowed a focus on a more homogeneous group that have demonstrated limited efficacy of epigenetic therapy in the clinic



(a)



(b)

Figure 3.42 ***EZH2* or *G9a* OE versus sensitivity.** Relationship between overexpression (OE) of (a) *EZH2* or (b) *EHMT2* (encoding G9a) and sensitivity to HKMTi-1-005, GSK343 (EZH2i), UNC0638 (G9ai) and cisplatin. Cell lines included are those from the GDSC2 cell line panel with sensitivity data for each of the four compounds and gene expression data, derived from any tumour type (n=694). OE was defined using a z-score of > 1.

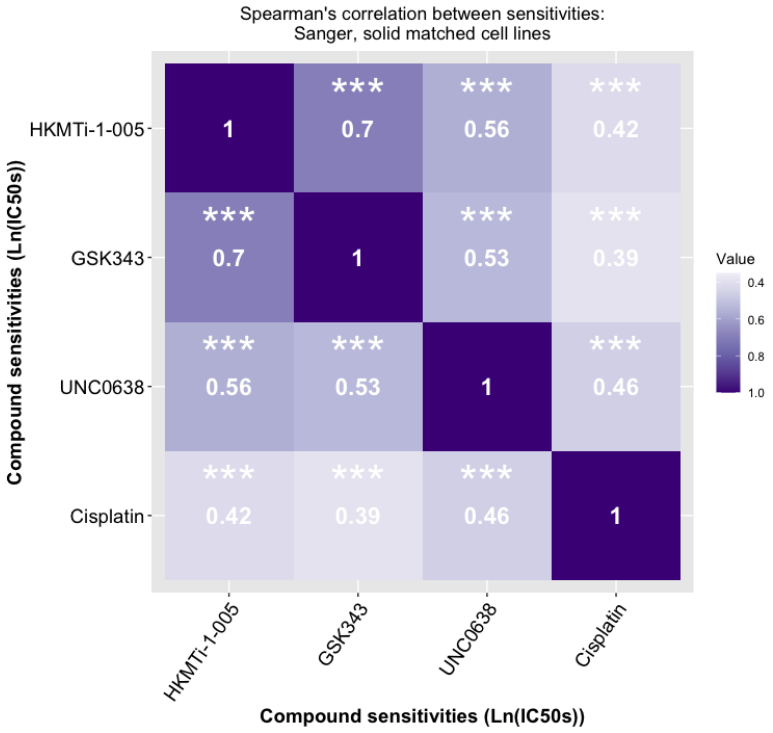
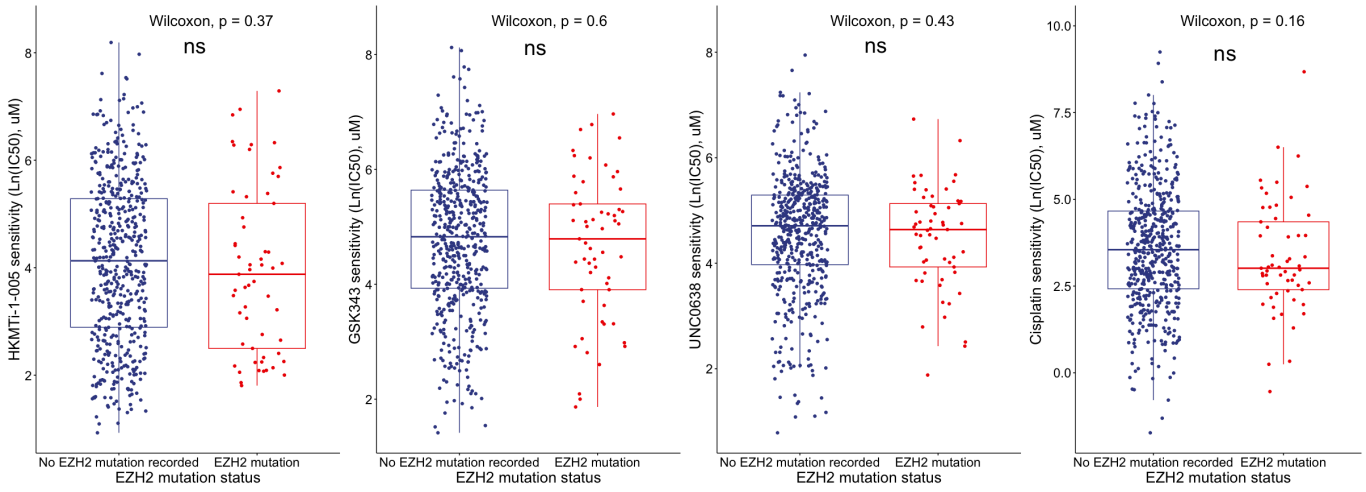
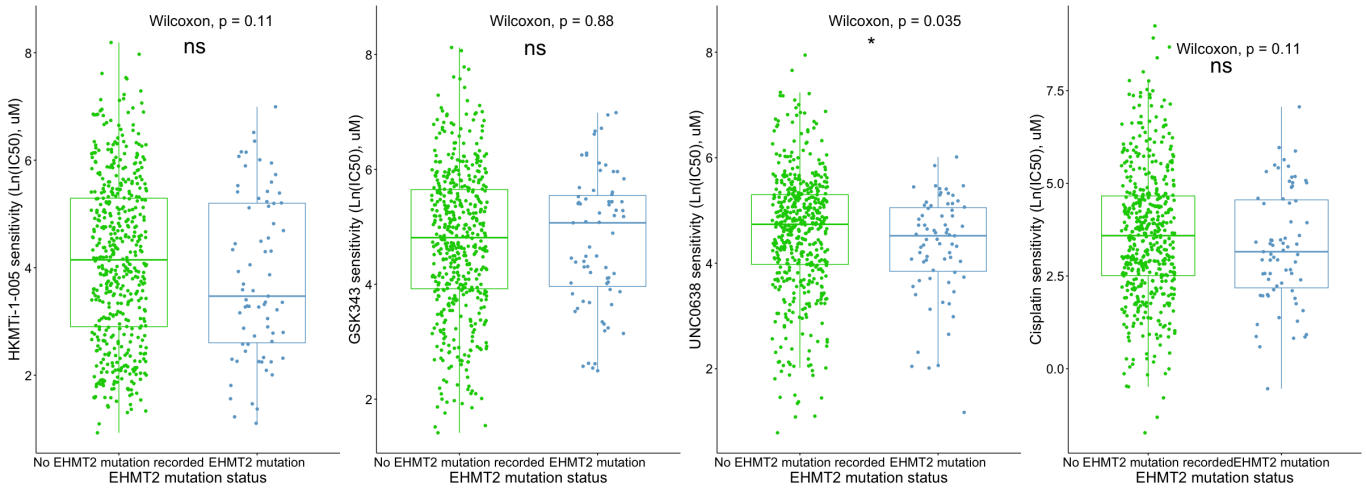


Figure 3.43 **Spearman's correlation between sensitivity to HKMTi-1-005, GSK343 (EZH2i), UNC0638 (G9ai) and cisplatin: solid cell lines.** Correlation across solid tumour cell lines only, with sensitivity data for each of the four compounds, n = 567. Ln(IC_{50}) = natural log of half-maximal inhibitory concentration.

to date (see 1.6.1) [136]. Correlation between sensitivities to each compound were plotted for the matched panel of 567 solid tumour cell lines, confirming the stronger relationship between HKMTi-1-005 sensitivity and single EZH2i or G9ai as compared to cisplatin observed in the larger cohort (figure 3.43) (HKMTi-1-005-GSK343 95% CI [0.66,0.74], HKMTi-1-005-UNC0638 95% CI [0.50,0.62], HKMTi-1-005-cisplatin 95% CI [0.35,0.49]). In this solid cell line cohort, neither *EZH2* nor *EHMT2* mutation was associated with greater sensitivity to the compounds, excepting *EHMT2* mutation and UNC0638, with a trend towards increased sensitivity to HKMTi-1-005 (p=0.1) (figure 3.44). The small number of *EZH2/EHMT2*-mutated lines may prevent mutation status associations being evident in haematological or solid cohorts separately. However, lines with OE of *EZH2* still demonstrated increased sensitivity to all compounds, and those with OE of *EHMT2* were more sensitive to all except the GSK343 EZH2i (figure 3.45). Hence, the relationship between *EZH2/EHMT2* status and sensitivity to HKMTi-1-005 is similar to that observed for single EZH2 and G9a inhibitors.

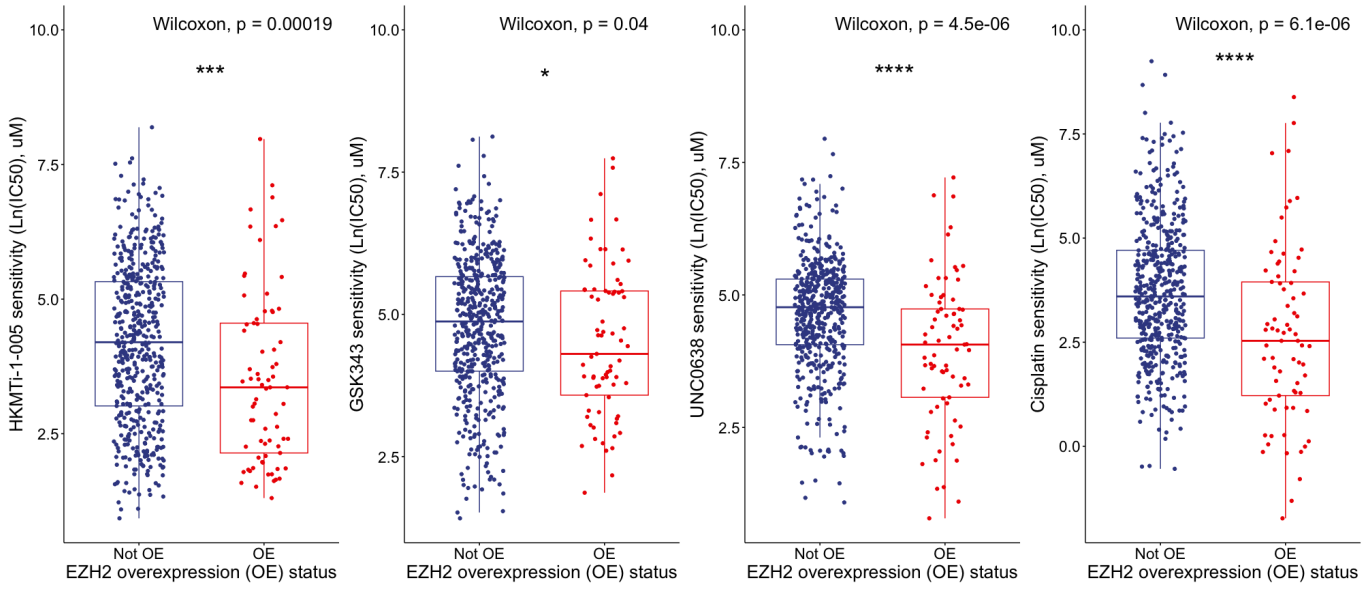


(a)

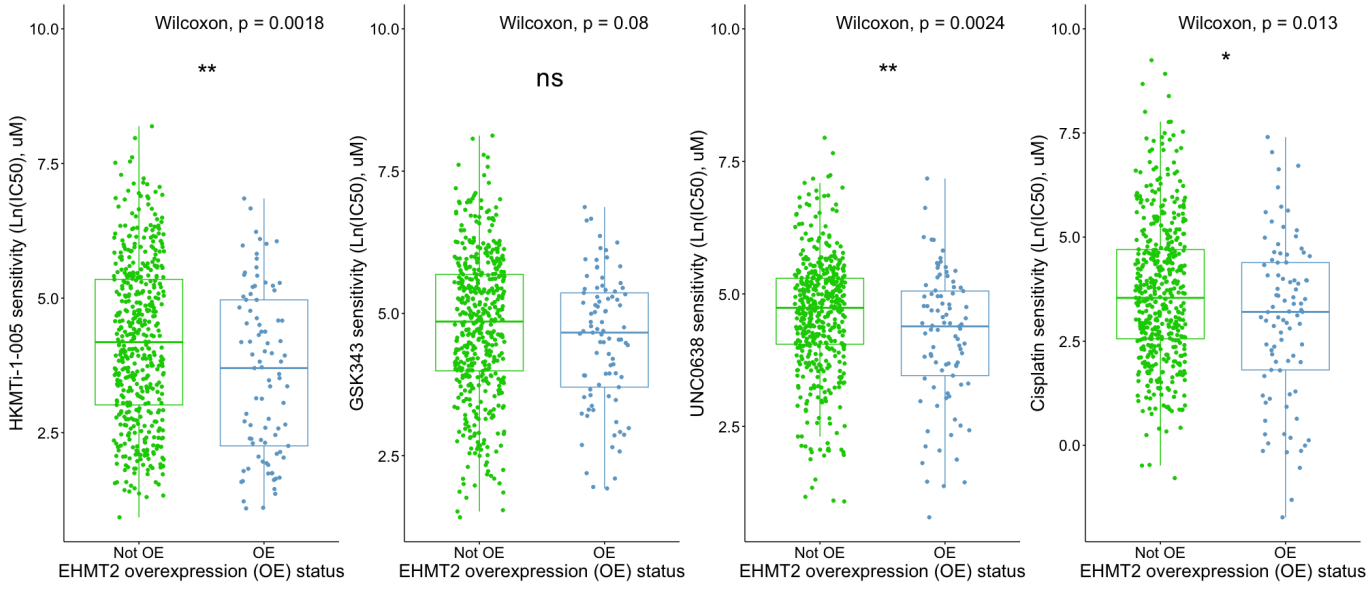


(b)

Figure 3.44 ***EZH2* or *G9a* mutations versus sensitivity: solid cell lines.** Relationship between mutations in (a) EZH2 or (b) EHMT2 (encoding G9a) and sensitivity to HKMTi-1-005, GSK343 (EZH2i), UNC0638 (G9ai) and cisplatin. Cell lines included are solid cell lines from the GDSC2 panel with sensitivity data for each of the four compounds (n=567). Mutation data were downloaded from COSMIC [407].



(a)



(b)

Figure 3.45 ***EZH2* or *G9a* OE versus sensitivity: solid cell lines.** Relationship between overexpression (OE) of (a) *EZH2* or (b) *EHMT2* (encoding G9a) and sensitivity to HKMTi-1-005, GSK343 (EZH2i), UNC0638 (G9ai) and cisplatin. Cell lines included are solid cell lines from the GDSC2 panel with sensitivity data for each of the four compounds and gene expression data (n=565). OE was defined using a z-score of > 1.

3.4.4 AstraZeneca: studying the biochemical activity of HKMTi-1-005

Curry et al. [109] initially identified HKMTi-1-005 based on a phenotypic screen for compounds that induced changes in EZH2-silenced gene expression. Transcriptional effects supporting EZH2 inhibition were observed both at specific target genes and globally, as were gene-specific and global changes in H3K9 and H3K27 methylation. These effects were also found to be due to substrate-competitive inhibition of EZH2, based on *in vitro* analysis utilising competition for a biotinylated peptide substrate. In collaboration with AstraZeneca (AZ), our group looked to validate the EZH2 inhibitory effects of HKMTi-1-005 and the potential mechanisms of action.

This was first assessed using H3K27me3 as a readout of EZH2 activity. To do so, AZ collaborators used a fluorescence resonance energy transfer (FRET) assay in G401 cells (an epithelial kidney rhabdoid tumour cell line). Probes that bind both H3 and K27me3 were introduced into cells, allowing induction of fluorescence when the probes exist in close proximity as a readout of H3K27me3 levels. Alongside the FRET assay, a cell viability assay was performed by AZ.

The results demonstrated that HKMTi-1-005 had an IC_{50} value of 8.7µM (first run) or 3.3µM (second run) (table 3.3), supporting cell-based assays by Curry et al. and confirming the reduction of H3K27me3 observed using both western Blot and ChIP approaches. The results were limited to some degree by difficulty in distinction between loss of cell viability and loss of H3K27me3 (personal communication). However, the results did suggest a distinct effect from the A-366 compound - with a distinct chemical scaffold to the three other G9ai compounds - where no loss of cell viability nor of H3K27me3 was observed at the concentrations tested.

The FRET cell-based approach was run in parallel with another investigation performed by AZ, utilising a cell-free model to assess EZH2i activity (see figure 2.2 for explanatory schematic). Whilst the original investigation by Curry et al. utilised a 3-component complex for cell-free investigations, AZ employed a larger 5-component complex. This consisted of EZH2, EED, SUZ12, RBBP4 (RbAp48) and AEBP2 subunits. The effect of HKMTi-1-005 on activity of this PRC2 complex was examined in the presence of peptide or nucleotide substrates. A very high IC_{50} of >100µM was

Compound	**IC_{50} (first run) (µM)**	**IC_{50} (second run) (µM)**
HKMTi-1-005	8.7	3.3
BIX-01294	ND	3.1
UNC0638	ND	4.3
A-366	ND	>100

Table 3.3 **IC_{50} values for H3K27me3 inhibition.** Based on FRET assays performed by AZ. Compounds include HKMTi-1-005 and comparison enzymatic G9a inhibitors BIX-01294, UNC0638 and A-366. ND = not done; only HKMTi-1-005 was included in the first run. In the second run, A-366 did not have any measurable effect on H3K27me3.

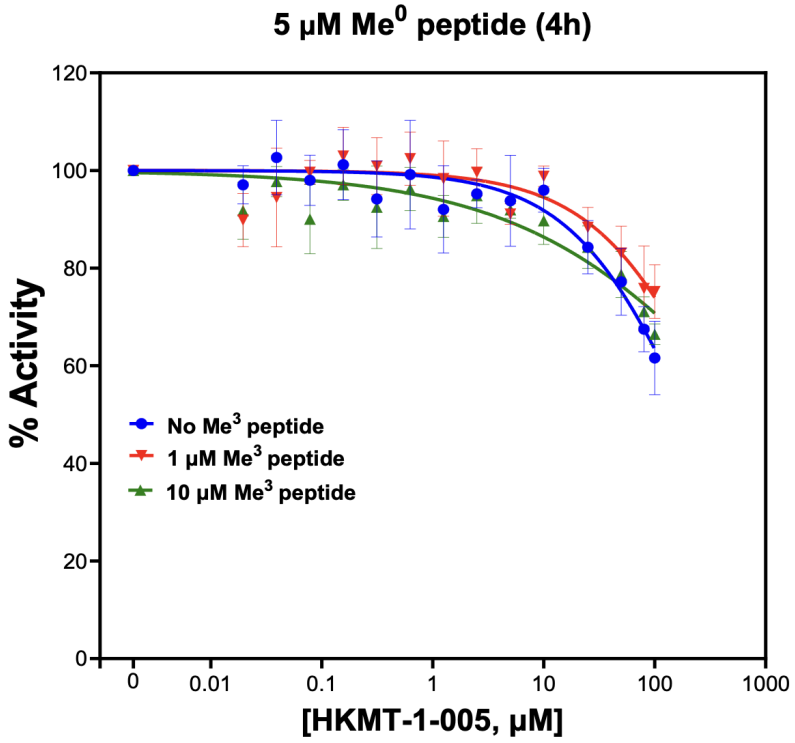


Figure 3.46 **PRC2 inhibition assay results.** Data from AZ demonstrating very weak inhibition of the 5-component PRC2 complex by HKMTi-1-005. Percentage activity is normalised to a control inhibitor of PRC2 (AZ14102599, an allosteric EED inhibitor); 100% activity represents no inhibition, 0% enzyme activity represents maximum inhibition by the control inhibitor. The assay uses 5μM unmethylated peptide (Me0) to allow enzymatic activity of the 5-component PRC2, with varying concentrations of trimethylated peptide (Me3) to test for any peptide-competitive effects of the compound.

observed in the cell-free assay (figure 3.46, personal communication), suggesting that the cellular environment is essential in coordinating the EZH2i effects of HKMTi-1-005. Taken together, the data gathered by AZ in combination with the original investigations suggest that HKMTi-1-005 has inhibitory effects on H3K9 and H3K27 methylation at EZH2 and G9a target genes, but that there may be limited direct inhibition of EZH2 enzymatic activity, with alternative mechanisms contributing to EZH2 target gene effects.

3.5 Discussion: conclusions, limitations and future work

The results presented in this chapter support EZH2i- and G9ai-like behaviour of the HKMTi-1-005 compound. Yet, they also suggest that some of the EZH2i effects may be indirect rather than direct enzymatic inhibition.

Changes in accessibility and transcription following treatment with HKMTi-1-005 were investigated in HGSC models, both *in vitro* and *in vivo*, to determine treatment effects at EZH2- and G9a-regulated loci. These analyses revealed changes at gene targets such as the *IL24* locus that was identified as a dual, but not single, EZH2/G9a silencing target by Casciello et al. [107]. Additionally, broader changes in genesets reflecting EZH2-regulated or dual EZH2/G9a-regulated genes supported inhibitory effects of HKMTi-1-005 against the activity of both enzymes. It must be acknowledged that a limitation of the fGSEA approaches for examining EZH2i/G9ai effects of the compound is the lack of published G9ai genesets. HKMTi-1-005 was originally based on modification of a G9ai, and the data from Curry et al. describe strong G9ai effects *in vitro* with proof of a substrate-specific mechanism of action [109]. The consistent changes in pathways such as cMYC signalling [110, 268, 287], cell cycle progression [137], apoptosis [138–140], metabolism and immune signalling [141, 142] also reflected current literature on the effect of EZH2i or G9ai in cancer cells. Hence, there exists strong evidence from multiple analyses for G9ai activity of the compound despite the lack of G9ai genesets.

Following treatment with HKMTi-1-005, there was a trend towards greater chromatin accessibility as anticipated. However, many sites were also significantly less accessible in each sample, despite the drug targeting the largely repressive H3K27/H3K9 marks. It is possible that these changes represent off-target effects of HKMTi-1-005 at non-H3K9/H3K27 loci. However, this effect may instead be explained by the extensive crosstalk between epigenetic modifiers (section 1.4), such as compensatory deposition or removal of alternative histone or DNA marks. For example, there was a significant transcriptional downregulation of the KDM3B lysine demethylase in all cell lines except OVCAR5: this enzyme removes methylation at H3K9 [135], hence its downregulation promotes chromatin compaction. Additionally, HDACs were altered across most samples following treatment, an effect that could promote changes in acetylation marks that are associated with increased accessibility. Aside from crosstalk effects, methylation loss at H3K9 and H3K27 residues also has the potential to directly promote decreased transcription and therefore accessibility. For example, whilst most H3K9 methylation – especially when located at promoter regions – is associated with transcriptional repression, H3K9me3 has also been associated with active gene expression when enriched across the coding region of genes [436]. Similarly, H3K27me1 is most often associated with constitutive heterochromatin, yet enrichment of the same mark over gene bodies is positively correlated with transcription

[71]. Accordingly, loss of H3K9me/H3K27me following treatment with a dual EZH2/G9a targeting compound should be expected to mediate an overall trend towards increased chromatin accessibility and gene expression, but with induction of the opposite effect at a subset of loci where methylation is associated with transcriptional activation as opposed to repression

ATACseq measures changes in chromatin accessibility, but without consideration of specific histone marks. Consequently, changes in accessibility at genomic coordinates associated with H3K27/H3K9 and other chromatin marks could not be measured directly. Instead, coordinates associated with these marks were obtained from public ChIPseq databases. In the absence of ChIPseq data derived from the relevant differentiated cell types, data from early developmental stages were chosen as a reference for histone mark coordinates. Given the highly cell type-specific nature of chromatin marks following differentiation [10, 12, 15] and the reversion of subsets of cancer cells to stem-like chromatin characteristics [328], these datasets were considered the best reference for analysis of H3K27/H3K9 domains in HGSC models.

It is likely that many histone mark coordinates overlap between the HGSC models and reference datasets. However, cell type- and cancer-specific H3K9/H3K27 domains would not have been accounted for, hence the exact specificity and extent of HKMTi-1-005 effects at H3K27/H3K9 domains could not be measured and may be underestimated using this approach. The use of ChIPseq in place of ATACseq would overcome this issue and allow direct measurement of H3K27/H3K9 methylation marks. This limitation in assessing specific accessibility changes at H3K27/H3K9 domains was compounded by the aforementioned crosstalk between epigenetic marks, which made selection of a control histone mark difficult: changes in H3K27/H3K9 methylation are not expected to be independent from any of the other examined marks, hence the specificity of HKMTi-1-005 effects at H3K27/H3K9 methylation domains is hard to prove.

Despite these limitations, the data did support some specificity of HKMTi-1-005 treatment. Stronger and more consistent increases in accessibility were observed over H3K27/H3K9-associated chromatin than over other domains. However, ATACseq also demonstrated increased accessibility at H3K36me3 and H3K4me1 coordinates. Assessment of enhancer accessibility and of intersected chromatin domains suggested some of these changes may be due to increased accessibility at enhancers and bivalent chromatin. Yet, changes at H3K36me3 and H3K4me1/3 domains are also likely to be due to knock-on effects of EZH2 or G9a inhibitory activity. For example, if loss of repressive H3K27/H3K9 methylation at gene promoters results in transcriptional upregulation, it is probable that this is accompanied by increased accessibility over H3K36me3 domains that are found across gene bodies [102] and over H3K4me1/3 that are associated with active promoters [97]. Taken together, HKMTi-1-005 effects over different chromatin domains support EZH2i and G9ai activity whilst emphasizing the highly interconnected nature of chromatin regulation, and demonstrate that

a subset of chromatin accessibility changes following dual EZH2/G9a targeting occur at enhancer and bivalent loci.

It must be acknowledged that the bulk nature of sequencing in the ID8 *in vivo* mouse model presents several limitations. First, it prevents distinction between on and off-target effects of HKMTi-1-005 treatment. The overall survival benefit accompanying treatment [155] and the tolerability profile of the drug suggest that non-tumour cell toxicity effects do not outweigh drug benefit. Additionally, Curry et al. showed increased sensitivity of tumour over normal cells to the compound [109], an observation that has been demonstrated with EZH2i and G9ai or their combination in other cell line models [107]. This specificity may in part be driven by the aberrant genetic and epigenetic profiles of tumour cells, with potentially lower thresholds for viral mimicry/inflammatory cell death [220, 235, 252–256]. Ongoing investigation into off-target effects are essential to ensure they are minimal and to help verify the safety of potential combination therapy approaches, especially given the importance of epigenetics in the maturation and polarisation of immune cell populations [28, 82].

Following assessment of the HKMTi-1-005 treatment effect in HGSC models, the sensitivity profile of HKMTi-1-005 was investigated. Utilising the GDSC2 panel, HKMTi-1-005 was compared to a single EZH2i, a single G9ai and to the first-line HGSC treatment cisplatin across pan-cancer cell lines. Overall sensitivity to the compound across the cell line panel was greater than to single EZH2 or G9a inhibitors. The compound demonstrated a similar sensitivity profile to single EZH2i/G9ai, including heightened sensitivity of haematological lines and cell lines with aberrant *EZH2* or *EHMT2/G9a*. Aside from these anticipated sensitive cell line cohorts, GDSC2 data suggested heightened HKMTi-1-005 sensitivity of subgroups of cell lines from SCLC, melanoma and Ewing's sarcoma tumour types, with a distinct distribution from that observed for single inhibitors or cisplatin. This finding recommends further investigation into drivers of sensitivity in these specific cell lines, whilst supporting both EZH2i- and G9ai-like behaviour of HKMTi-1-005.

Despite sensitivity data and treatment effects suggesting EZH2i and G9ai activity, biochemical investigation by AZ suggested there was no direct effect of HKMTi-1-005 on EZH2 enzymatic activity in a cell-free model. This contrasted with assays measuring H3K27me3, which demonstrated inhibitory HKMTi-1-005 effect at IC_{50} values equivalent to those observed in cell-based assays. Such observations imply that the cell-free model is unable to recapitulate aspects of the cellular environment required for HKMTi-1-005 EZH2 target gene effects.

The cell-free assay utilised a 5-component PRC2 complex and a peptide or nucleosome substrate. PRC2 subunits and chromatin components may be lacking in this model that are essential for the inhibitory activity observed in cells, providing insight into potential indirect mechanisms of action of the drug. Additionally, evidence from the RNAseq and ATACseq analyses suggested HKMTi-1-005 effects at EZH2 targets may be distinct from direct enzymatic inhibition: despite both EZH2-activated

genes and EZH2-silenced genes becoming more accessible following HKMTi-1-005 treatment, the former were transcriptionally downregulated whilst the latter were upregulated. Genesets defined using knockdown of EZH2 – rather than EZH2 enzymatic inhibition – were also significantly altered by HKMTi-1-005 treatment. These effects are difficult to explain assuming direct enzymatic activity of the compound against EZH2, without accounting for inhibition of non-canonical EZH2 activity. Indeed, enzymatic inhibitors have been proven insufficient for immune stimulation [250] and cMYC signalling disruption [292, 300, 301] by some groups, due to a failure to target non-canonical EZH2 activity, yet HKMTi-1-005 induces these pathways.

Several explanatory hypotheses exist. HKMTi-1-005 could mediate its activity at EZH2 targets via 1) disruption of EZH2-G9a crosstalk, 2) prevention of PRC2 chromatin targeting, or 3) direct depletion of PRC2 components at the RNA or protein level. Disruption of EZH2-G9a crosstalk would not be captured in the assay, including effects such as the loss of G9a-mediated H3K27 methylation (which could subsequently reduce PRC2 chromatin targeting [69], amplifying H3K27 methylation loss) or of physical EZH2-G9a interaction required for methylation of certain EZH2 targets. An example of this includes the G9a-PALI1-PRC2 super-complex described in section 1.4.3, with both H3K9 and H3K27 methyltransferase activity: inhibition of this complex would not be accounted for by the cell-free assay.

Whilst thorough investigation of these hypotheses could not be undertaken, a preliminary assessment of HKMTi-1-005 effects on PRC2 components did not rule out a depletion effect. First, analysis of H3K9 and H3K27 methylation machinery genesets showed a strong trend towards downregulation. Further, analysis of transcriptional changes in PRC2 components suggested downregulation of several components in each sample. PRC2-recruiting lncRNAs did not demonstrate any consistent downregulation and the effect on PRC2 components themselves was modest, hence no definitive conclusions can be drawn from this data. Additionally, many PRC2 recruitment factors cannot be measured by RNAseq analyses, such as specific chromatin characteristics or occupancy by other proteins. Further investigation into the mechanism of action of HKMTi-1-005 is warranted, including validation of PRC2 component or cofactor depletion at the protein level, to determine how treatment effects at EZH2-regulated targets are achieved.

Such investigations are worth pursuing: evidence from the literature suggests that some of the anti-tumour activity observed with EZH2 depletion therapies are not achieved using enzymatic EZH2 inhibitors [250, 292, 300, 301]. Supporting the findings presented here, Sbirkov et al. recently demonstrated that HKMTi-1-005 treatment but not use of the enzymatic EZH2i GSK343 phenocopied EZH2 knockdown in acute myeloid leukemia cell lines, despite GSK343 significantly reducing H3K27me3 [437]. Initial research to resolve EZH2i effects could involve examination of PRC2 components at the protein level using western blots, followed by co-immunoprecipitation (co-IP) experiments to assess

the compound's capacity to interfere with EZH2-G9a interaction, or to interfere with PRC2 complex formation. Additionally, ChIPseq to investigate PRC2 chromatin occupancy in control versus HKMTi-1-005-treated cells would help to determine whether treatment affects PRC2 recruitment to target sites.

In conclusion, the results presented in this chapter illustrate a convincing EZH2i- and G9ai-like behaviour of HKMTi-1-005 in HGSC models. They further demonstrate a sensitivity profile similar to both EZH2i and G9ai in pan-cancer cell line analyses, and evidence heightened sensitivity of haematological, melanoma, Ewing's sarcoma and SCLC cell lines. The data recommend further investigation into the mechanism of action of HKMTi-1-005 against EZH2 activity, as direct enzymatic inhibition is not supported by cell-free assays. Additionally, an analysis into the drivers of sensitivity in cell lines may provide valuable insight into which tumour types are potential targets for clinical efficacy.

Chapter 4

Results Section 2: Characterisation of the viral mimicry response induced by HKMTi-1-005

4.1 Introduction

The viral mimicry (VM) response is thought to provide anti-tumour immune stimulation [213, 229]. This effect is driven by the de-repression of transposable elements (TEs), particularly endogenous retroviruses (ERVs). TEs are recognised by cytoplasmic receptors, with subsequent stimulation of intracellular innate immune signalling and cell death pathways [225–228, 251] (see 1.7.5). Production of signalling molecules such as IFNs amplifies the response via autocrine and paracrine signalling effects [251].

Most literature assessing VM induction is focused on DNMTi use, as DNA methylation is a central regulator of TE silencing [190, 191]. However, numerous other epigenetic modifications are involved in control of TE expression [189] including histone modifications; targeting of EZH2 or G9a enzymes has also been shown to induce a VM response [110, 153, 226, 241, 243]. Yet, individual targeting of any of these epigenetic enzymes can be restricted by compensatory mechanisms employed to prevent TE expression. For example, the use of DNMTi individually may be limited by a compensatory increase in H3K27me3 over TE loci following DNA hypomethylation [195, 438], and stronger VM is observed when EZH2i is used in combination with DNMTi [153]. Similarly, it has been reported that targeting of either EZH2 or G9a individually is insufficient to induce the VM response [152, 153]. Whilst combinations of EZH2i or G9ai with DNMTi have proven successful in some tumour models for immune stimulation [120, 128, 151, 152], this has not always been evident: in melanoma, breast and ovarian cancer lines, no combinatorial effect was observed between DNMTi and either EZH2i or

G9ai [107]. Instead, the combination of EZH2 and G9a targeting was required to achieve efficacy.

As detailed in chapter 3, HKMTi-1-005 has EZH2i- and G9ai-like effects in cell line and *in vivo* models of HGSC. Our group has previously shown induction of an immune response by the compound in the ID8 *in vivo* model, with accompanying expression of ERVs [155]. These initial findings motivated further investigation into induction of a VM response by the compound in HGSC models, particularly into aspects of the VM response that may differ from, or complement, the DNMTi-induced response. The literature suggests that histone modifications silence different types of ERVs compared to DNA methylation: older/intermediate age ERVs are silenced largely by histone modifications, rather than the DNA methylation that prevails over young ERVs, due to the loss of CpG sites over time [153]. Hence, the targeting of histone methylation marks may induce distinct TEs and possibly distinct immune effects as compared to DNMTi use.

Several factors have been identified that may affect the extent of the VM response induced by epigenetic or other compounds (see 1.7.7). One such factor is the basal expression of ERVs. However, whilst high basal ERV expression has been proposed to "prime" or sensitise cells to VM in some instances [220], contrasting data suggest that high basal ERV expression may result in lower sensitivity to VM induction, with a higher concentration of epigenetic drugs needed for treatment effect [214]. Potentially, expression of the downstream interferon-stimulated genes (ISGs) may provide an alternative sensitivity marker; where high basal ERV expression was associated with insensitivity to VM induction, it was accompanied by high basal ISG expression that could limit further induction of VM response genes or indicate IFN tolerance in these cells [214]. Hence, additional investigation into the nature of the ERV/ISG-VM relationship may provide insight into potential patient stratification for VM sensitivity.

Taken together, the literature highlights several features of the VM response to HKMTi-1-005 treatment that warrant investigation. I first looked to characterise the type of immune response induced by the compound, to provide insight into the signalling pathways implicated in treatment response. To do so, immune gene expression following HKMTi-1-005 treatment was investigated in a panel of HGSC cell lines alongside re-analysis of the ID8 *in vivo* data. As part of this characterisation, I aimed to identify a subset of genes that act as central markers of the VM response specific to HKMTi-1-005.

Subsequently, I expanded the initial TE analysis performed on the ID8 *in vivo* model to include assessment of expression changes across all TE superfamilies following treatment, to determine whether ERV elements are the relevant TE targets of the HKMTi-1-005 compound. Mapping of reads to these repetitive regions is complex [159], therefore analysis of treatment effect was also modified to use specialised tools rather than the standard RNAseq pipeline. These specialised tools use expectation-maximisation algorithms to reassign multimapping reads to the most likely TE sources. Within TE

analyses, I looked to assess the hypothesis that older/intermediate age ERVs are particularly reliant on histone methylation silencing and are therefore disproportionately represented in HKMTi-1-005 TE targets. Finally, gene and TE expression data were investigated to determine whether our data provide support for a relationship between basal ERV/ISG status and VM induction, and in which direction the relationship lies.

In summary, chapter 4 aims to investigate and characterise any viral mimicry effects downstream of HKMTi-1-005 treatment in HGSC models. The aims and objectives of this chapter are summarised by the following questions:

- Does HKMTi-1-005 induce immune signalling pathways in HGSC cell lines as well as the ID8 *in vivo* mouse model, and how can this immune signalling be characterised?
- Is HKMTi-1-005-induced immune signalling accompanied by transcriptional upregulation of TEs? If so, which types of TEs are induced?
- Within the ERV superfamily, are ERV elements from older families more likely than those from younger ERV families to be induced by HKMTi-1-005 treatment?
- Are transcriptional changes in TEs accompanied by changes in chromatin accessibility?
- Do the HGSC cell line data lend support to the hypothesis that basal ERV/ISG status relates to sensitivity to VM induction?

4.2 HKMTi-1-005 induces immune stimulation in HGSC models

4.2.1 Transcriptional changes in immune pathways induced by HKMTi-1-005

The initial investigations into HKMTi-1-005 effects on HGSC models suggested an immune stimulation effect (see chapter 3), supporting the current literature examining EZH2i and G9ai effects on cell lines and tumours. Additionally, all samples except PEO4 demonstrated overrepresentation of pathways relating to viral processes (figure 3.7). These preliminary findings suggested potential viral mimicry induced by the HKMTi-1-005 compound.

The nature of the immune response induced was investigated using the fGSEA approach, beginning with an examination of immune pathways in the Hallmark database (figure 4.1). These data supported the overall skew towards immune geneset upregulation, with a very strong upregulation of immune genesets in the ID8 *in vivo* sample. A mixed effect was seen across the HGSC cell lines. However, certain genesets such as complement, IL2-STAT5, inflammatory response, and IFNG signalling showed a more consistent upregulation following HKMTi-1-005 treatment. IL6-JAK-STAT3, interferon alpha (IFNA), and TNFA hallmark pathways were the least consistently upregulated immune genesets across cell lines.

Given the literature describing innate immune interferon signalling as a central component of the VM response [222], a refined set of pathways reflecting this response were examined separately. This included the hallmark interferon alpha (IFNA) and gamma (IFNG) genesets, alongside an additional geneset for interferon beta (IFNB) signalling (GOBP response to interferon beta) and a

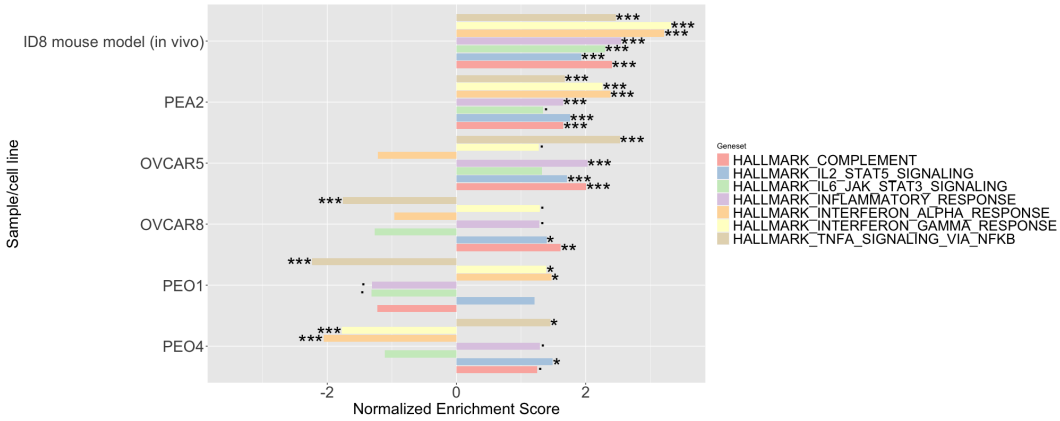


Figure 4.1 **fGSEA of hallmark genesets, HGSC samples.** fGSEA results for the ID8 *in vivo* HGSC model, PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines, comparing RNAseq data from control and HKMTi-1-005-treated samples. Genesets are from the Hallmark collection hosted by MSigDB [416], narrowed to those directly relating to immune signalling. . = p-adj < 0.1, * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001.

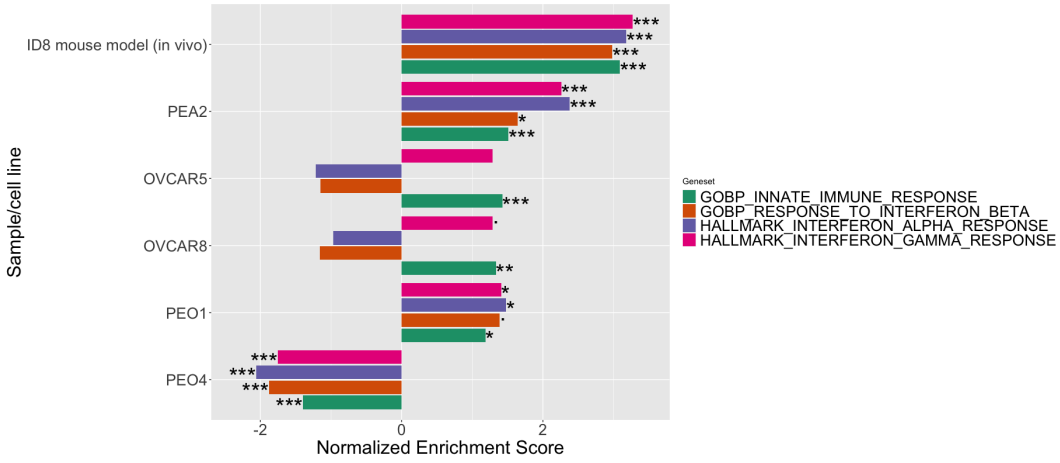


Figure 4.2 **fGSEA of immune genesets, HGSC samples.** fGSEA results for the ID8 *in vivo* HGSC model, PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines, comparing RNAseq data from control and HKMTi-1-005-treated samples. Genesets include interferon alpha and gamma response from the Hallmark collection, alongside two additional genesets selected from the GOBP database, describing the innate immune response and the response to interferon beta. All genesets are hosted by MSigDB [416]. . = p-adj < 0.1, * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001.

general innate immune response geneset (GOBP innate immune response), both from the GOBP database. The results demonstrated a consistent upregulation of innate immune signalling in all samples except for PEO4 (figure 4.2), alongside upregulation of the hallmark IFNG response geneset in most samples. Strikingly, the PEO4 line exhibited a marked downregulation of all four genesets examined. In contrast to innate immune and IFNG genesets, IFNA/IFNB response genesets were variable between samples.

As the innate immune response and IFNG genesets were the most consistently induced, I next looked to identify which signalling sub-pathways were underlying their upregulation. Utilising data from the MSigDB [416], I categorised genes from the GOBP innate immune and hallmark IFNG genesets that were significantly upregulated (figure 4.3). It is important to note that classification is not precise; many genes may be placed in multiple different categories due to the intersecting nature of the immune signalling pathways, and multiple genesets describing each signalling pathway are hosted by MSigDB. However, the results demonstrated upregulation of genes involved in antigen presentation, chemokine signalling, inflammasome signalling, STING signalling, RIG-I signalling, and TLR signalling in different samples. Particularly, genes in the 'defense response to virus' pathway were among the upregulated genes in every sample (pathway named 'response to virus' in the ID8 sample).

Of these IFNG and innate immune response genes, I next looked to define those genes that were most consistently upregulated across cell lines. There were 107 genes that were significantly upregulated in at least 2 of the 5 cell lines (log2FC> 0, p-adj < 0.05), and not significantly downregulated in any lines (log2FC< 0, p-adj < 0.05) (figures 4.4a and 4.4b). Most of these genes were also upregulated in the orthologous mouse genes in the ID8 *in vivo* sample (figure 4.4c).

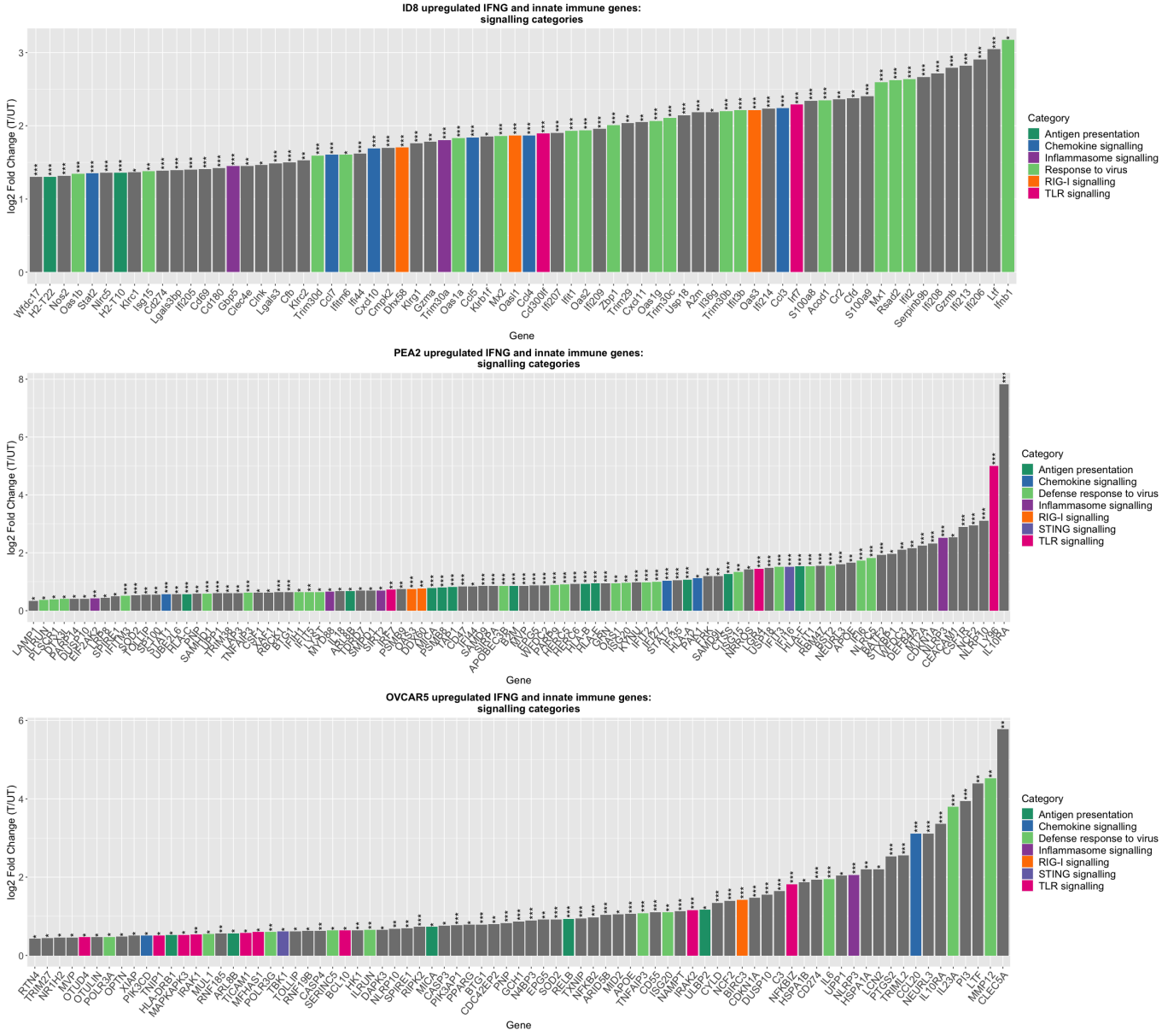


Figure 4.3 **IFNG and innate immune gene expression changes.** Continued overleaf.

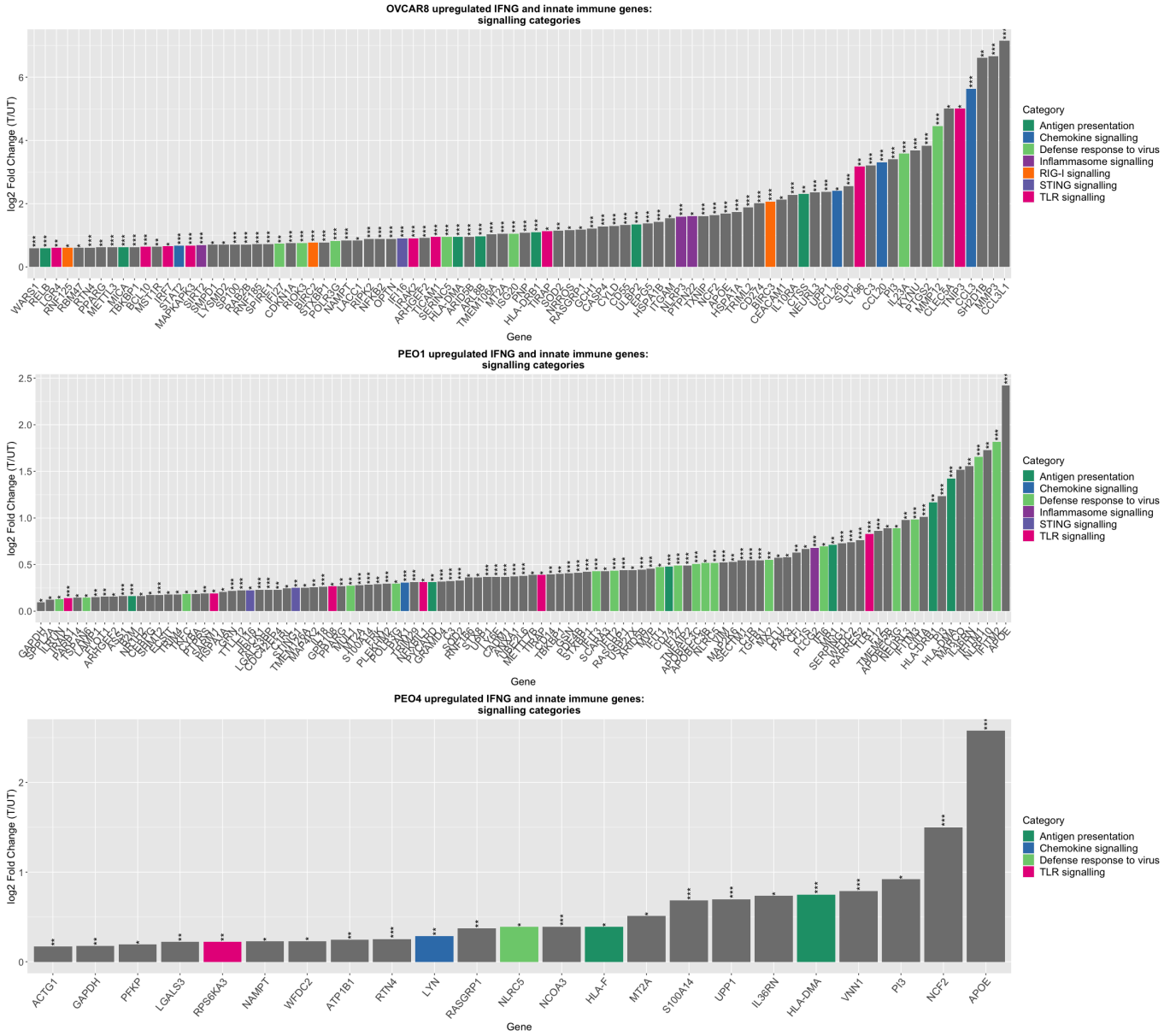
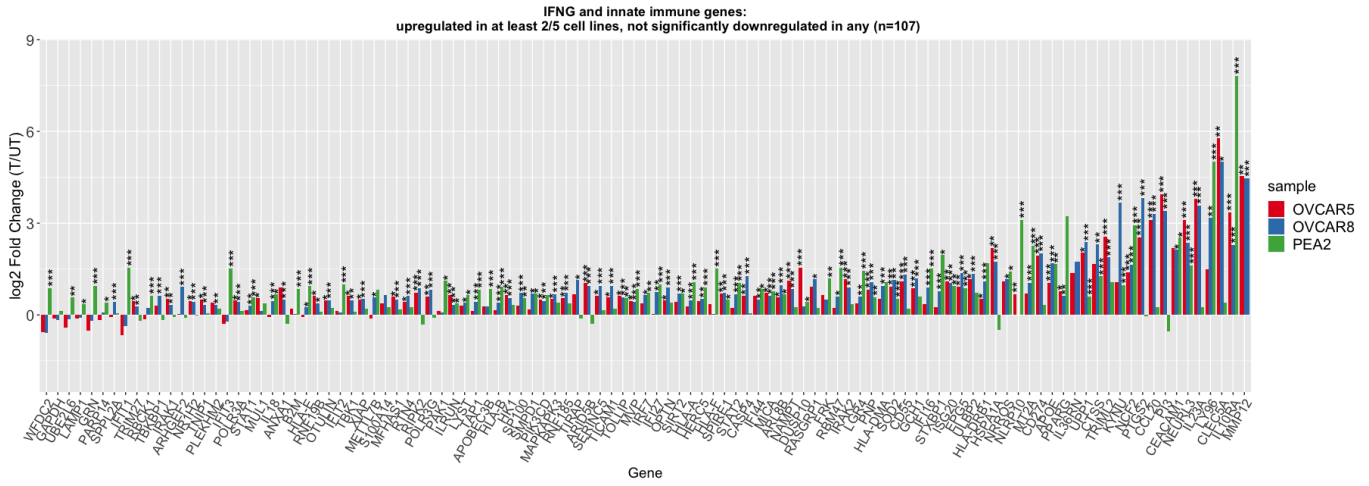
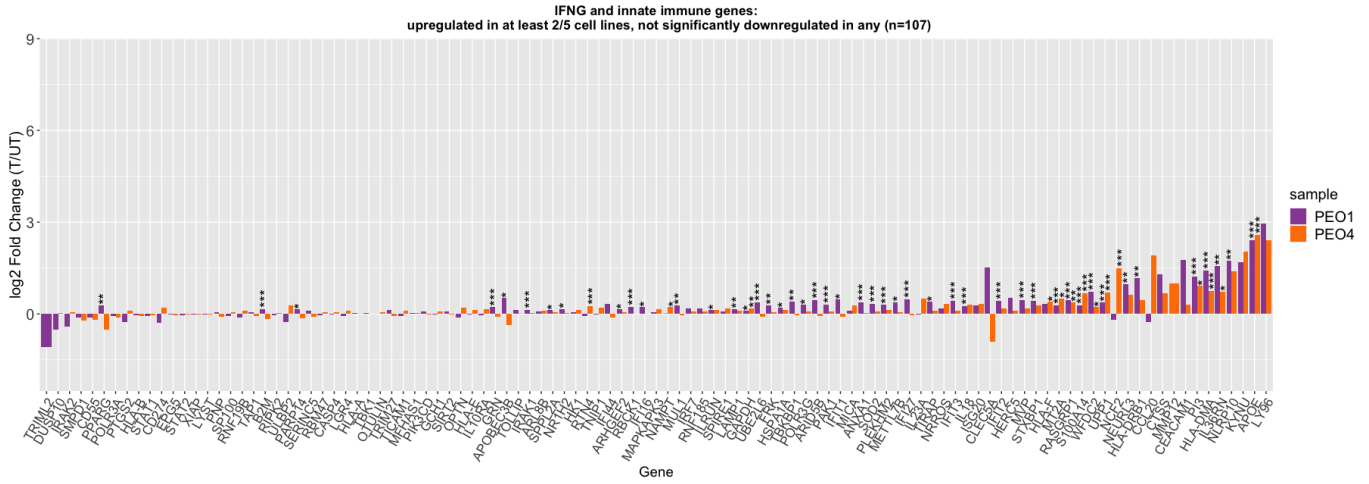


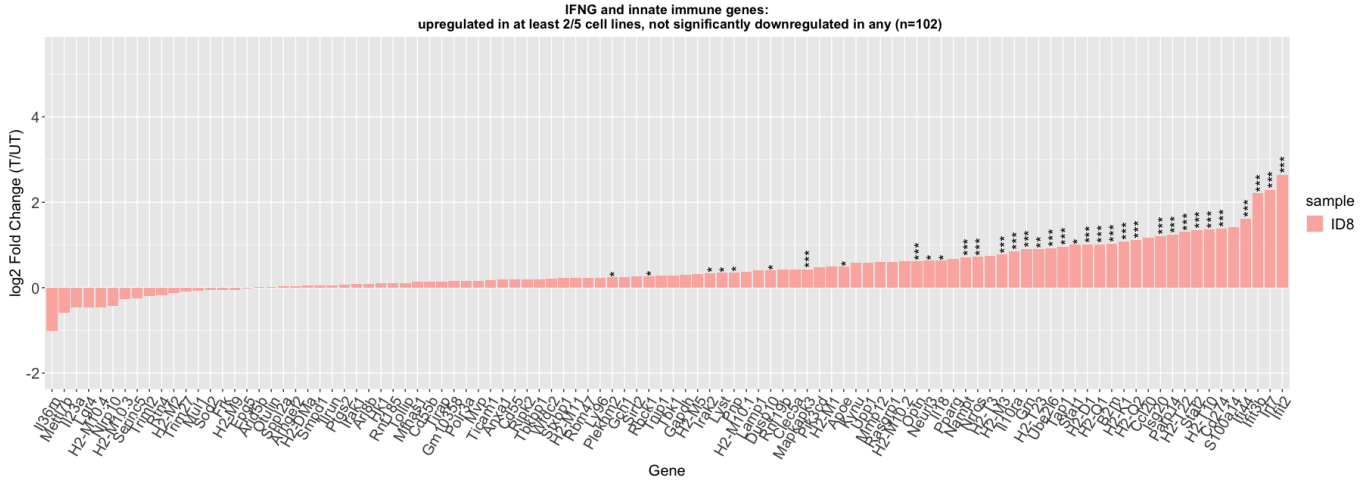
Figure 4.3 **IFNG and innate immune gene expression changes.** Visualisation of genes from the GOBP pathway "innate immune response" and the hallmark pathway "interferon gamma response" that are significantly upregulated (log2FC > 0, p-adj < 0.05) in HGSC samples following HKMTi-1-005 treatment. Genes are labelled by different innate immune signalling categories, based on genesets from the MSigDB [416]. Samples (from top to bottom) include: the ID8 *in vivo* HGSC model, PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines. RIG-I = Retinoic Acid Inducible Gene I, STING = Stimulator Of Interferon Response CGAMP Interactor 1, TLR = Toll-like receptor. RIG-I, STING, and TLRs all act as pattern recognition receptors (PRRs), involved in the sensing of pathogens. * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001. N.B. ID8 *in vivo* and OVCAR8 samples are visualized with an additional log2FC cut-off of 0.58 and 1.3 respectively, equivalent to a fold-change of 1.5 and 2.5 respectively, to reduce the number of genes plotted.



(a)



(b)



(c)

Figure 4.4 **Consistently upregulated immune genes, HGSC samples.** Innate immune response and IFNG response genes upregulated in at least 2/5 HGSC cell lines following HKMTi-1-005 treatment, and not significantly downregulated in any line (log2FC > 0 or < 0 respectively, p-adj < 0.05). (a) Changes in these genes (n=107) in the PEA2, OVCAR5 and OVCAR8 cell lines. (b) Changes in these genes (n=107) in the PEO1 and PEO4 cell lines. (c) Changes in mouse orthologues of these genes (n=102) in the ID8 *in vivo* HGSC model. * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001.

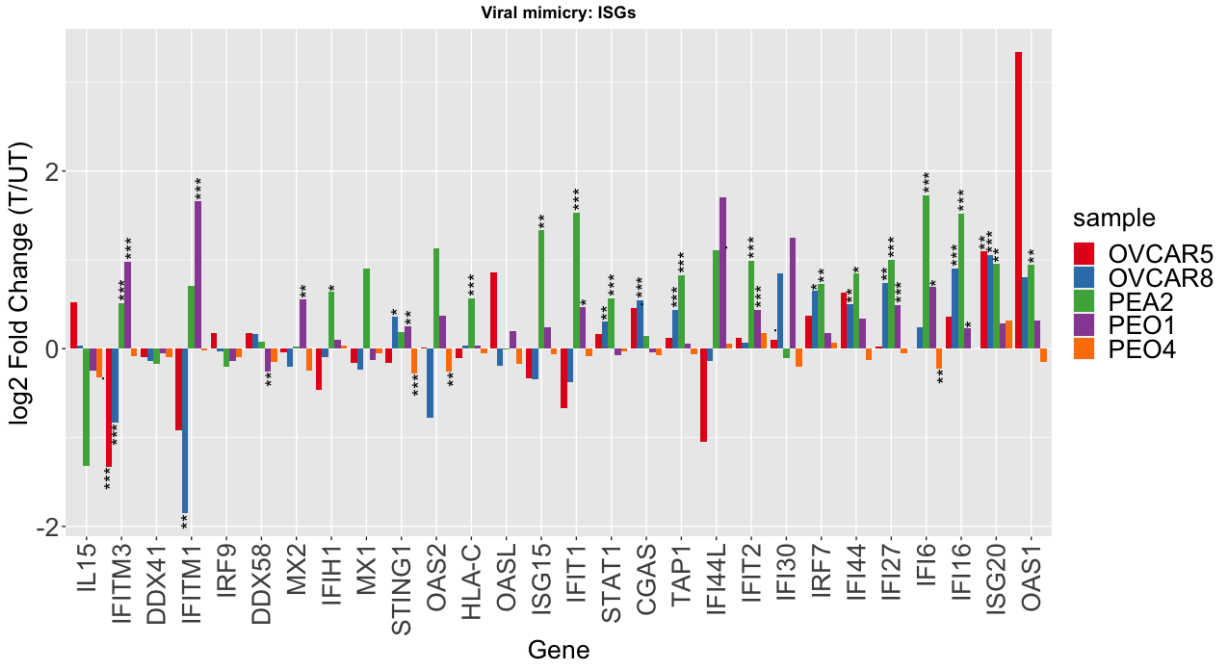
The data demonstrate consistent upregulation of many innate immune and IFNG response genes, signalling pathways that are known to be engaged in VM. A subset of these IFNG and innate immune genes are ISGs, which are important downstream signalling components of the VM response [251]. The extensive effects of IFNs make definition of ISGs difficult. As a result, ISGs here were defined as those in a published panel of 28 genes previously used in the evaluation of the VM response in ovarian tumours [439] (table 4.1: 27 of the 28 genes are in the broader IFNG/innate immune sets. DDX41 is the single additional gene). Examining expression changes of these genes confirmed upregulation of VM ISGs in most samples (figure 4.5a and 4.5b). Once again, the ID8 *in vivo* sample showed very strong immune stimulation, and PEO4 was the outlier with a very limited effect of treatment on ISGs, and with more significantly down- than up-regulated genes.

From these 28 VM ISGs, I looked to define a reduced signature of genes involved in the HKMTi-1-005 viral mimicry immune response. To select genes that were strongly or consistently upregulated, the criteria were as follows: fold-change > 1 and p-adj < 0.01 in at least 2 of the 5 cell lines, or fold-change > 1 in every cell line, and excluding any genes that were significantly downregulated in any cell line with p-adj < 0.01 (appendix figure A.1). This resulted in selection of 7 genes: *IFI16*, *IFI27*, *IFIT2*, *IRF7*, *ISG20*, *STAT1*, and *TAP1* (figure 4.6), from here on referred to as the 7ISG. Analysis of changes in the ID8 sample demonstrated a consistent upregulation of mouse orthologues of the 7ISG *in vivo* too (figure 4.7). Of note, 5 of these 7 genes (*IFI27*, *IFIT2*, *IRF7*, *ISG20* and *STAT1*) exist in both IFNG and innate immune response genesets, and the remaining two exist in either the IFNG response set (*TAP1*) or the innate immune response set (*IFI16*).

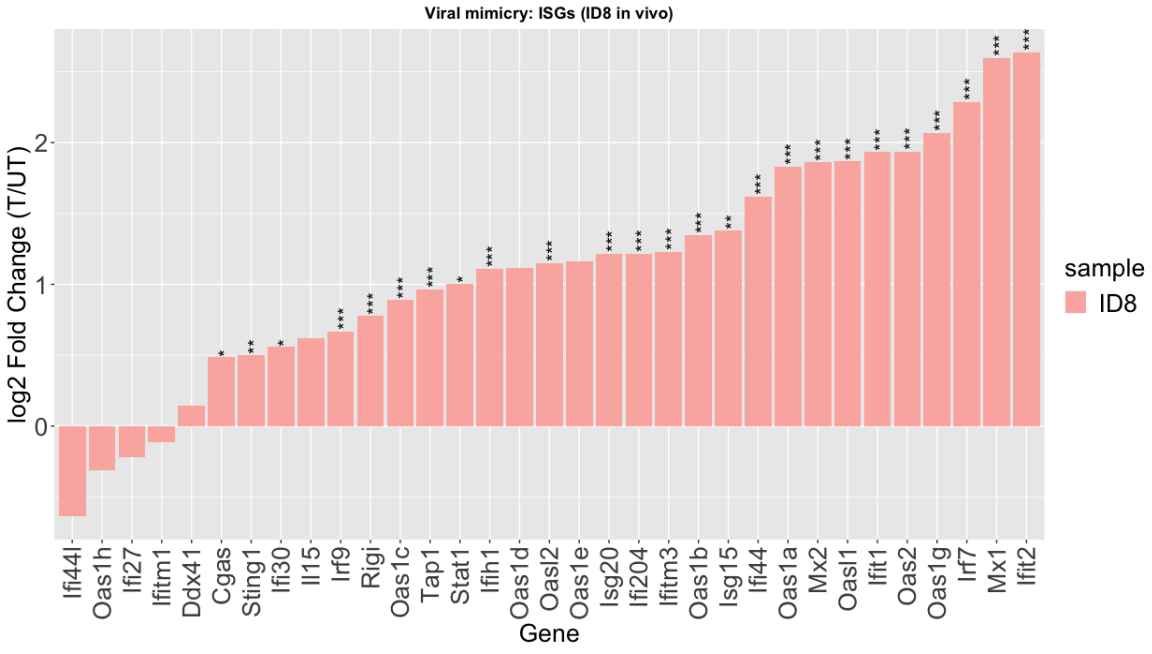
The data presented demonstrate that both innate immune and IFNG responses, including many ISGs, are upregulated by HKMTi-1-005 treatment in HGSC models. The 7ISG represents a reduced signature geneset describing strongly and consistently HKMTi-1-005-upregulated VM genes. Yet, the results also demonstrate a discordant treatment effect on PEO4 cells compared with other lines, despite its derivation from the same patient as PEO1. This effect is further investigated in section 4.4.

Gene	Description	Alias
CGAS	cyclic GMP-AMP synthase	*C6orf150, D4, MB21D1, h-cGAS*
DDX41	DEAD-box helicase 41	*ABS, MPLPF*
DDX58	RNA sensor RIG-I	*DDX58, RIG-I, RIG1, RLR-1, SGMRT2*
HLA-C	major histocompatibility complex, class I, C	*D6S204, HLA-JY3, HLAC, HLC-C, MHC*
IFI16	interferon gamma inducible protein 16	*IFNGIP1, PYHIN2*
IFI27	interferon alpha inducible protein 27	*FAM14D, ISG12, ISG12A, P27*
IFI30	IFI30 lysosomal thiol reductase	*GILT, IFI-30, IP-30, IP30*
IFI44	interferon induced protein 44	*MTAP44, TLDC5, p44*
IFI44L	interferon induced protein 44 like	*C1orf29, GS3686, TLDC5B*
IFI6	interferon alpha inducible protein 6	*6-16, FAM14C, G1P3, IFI-6-16, IFI616*
IFIH1	interferon induced with helicase C domain 1	*AGS7, Hlcd, IDDM19, IMD95, MDA-5*
IFIT1	interferon induced protein with tetratricopeptide repeats 1	*C56, G10P1, IFI-56, IFI-56K, IFI56*
IFIT2	interferon induced protein with tetratricopeptide repeats 2	*G10P2, GARG-39, IFI-54, IFI-54K, IFI54*
IFITM1	interferon induced transmembrane protein 1	*9-27, CD225, DSPA2a, IFI17, LEU13*
IFITM3	interferon induced transmembrane protein 3	*1-8U, DSPA2b, IP15*
IL15	interleukin 15	*IL-15*
IRF7	interferon regulatory factor 7	*IMD39, IRF-7, IRF-7H, IRF7A, IRF7B*
IRF9	interferon regulatory factor 9	*IRF-9, ISGF3, ISGF3G, p48*
ISG15	ISG15 ubiquitin like modifier	*G1P2, IFI15, IMD38, IP17, UCRP*
ISG20	interferon stimulated exonuclease gene 20	*CD25, HEM45*
MX1	MX dynamin like GTPase 1	*IFI-78K, IFI78, MX, MxA, lncMX1-215*
MX2	MX dynamin like GTPase 2	*MXB*
OAS1	2'-5'-oligoadenylate synthetase 1	*E18/E16, IFI-4, IMD100, OIAS, OIASI*
OAS2	2'-5'-oligoadenylate synthetase 2	
OASL	2'-5'-oligoadenylate synthetase like	*OASL1, OASLd, TRIP-14, TRIP14, p59 OASL*
STAT1	signal transducer and activator of transcription 1	*CANDF7, IMD31A, IMD31B, IMD31C, ISGF-3*
STING1	stimulator of interferon response cGAMP interactor 1	*STING, TMEM173, ERIS, MITA, MPYS, NET23, SAVI*
TAP1	SEC14 like lipid binding 2	*C22orf6, SPF, TAP*

Table 4.1 **Genes in the ISG geneset defined in [439].** Gene descriptions from g:Profiler (`https://biit.cs.ut.ee/gprofiler/convert`).



(a)



(b)

Figure 4.5 **ISG expression changes.** Change in expression of interferon-stimulated genes (ISGs) following HKMTi-1-005 treatment, utilising those ISGs defined in [439] assessing the viral mimicry response in ovarian tumours. (a) Changes in ISG expression in the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines. (b) Changes in expression of mouse orthologues of ISGs, in the ID8 *in vivo* HGSC model. * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001.

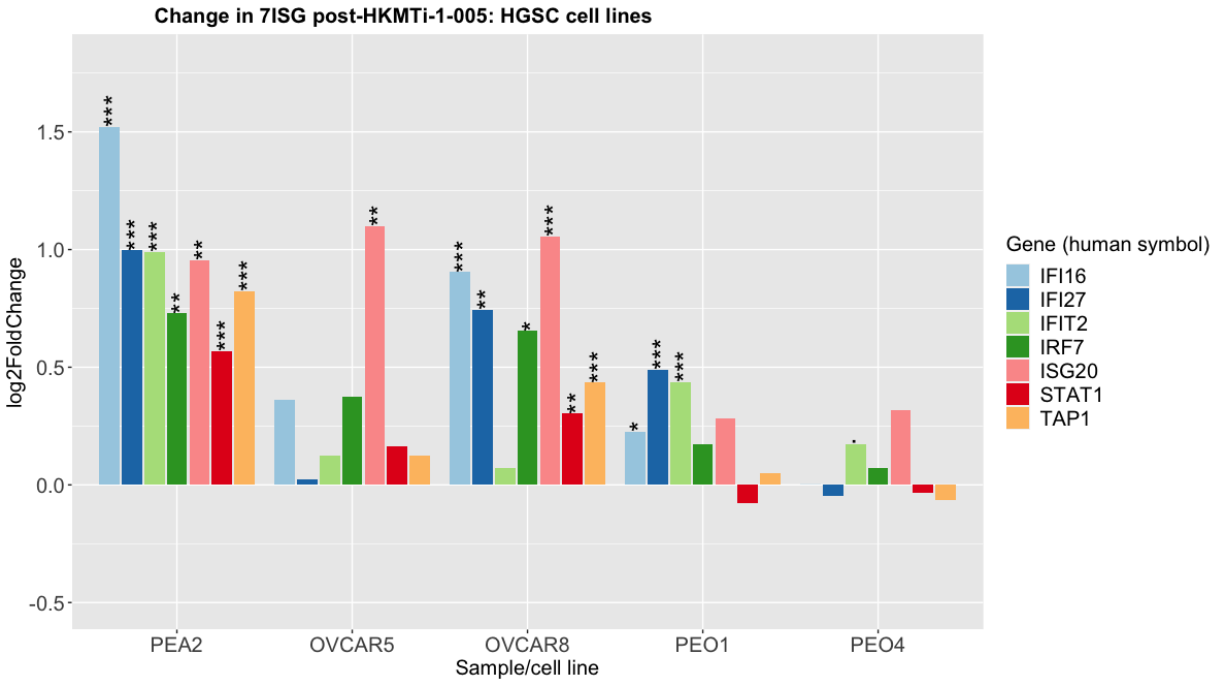


Figure 4.6 **7ISG expression changes, HGSC cell lines.** Bar plots showing transcriptional changes in the 7ISGs following HKMTi-1-005 treatment, across the five HGSC cell lines PEA2, OVCAR5, OVCAR8, PEO1, and PEO4. 7ISGs were defined as upregulated genes from the 28-gene VM ISG set [439] with p-adj < 0.01 in at least 2 of the 5 cell lines, or fold-change > 1 in every cell line, excluding any genes that were significantly downregulated in any lines (p-adj < 0.01). . = p-adj < 0.1, * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001.

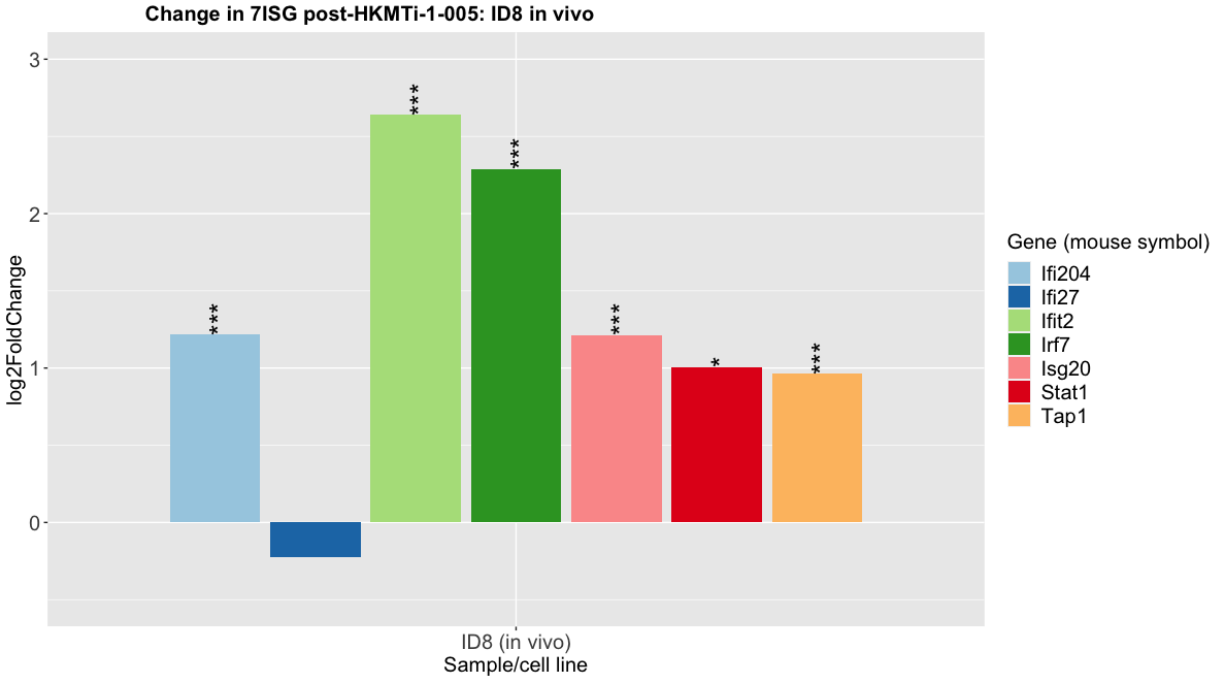


Figure 4.7 **7ISG expression changes, ID8 *in vivo* model.** Bar plots showing transcriptional changes following HKMTi-1-005 treatment in mouse orthologues of the 7ISGs, in the ID8 *in vivo* sample. 7ISGs were defined as upregulated genes from the 28-gene VM ISG set [439] with p-adj $<$ 0.01 in at least 2 of 5 HGSC cell lines following HKMTi-1-005 treatment, or fold-change > 1 in every cell line, excluding any genes that were significantly downregulated in any lines (p-adj < 0.01). . = p-adj < 0.1, * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001. Note that Ifi204 is the mouse orthologue of human IFI16.

4.3 HKMTi-1-005 induces transposable elements in HGSC models

4.3.1 De-repression of transposable element transcription: sub-family level analysis

Having established the transcriptional induction of innate immune and IFNG response pathways, including ISGs, I next investigated whether transposable elements were also upregulated by HKMTi-1-005.

To do so, I first used the TEtranscripts tool [382, 383]. This assesses transposable element expression at the sub-family level, combining all counts from individual loci within a sub-family. Figure 4.8 gives an overview of the TE nomenclature used in this section. Note that the terms ERV and LTR are used interchangeably in this section; LTRs flank the internal coding domains of ERVs in their canonical structure (see figure 1.10b), but ERVs exist most often as solo LTRs in the genome due to recombination and mutation over time [164]. Therefore, treatment-induced changes at LTRs are annotated within the broader category of ERVs.

Including TEs of all superfamilies – the most abundant categories of which are SINEs, LINEs, DNA elements and LTRs/ERVs – there was upregulation of between 3 and 18 sub-families in each sample (p-adj < 0.05) (figures 4.9, 4.10, 4.11 and 4.12a). Many other TE sub-families were upregulated, but most changes were not significant at the specified threshold (figure 4.9 and 4.10). In all samples except the ID8 *in vivo* sample, there were more up- than down-regulated TEs (figure 4.11a). Intriguingly, the sample with the smallest total number of upregulated TEs at the sub-family level was the ID8 *in vivo* sample that had demonstrated the most extensive immune stimulation effect.

LTR/ERV superfamily elements made up the vast majority of transcriptionally upregulated TEs across all samples (figure 4.12b). Indeed, in the ID8 and PEO4 samples, 100% of upregulated TEs

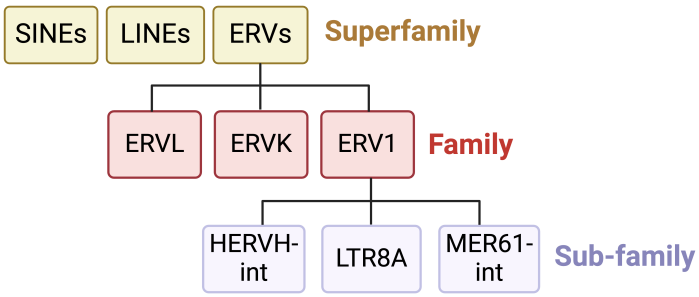


Figure 4.8 **TE and ERV nomenclature used in analyses.** TE superfamilies include element types such as SINEs, LINEs and ERVs. The ERV superfamily can be divided into families such as ERVL, ERVK and ERV1 (otherwise known as class III, II and I respectively). ERV families can be further divided into sub-families, including e.g. HERVH-int.



Figure 4.9 **TE expression changes by superfamily, at the sub-family level.** Distribution of transcriptional fold-changes in TEs following HKMTi-1-005 treatment, across the the ID8 *in vivo* mouse model and the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines. TE differential expression measured at the sub-family level, using the TEtranscripts tool [382, 383]. TE superfamilies are labelled by colour according to the legend on the right-hand side ("class_id"). Classification: short-interspersed nuclear elements (SINE), long-interspersed nuclear elements (LINE), long terminal repeat elements (LTR), retroposons, DNA repeat elements (DNA), satellite repeats, Rolling Circle (RC). ? indicates uncertain classification.

were LTRs/ERVs. This contrasted with transcriptionally downregulated TEs, which were represented by varying proportions of different TE superfamilies (figure 4.12b). Importantly, the proportion of upregulated TEs that were of the LTR/ERV superfamily was much greater than the reference genome distribution in all samples (figures 4.12b and 4.12c), suggesting a selective and concordant upregulation of elements from this versus other TE superfamilies and versus a background effect. As observed when considering all TE types combined, all samples except the ID8 *in vivo* demonstrated a greater number of significantly up- versus down-regulated LTRs/ERVs (figure 4.11b).

Few TE sub-families were commonly upregulated across cell lines. However, the overlap of up-regulated TEs was significant between cell lines despite the low overall number induced (figure 4.13. Note: ID8 sample cannot be compared). The most consistent changes were all in LTRs/ERVs: HERVH-int, LTR7Y (an LTR sequence of the full HERVH provirus) and LTR1C1 sub-families were upregulated across 5, 4, and 3 of the HGSC cell lines respectively, each of which are ERV1 family elements.

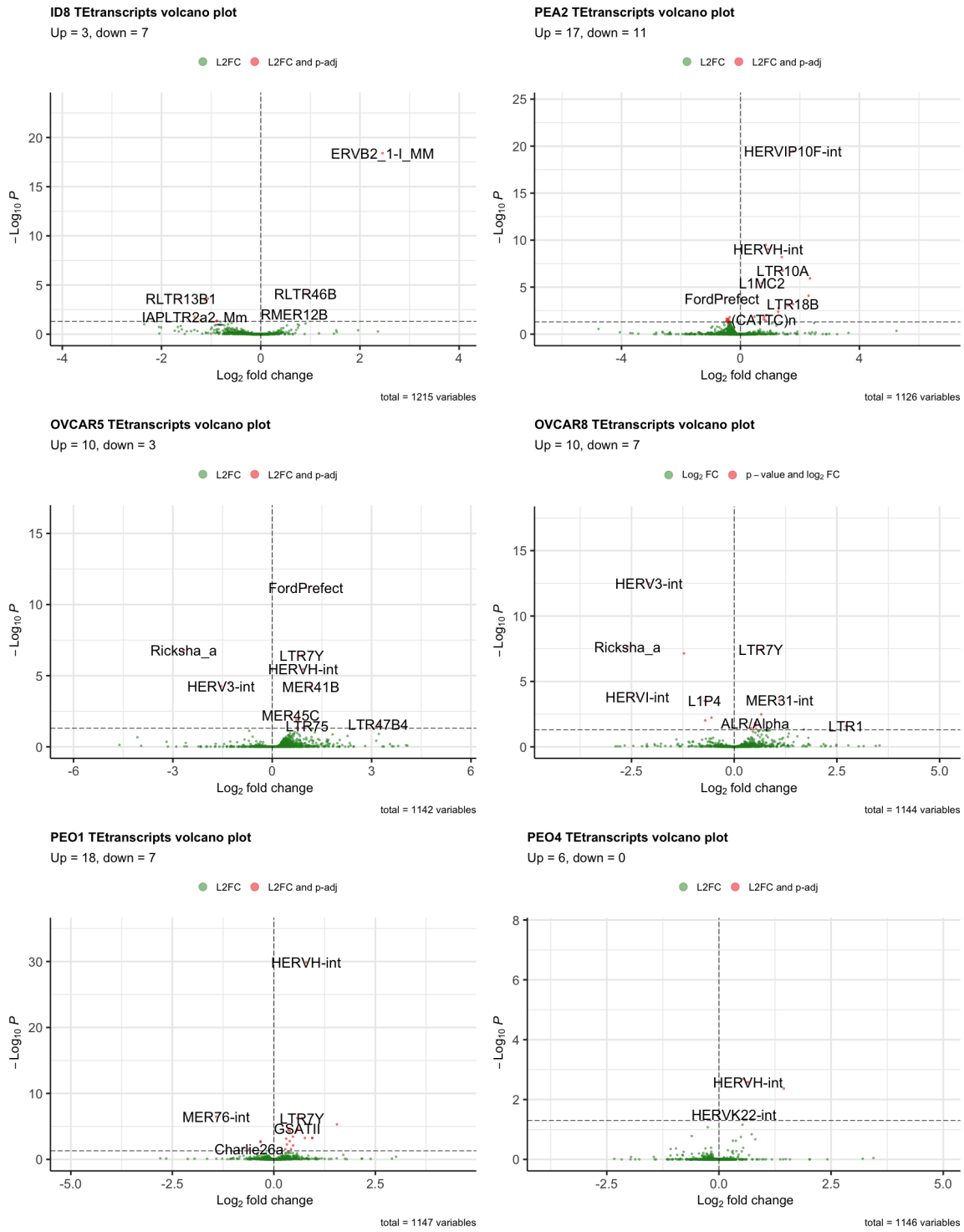
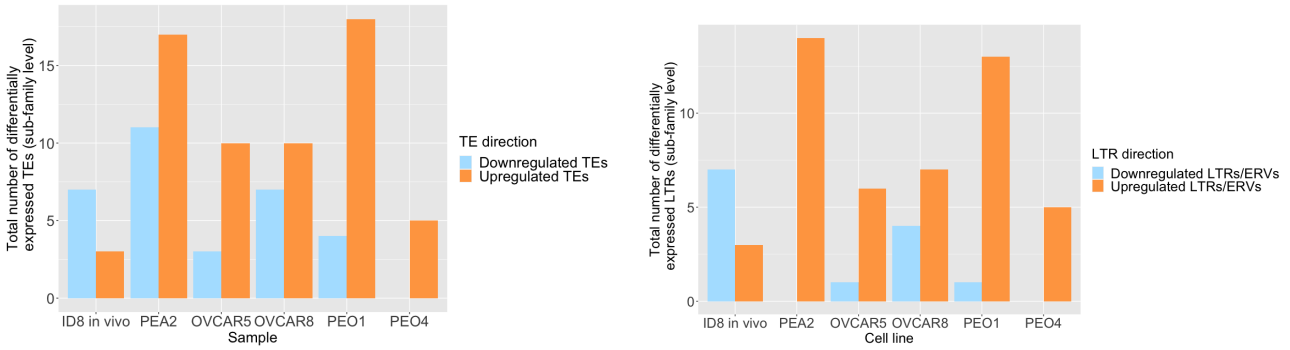


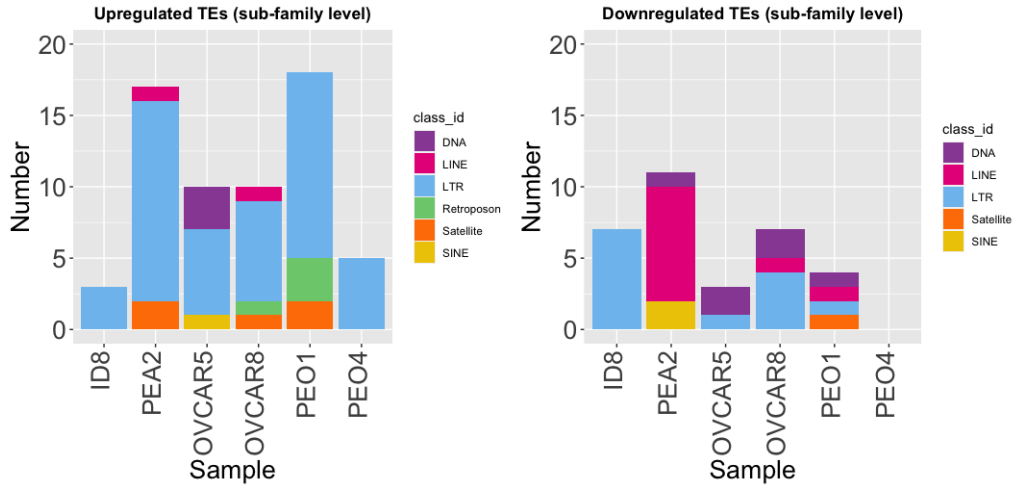
Figure 4.10 **Volcano plot of TE changes in HGSC samples, sub-family level.** Volcano plot examining differential expression of TEs at the sub-family level, in control versus HKMTi-1-005-treated conditions, using the TEtranscripts tool [382, 383]. Samples include (top to bottom, left to right): the ID8 *in vivo* mouse model and the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines. TEs are labelled by colour, based on significance and fold-change: log2FC (green) = absolute log2FC >0, log2FC and p-adj (red) = absolute log2FC > 0 and p-adj < 0.05.



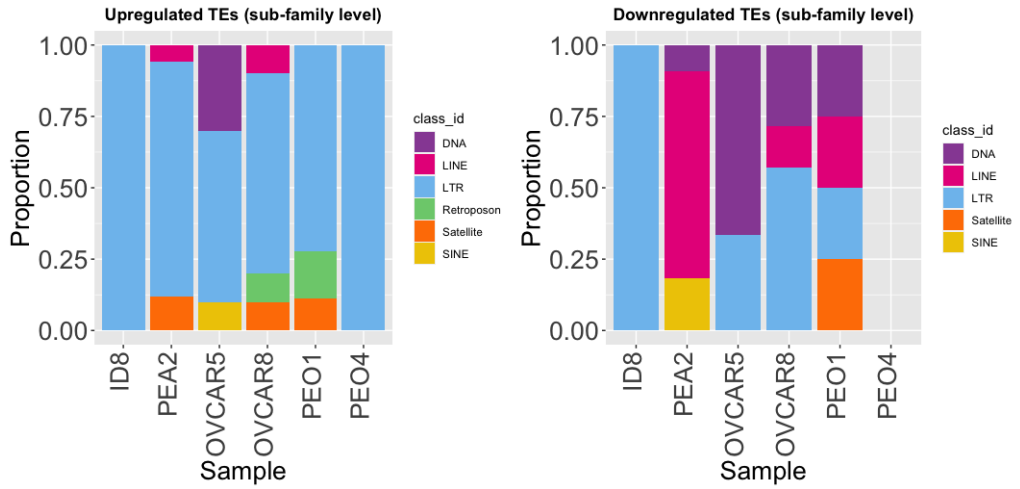
(a) (b)

Figure 4.11 **Total up- versus down-regulated TEs in HGSC samples, sub-family level.** Bar plots showing total number of transcriptionally up- versus down-regulated (a) TEs or (b) LTRs/ERVs across the the ID8 *in vivo* mouse model and the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines, in control versus HKMTi-1-005-treated conditions. TEs measured at the sub-family level, with a threshold of p-adj < 0.05 for differential expression. Blue = downregulated, orange = upregulated.

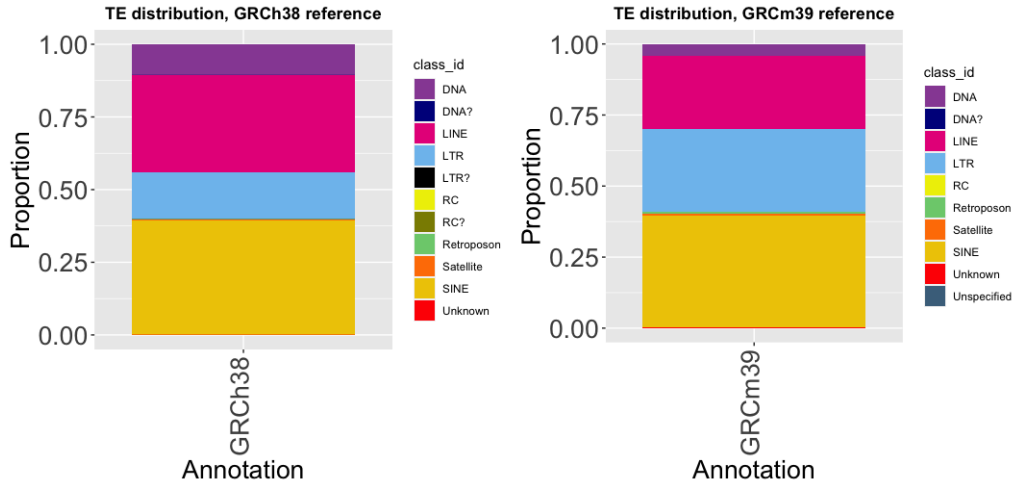
Indeed, considering changes within the LTR/ERV superfamily only, the older/intermediate age ERV1 family made up the majority across all cell lines (figure 4.14a and 4.14b). No other ERV families were consistently overrepresented in upregulated ERVs in this sub-family level analysis. ERV1 family elements made up at least 50% of upregulated ERVs across each of the HGSC lines, as compared to their proportion in the reference genome of roughly 25% (figure 4.14c). However, a stark difference was notable in the ID8 *in vivo* sample. Of the very few LTRs that were induced at the sub-family level, each one was of the ERVK family rather than the ERV1. Despite differences in the genomic distribution of ERV families in mouse genomes (figure 4.14c), ERVK were still notably overrepresented as compared to the reference, whilst no ERV1 family elements were detected. As anticipated, the oldest ERVL-MaLR elements made up a very small proportion of differentially expressed ERVs across all samples; these elements are the most likely to be permanently transcriptionally silenced, and uninducible [153].



(a)



(b)



(c)

Figure 4.12 **TE changes at sub-family level, HGSC samples.** Bar plots demonstrating transcriptional changes in TEs measured at the sub-family level in the the ID8 *in vivo* mouse model and the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines, following HKMTi-1-005 treatment. Caption continued overleaf.

Figure 4.12 (a) Significantly upregulated TEs (left) and downregulated TEs (right) across each sample, coloured by TE superfamily ("class_id"). (b) Same as in (a), but with TE superfamilies shown as a proportion of total up- (left) or down- (right) regulated TEs. (c) Distribution of TE superfamilies in the human GRCh38 reference genome (left) and the mouse GRCm39 reference genome (right), coloured by TE superfamily ("class_id"). TE superfamilies are labelled by colour according to the legend on the right-hand side ("class_id"). Classification: short-interspersed nuclear elements (SINE), long-interspersed nuclear elements (LINE), long terminal repeat elements (LTR), retroposons, DNA repeat elements (DNA), satellite repeats, Rolling Circle (RC). ? indicates uncertain classification.

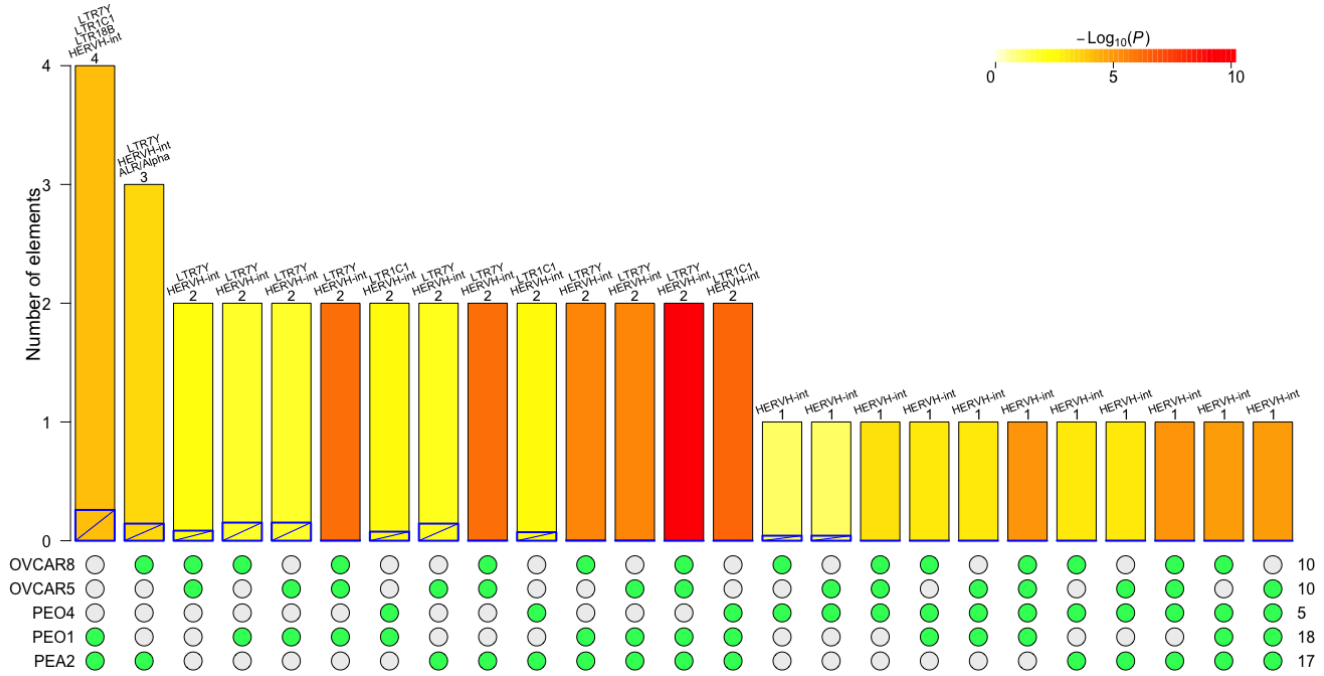
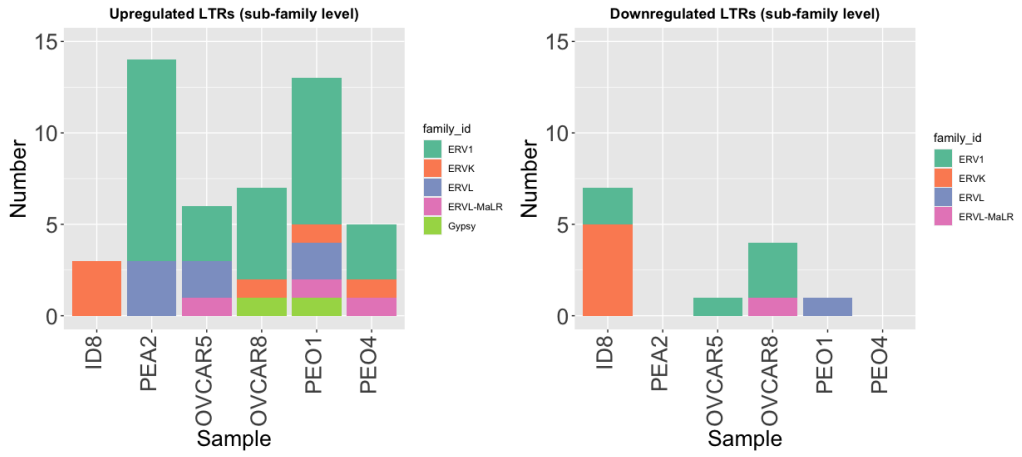
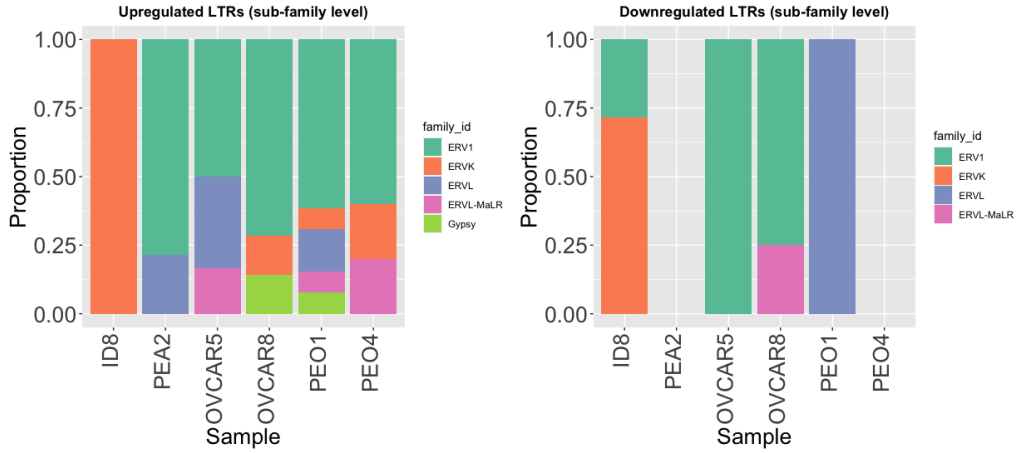


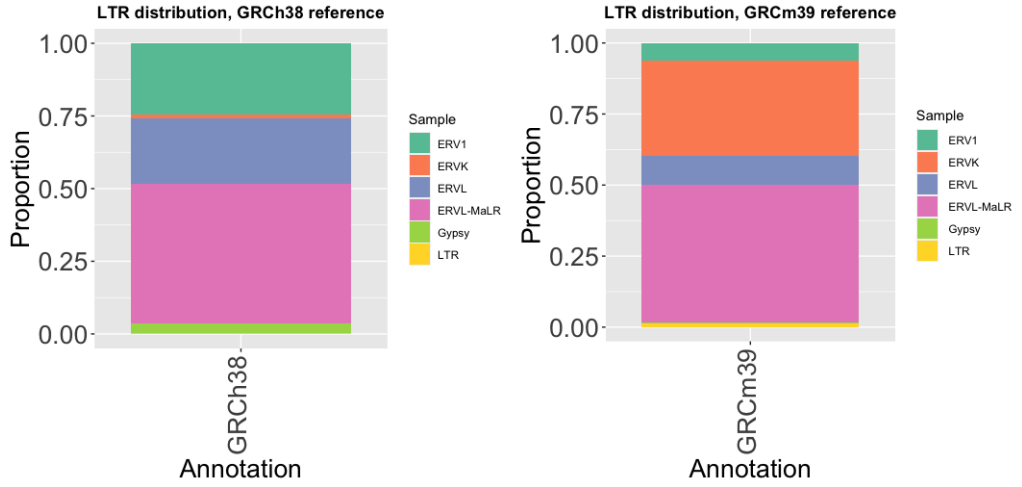
Figure 4.13 **Comparison of TEs induced in common across samples, sub-family level.** Statistical assessment of overlap in TEs induced at the sub-family level across the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines. Blue hatched bars represent the expected overlap between samples. TE sub-families in common are found above the corresponding bar. Total number of upregulated TE sub-families are found on the right-hand side of the figure for each cell line. Statistical test and visualisation performed using the SuperExactTest package [419].



(a)



(b)



(c)

Figure 4.14 **LTR/ERV changes at sub-family level, HGSC samples.** Bar plots demonstrating changes in LTRs/ERVs measured at the sub-family level in the the ID8 *in vivo* mouse model and the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines, following HKMTi-1-005 treatment. (a) Significantly upregulated ERVs (left) and downregulated ERVs (right) across each sample, coloured by ERV family ("family_id"). (b) Same as in (a), but with ERV families shown as a proportion of total up- (left) or down- (right) regulated ERVs. (c) Distribution of ERV families in the human GRCh38 reference genome (left) and the mouse GRCm39 reference genome (right), coloured by ERV family ("family_id").

4.3.2 De-repression of transposable element transcription: locus-specific analysis

Having assessed TEs at the sub-family level, I next analysed changes at the locus-specific level. This allows for distinction between TEs of the same sub-family at different genomic loci that may show distinct transcriptional effects. The Telescope tool [384] was used to investigate locus-specific changes. Telescope uses a very similar approach to TEtranscripts [383] including implementation of an expectation-maximisation algorithm to address multimapping reads. However, an additional parameter is included to correct for ambiguous read alignments [384], allowing each read to be assigned to its most probable TE locus in place of summing all reads at the sub-family level.

In each sample, many more differentially expressed TEs were detected at the locus-specific level than at the sub-family level, with a greater range of fold-changes (figures 4.15, 4.16 and 4.17a). Between 44 and 2490 TE loci were upregulated, with a greater number of up- compared to down-regulated loci in every sample except OVCAR8 (figures 4.17a, 4.19a, and 4.19b). The overlap of significantly upregulated TEs was statistically significant between most samples (figure 4.18). Particularly, PEO1 and PEO4 lines showed strong concordance in upregulated TEs at the locus-specific level, as did OVCAR5 and OVCAR8 lines. Interestingly, there was a markedly higher number of upregulated TEs in the ID8 *in vivo* sample as compared to the HGSC cell lines using locus-specific analyses, despite having the lowest count at the sub-family level. This effect suggests differential expression of many individual TE loci, but with little concordance in differential expression across TE sub-families in the sample.

As compared to the results obtained at the sub-family level, there was a less evident skew towards LTR/ERV superfamily elements amongst upregulated TEs in this locus-specific analysis (figure 4.19c). However, all samples demonstrated many changes at LTR/ERV loci, and all cell lines showed a greater number and proportion of up- rather than down-regulated LTR/ERV loci (figures 4.17b and 4.19c). This included the OVCAR8 sample, where total downregulated TEs outnumbered upregulated TEs. Most samples also still demonstrated a greater proportion of LTR/ERV superfamily changes than the reference distribution (figure 4.19c, reference figure 4.12c).

Amongst the upregulated LTRs/ERVs, locus-specific analyses corroborated the skew towards the older/intermediate age ERV1 family elements observed at sub-family level (figure 4.20). In all samples, a greater proportion of upregulated LTRs/ERVs were of the ERV1 family than in the reference (figure 4.20c, reference figure 4.14c). As with sub-family analyses, this effect was less evident in the ID8 *in vivo* sample. Here, ERVK made up a larger proportion of upregulated ERVs than expected based on the reference (figure 4.20c, reference figure 4.14c). However, ERV1 elements did still make up more than the reference in the ID8 *in vivo* sample, with the fewest changes in the

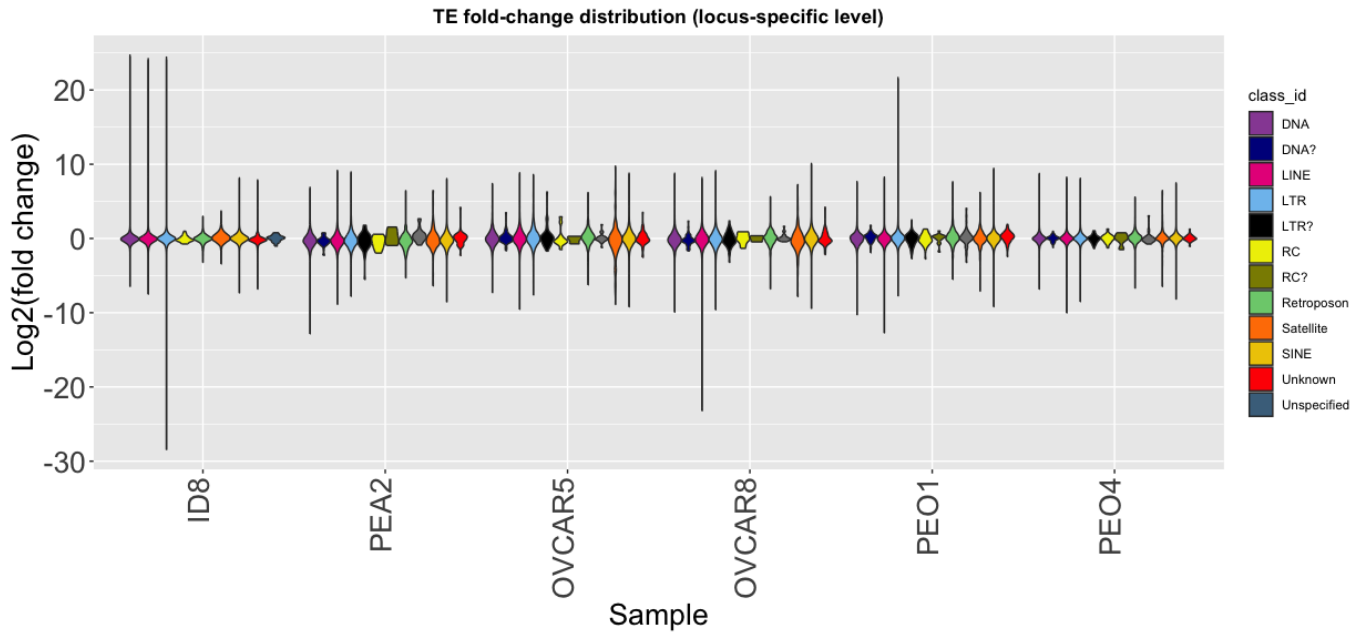


Figure 4.15 **TE expression changes by superfamily, at the locus-specific level.** Distribution of fold-changes of TEs following HKMTi-1-005 treatment, across the the ID8 *in vivo* mouse model and the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines. TE differential expression measured locus-specific level, using the Telescope tool [384]. TE superfamilies are labelled by colour according to the legend on the right-hand side ("class_id"). Classification: short-interspersed nuclear elements (SINE), long-interspersed nuclear elements (LINE), long terminal repeat elements (LTR), retroposons, DNA repeat elements (DNA), satellite repeats, Rolling Circle (RC). ? indicates uncertain classification.

older ERVL-MaLR elements.

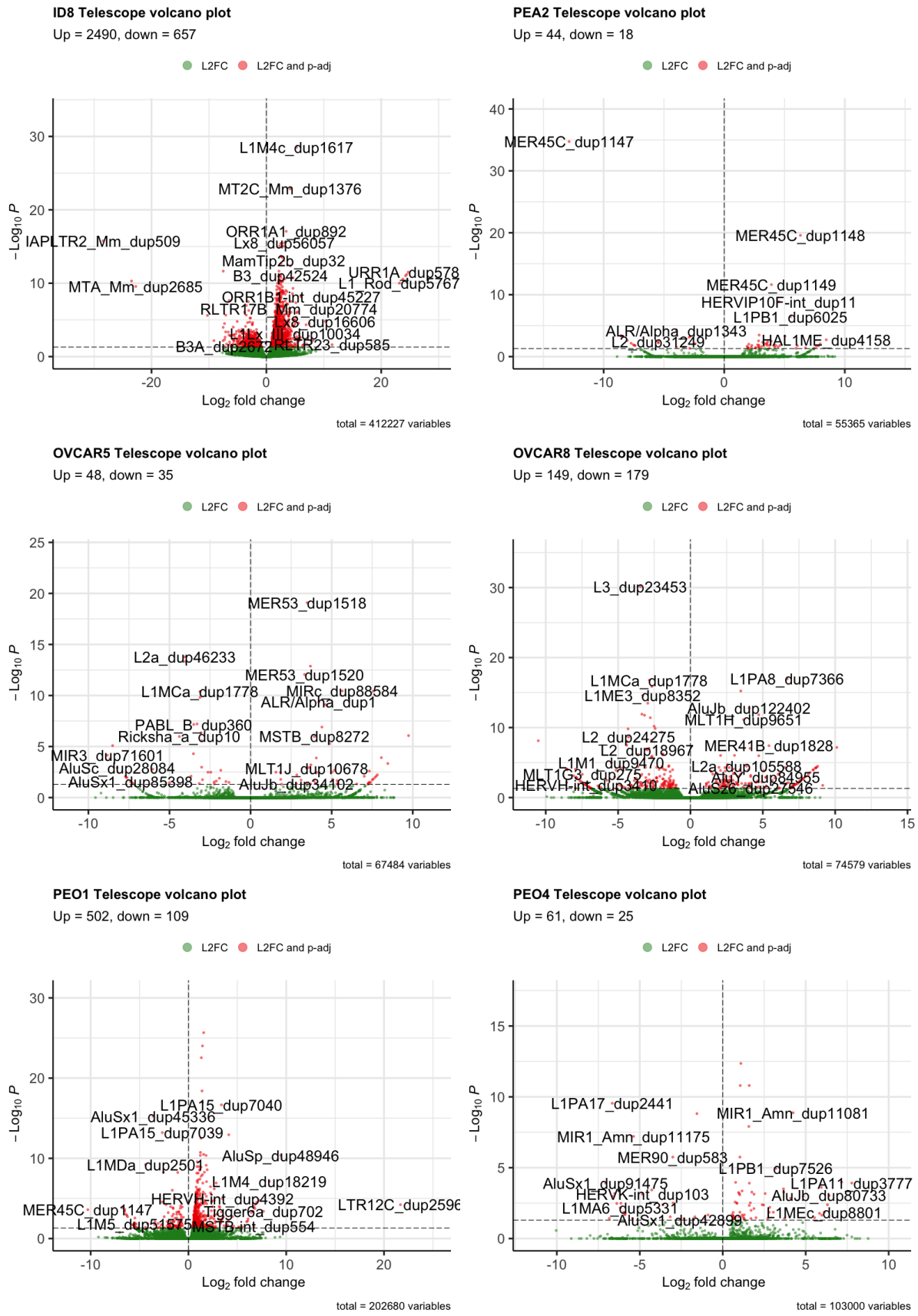
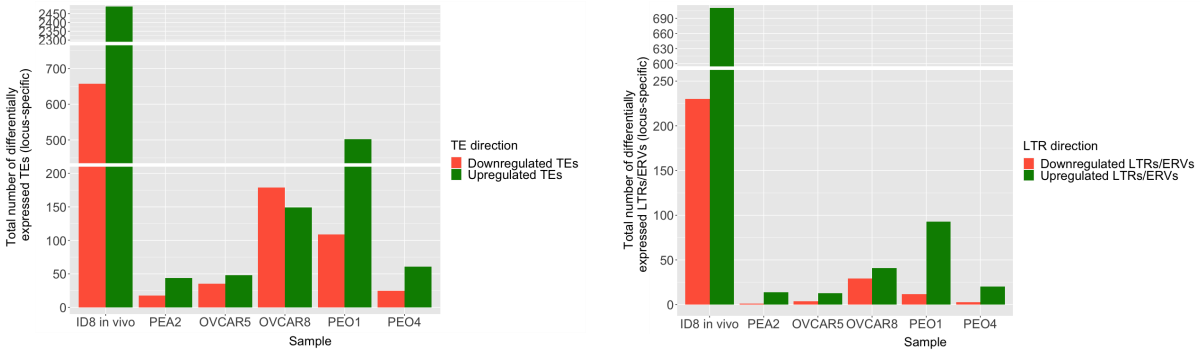


Figure 4.16 **Volcano plot of TE changes in HGSC samples, locus-specific level.** Volcano plot examining differential expression of TEs at the locus-specific level, in control versus HKMTi-1-005-treated conditions, using the Telescope tool [384]. Samples include (top to bottom, left to right): the ID8 *in vivo* mouse model and the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines. TEs are labelled by colour, based on significance and fold-change: log2FC (green) = absolute FC >0, log2FC and p-adj (red) = absolute log2FC > 0 and p-adj < 0.05. The "_dup" label identifies the specific TE locus being described, from the many loci within any particular sub-family.



(a) (b)

Figure 4.17 **Total up- versus down-regulated TEs and LTRs/ERVs in HGSC samples, locus-specific level.** Bar plots showing total number of up- versus down-regulated (a) TEs or (b) LTRs/ERVs across the the ID8 *in vivo* mouse model and the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines, in control versus HKMTi-1-005-treated conditions. TEs measured at the locus-specific level, with a threshold of p-adj < 0.05 for differential expression. Red = downregulated, green = upregulated.

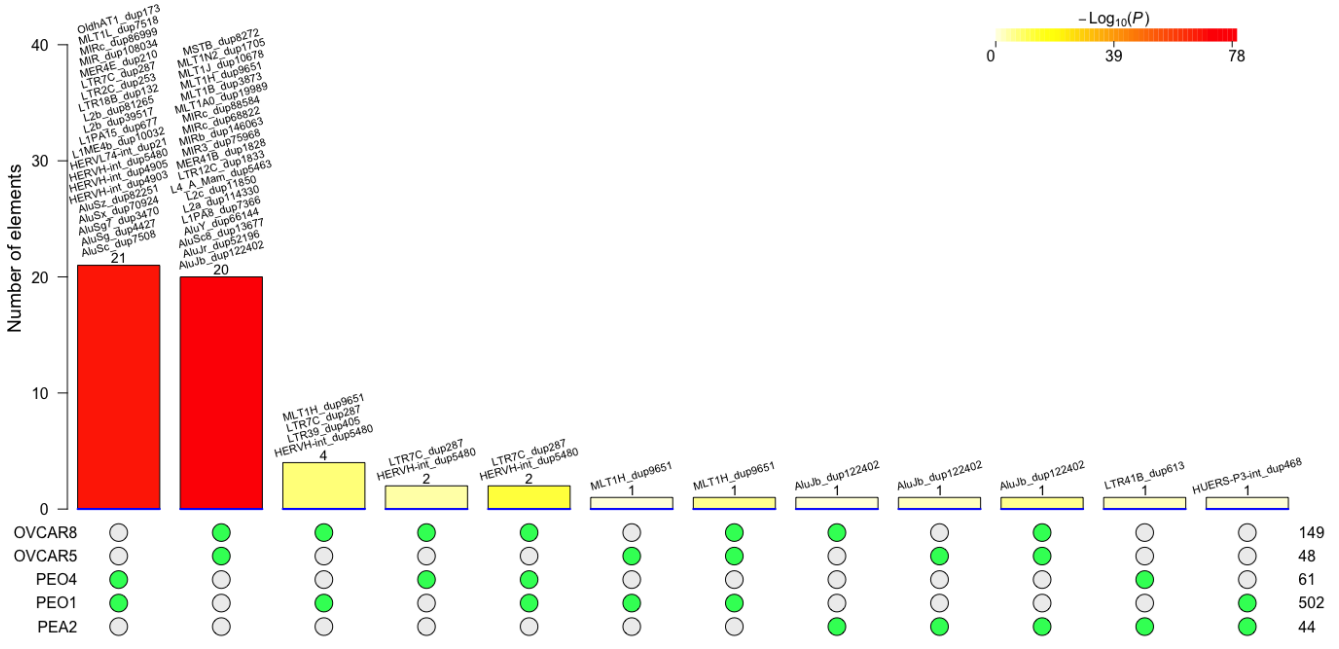
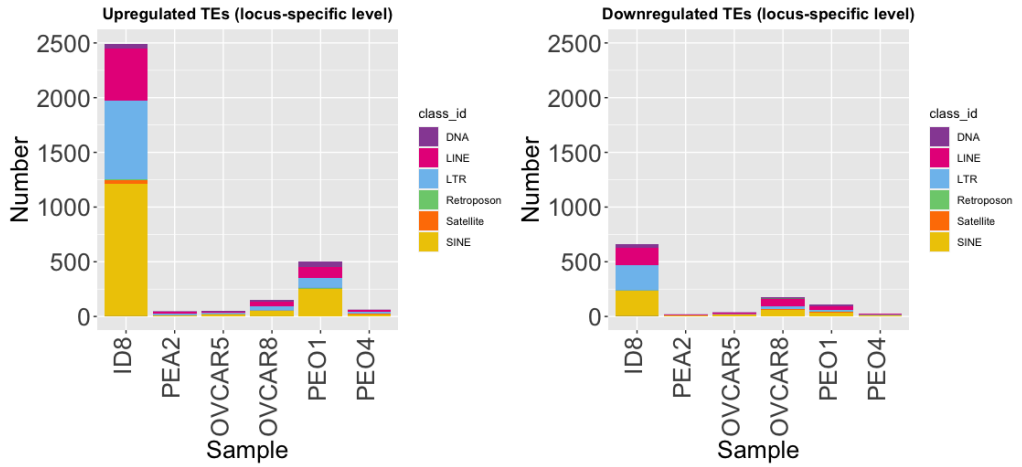
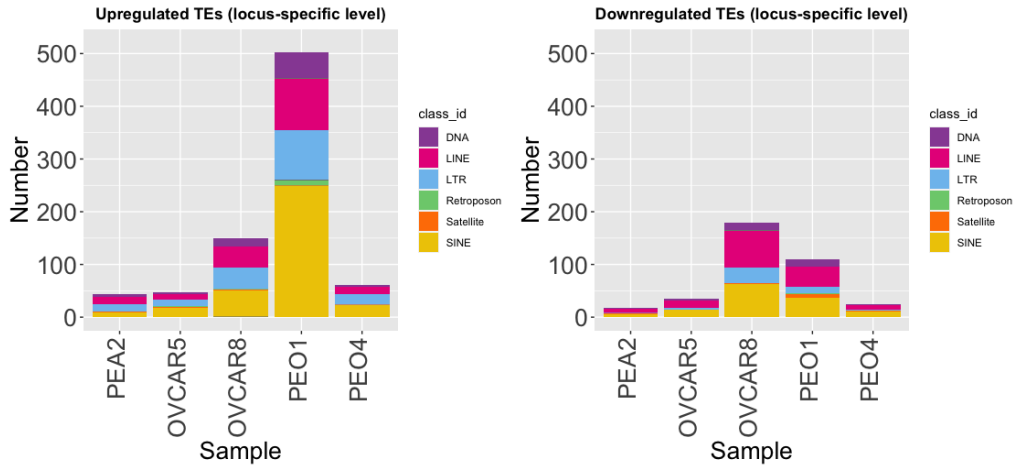


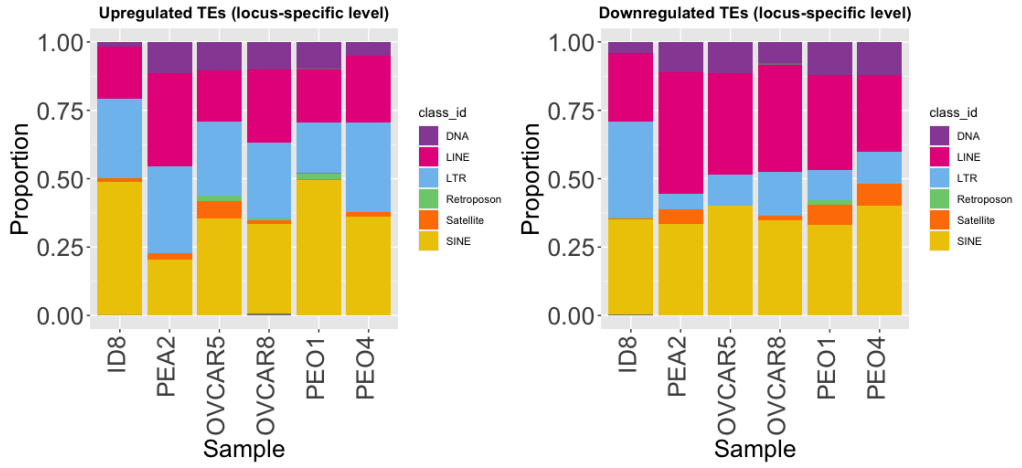
Figure 4.18 **Comparison of TEs induced in common across samples, locus-specific level.** Statistical assessment of overlap in TEs induced at the locus-specific level across the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines. Blue hatched bars represent the expected overlap between samples. TE loci in common are found above the corresponding bar. Total number of upregulated TE loci are found at the right-hand side, for each cell line. Statistical test and visualisation performed using the SuperExactTest package [419].



(a)

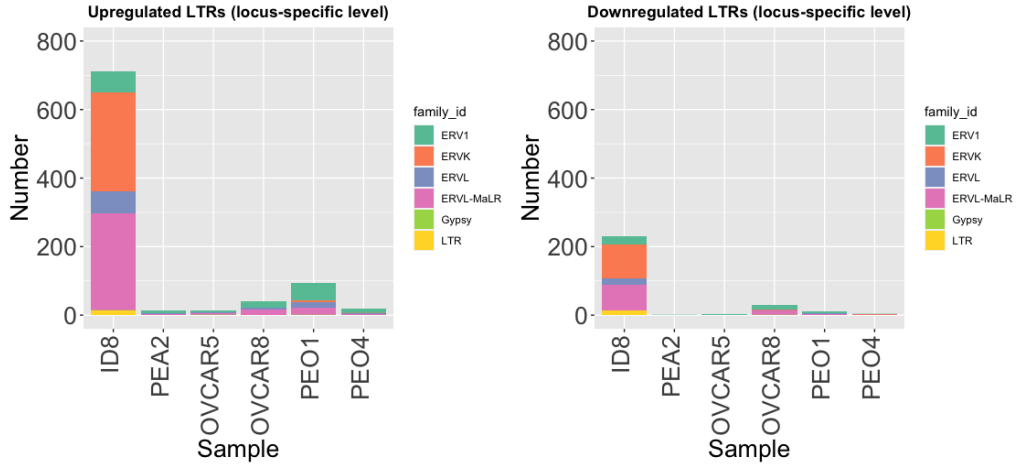


(b)

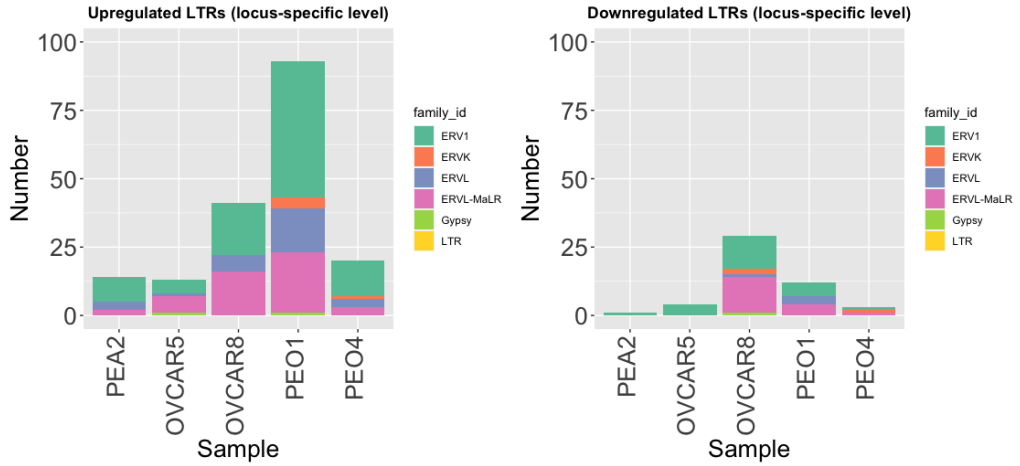


(c)

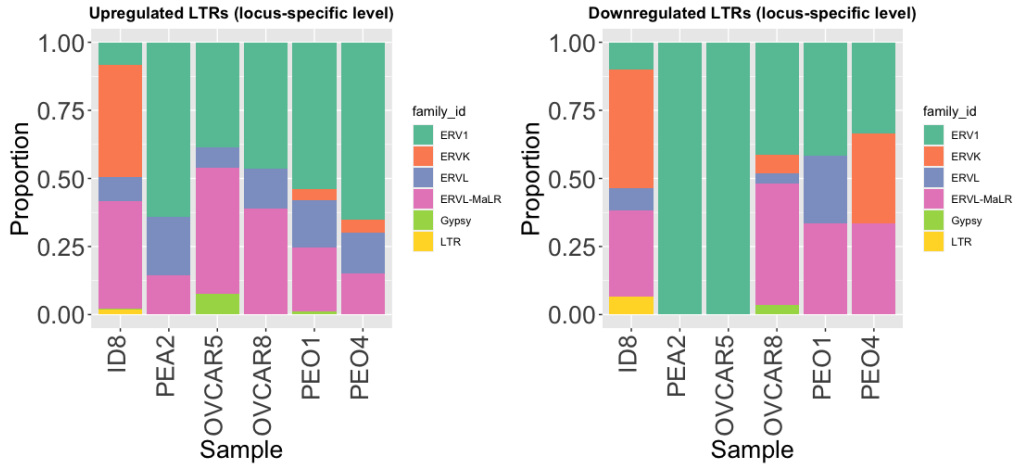
Figure 4.19 **TE changes at locus-specific level, HGSC samples.** Bar plots demonstrating changes in TEs measured at the locus-specific level in the the ID8 *in vivo* mouse model and the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines, following HKMTi-1-005 treatment. (a) Significantly upregulated TEs (left) and downregulated TEs (right) across each sample, coloured by TE superfamily ("class_id"). (b) Same as in (a), but excluding the ID8 *in vivo* sample to allow clearer visualisation of HGSC cell line effects. (c) Same as in (a), but with TE superfamilies shown as a proportion of total up- (left) or down- (right) regulated TEs.



(a)



(b)



(c)

Figure 4.20 **LTR/ERV changes at locus-specific level, HGSC samples.** Bar plots demonstrating changes in LTRs/ERVs measured at the locus-specific level in the the ID8 *in vivo* mouse model and the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines, following HKMTi-1-005 treatment. (a) Significantly upregulated ERVs (left) and downregulated ERVs (right) across each sample, coloured by ERV family ("family_id"). (b) Same as in (a), but excluding the ID8 *in vivo* sample to allow clearer visualisation of effects in the HGSC cell lines. (c) Same as in (a), but with ERV families shown as a proportion of total up- (left) or down- (right) regulated ERVs.

4.3.3 Increased accessibility of transposable elements

The data presented demonstrate upregulation of TEs at the transcriptional level. To accompany these findings, I investigated the changes in chromatin accessibility of TEs. The two ATACseq analysis methods described in chapter 3 were utilised: the DAC and subtraction approaches (see figure 3.15 for explanatory diagram). The ID8 *in vitro* samples – BRCA1, BRCA2 and F3 – were included alongside the PEO1 and PEO4 HGSC cell lines and the ID8 *in vivo* sample.

I first used the DAC approach. This indicated significantly increased accessibility of many TEs in each sample (figure 4.21). The effect was most evident in the ID8 *in vivo* sample, where over 49,000 upregulated DACs were associated with TE loci. However, there were not consistently more up- than downregulated DACs across all TEs, or across LTRs/ERVs separately (figure 4.22 and 4.23). The ID8 *in vivo* sample and ID8 *in vitro* BRCA1 and F3 samples showed the expected pattern of increased accessibility at TEs/LTRs, whilst PEO1, PEO4 and ID8 BRCA2 had a greater number of downregulated DACs at TE and LTR loci following treatment.

The more accessible TEs were of mixed superfamilies (figure 4.24 and 4.25). Whilst LTRs/ERVs were not markedly overrepresented, SINEs were overrepresented as compared to the reference distribution in almost all samples (see figure 4.12c for reference). The ID8 *in vivo* sample was the exception to both trends, with a strong skew towards upregulation of LTRs/ERVs over other TE superfamilies and with fewer more accessible SINEs than the reference distribution. Within the more accessible LTRs/ERVs, the expected trends were observed: upregulated DACs demonstrated overrepresentation of the older/intermediate aged ERV1 family in most samples, including in the ID8 *in vitro* samples where the oldest ERVL-MaLR elements also appeared to be modestly overrepresented (figure 4.26 and 4.27). The ID8 *in vivo* sample was once again an anomaly, with a far higher proportion of upregulated DACs at ERVK family TEs than expected based on the reference.

At the individual locus level, overlaps between more accessible TEs and transcriptionally upregulated TEs were assessed where the relevant RNAseq data were available (PEO1, PEO4 and ID8 *in vivo* samples). These analyses demonstrated that overall, there were very few overlaps between the TEs that were transcriptionally upregulated and those that gained chromatin accessibility using this method, when considering individual loci. However, for the PEO1 and PEO4 cell lines, this modest overlap still reached statistical significance (appendix table A.6).

I next used the subtraction approach to investigate unique HKMTi-1-005 treatment peaks of accessibility (figure 4.28 and 4.29). These demonstrated similar results to the DAC approach; SINEs were often overrepresented as compared to the reference, and ERV1 family were overrepresented amongst LTR/ERV superfamily TEs, with the ID8 *in vivo* sample as the exception to both trends. There was a significant overlap in all three PEO1, PEO4 and ID8 *in vivo* samples between TEs with increased accessibility using this approach and transcriptionally upregulated TEs (appendix table

A.6).

Taken together, these data suggest significantly increased accessibility accompanies transcriptional upregulation of TEs at only a small subset of TEs. Whilst the relationship between accessibility and transcription of TEs was statistically significant, the vast majority of TEs that were transcriptionally upregulated were not more accessible using either the DAC or subtraction approach. However, despite limited overlap between accessibility and transcriptional upregulation at specific TE loci, the trends observed in accessibility over different ERV families mimic the effects observed at the transcriptional level: HKMTi-1-005 effects are skewed towards the older/intermediate age ERV1 family elements, over the younger or older ERVK and ERVL families respectively.

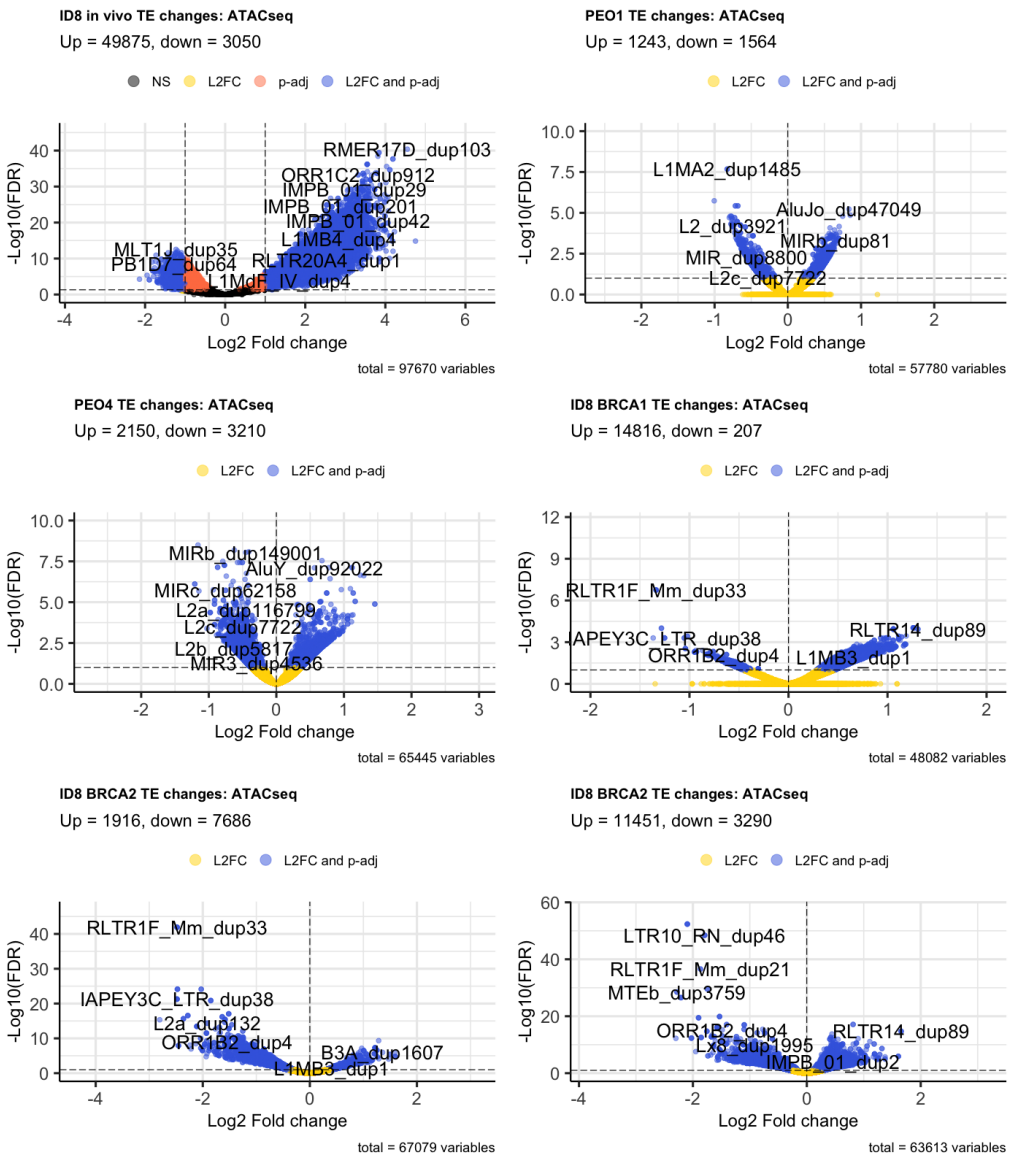
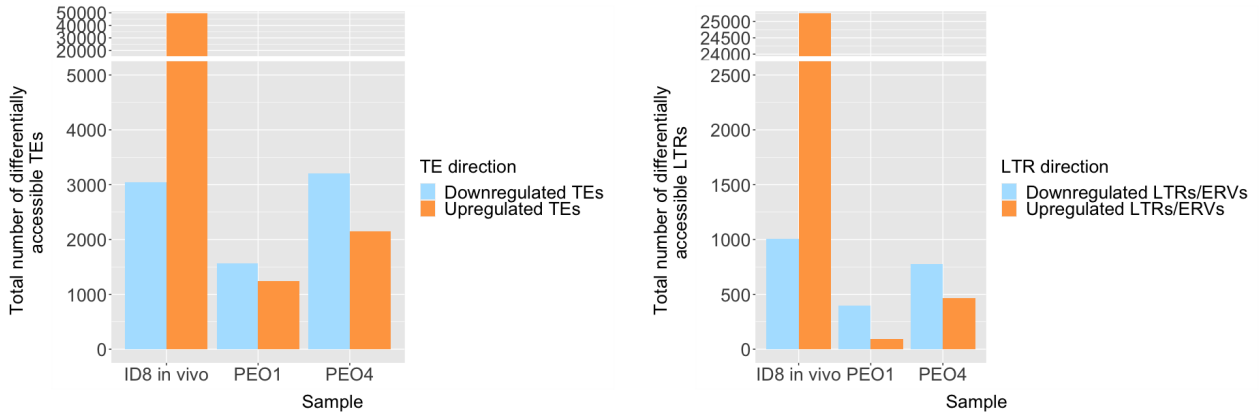
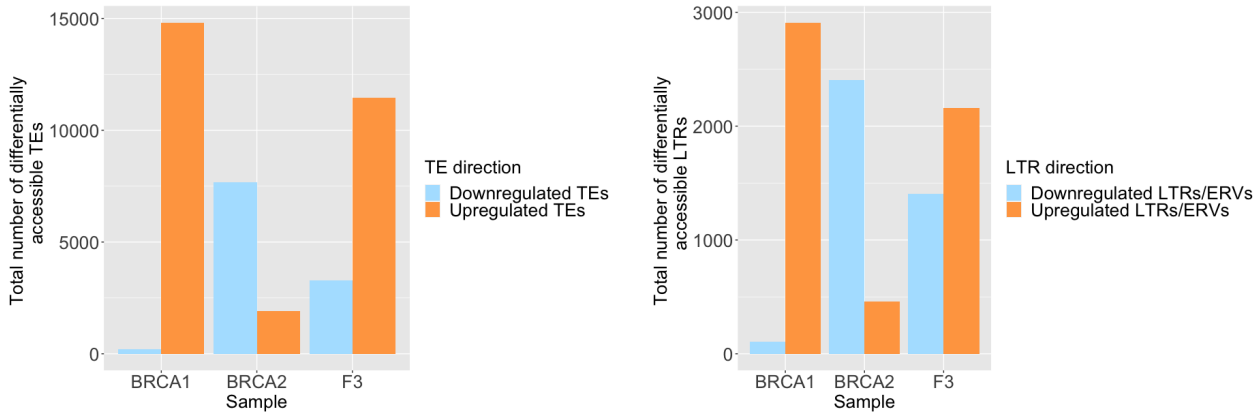


Figure 4.21 **Volcano plot of TE accessibility in HGSC samples.** Volcano plot examining differential accessibility of TEs as measured using DiffBind [390] following HKMTi-1-005 treatment. Samples include (left to right, top to bottom): ID8 *in vivo* mouse model, PEO1, PEO4, BRCA1 ID8, BRCA2 ID8, and F3 ID8 cell lines. TEs are labelled by colour, based on significance and fold-change: NS (black) = non-significant, L2FC (yellow) = absolute fold-change > 1 (absolute log2FC > 0), p-adj (orange) = p-adj < 0.05, L2FC and p-adj (blue) = absolute L2FC > 0 and p-adj < 0.05. For the ID8 *in vivo* sample, an additional absolute log2FC threshold of > 1 is applied.



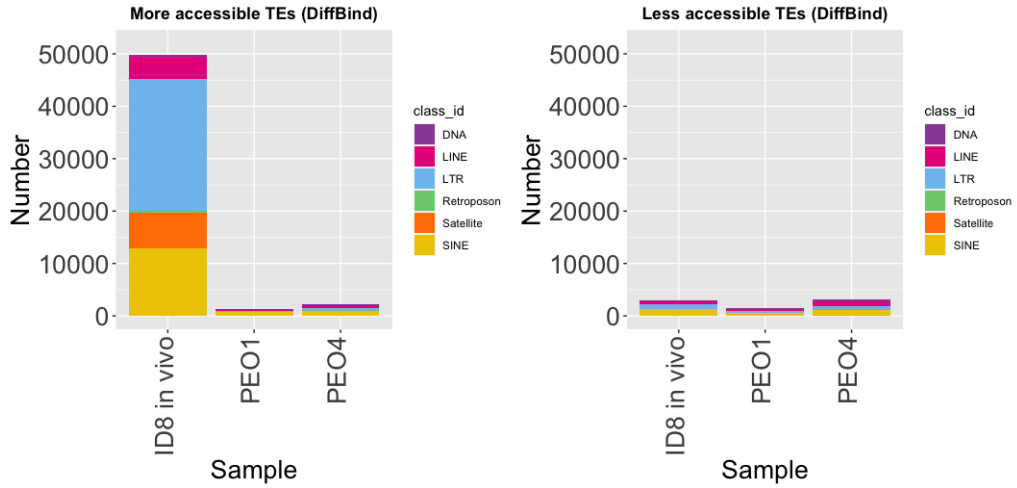
(a) (b)

Figure 4.22 **Total up- versus downregulated DACs at TEs and LTRs/ERVs in PEO1, PEO4 and ID8 *in vivo* samples.** Bar plots showing total number of more versus less accessible (a) TEs or (b) LTRs/ERVs across the PEO1 and PEO4 cell lines and the ID8 *in vivo* mouse model, in control versus HKMTi-1-005-treated conditions. Accessibility was measured using DiffBind with subsequent annotation of TEs, and differentially accessible TEs were defined using a threshold of p-adj < 0.05. For the ID8 *in vivo* sample, an additional absolute log2FC threshold of > 1 is applied.

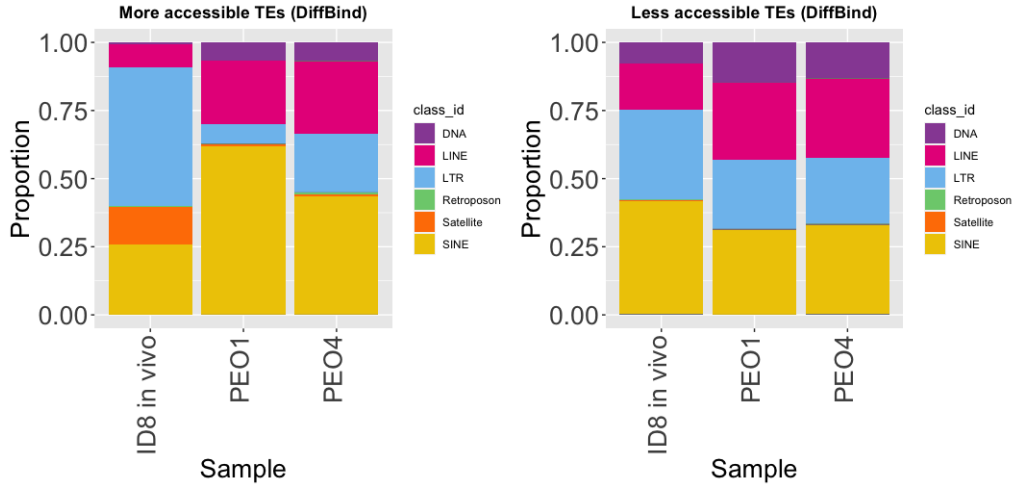


(a) (b)

Figure 4.23 **Total up- versus downregulated DACs at TEs and LTRs/ERVs in ID8 *in vitro* samples.** Bar plots showing total number of more versus less accessible (a) TEs or (b) LTRs/ERVs across the *in vitro* BRCA1, BRCA2 and F3 ID8 lines, in control versus HKMTi-1-005-treated conditions. Accessibility was measured using DiffBind with subsequent annotation of TEs, and differentially accessible TEs were defined using a threshold of p-adj < 0.05.

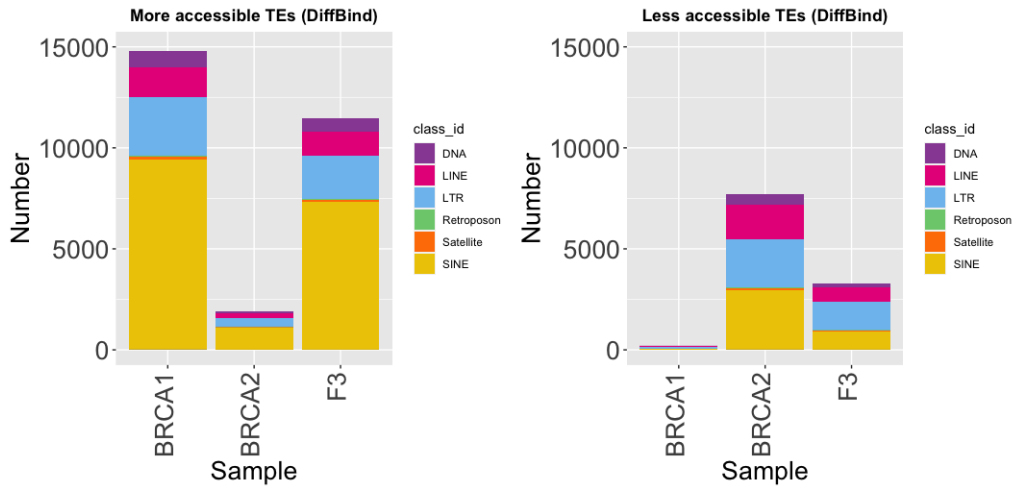


(a)

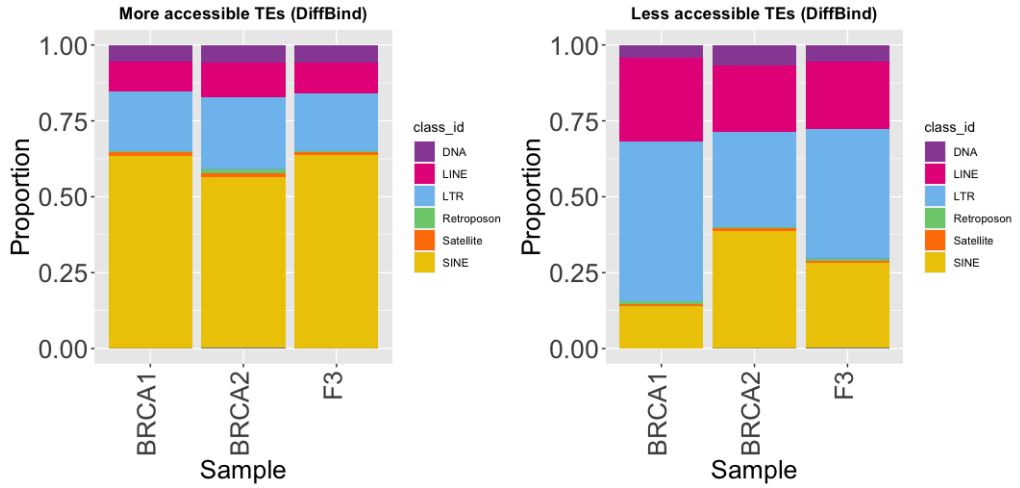


(b)

Figure 4.24 **Accessibility changes across TE superfamilies, in PEO1, PEO4 and ID8 *in vivo* samples: DAC method.** Bar plots demonstrating chromatin accessibility changes in TEs in the PEO1 and PEO4 cell lines and ID8 *in vivo* samples, following HKMTi-1-005 treatment. (a) More (left) and less (right) accessible TEs, coloured by TE superfamily ("class_id"). (b) Same as in (a), but with TE superfamilies shown as a proportion of total more (left) or less (right) accessible TEs. Threshold for differential accessibility is p-adj < 0.05. For the ID8 *in vivo* sample, an additional absolute log2FC threshold of > 1 is applied.

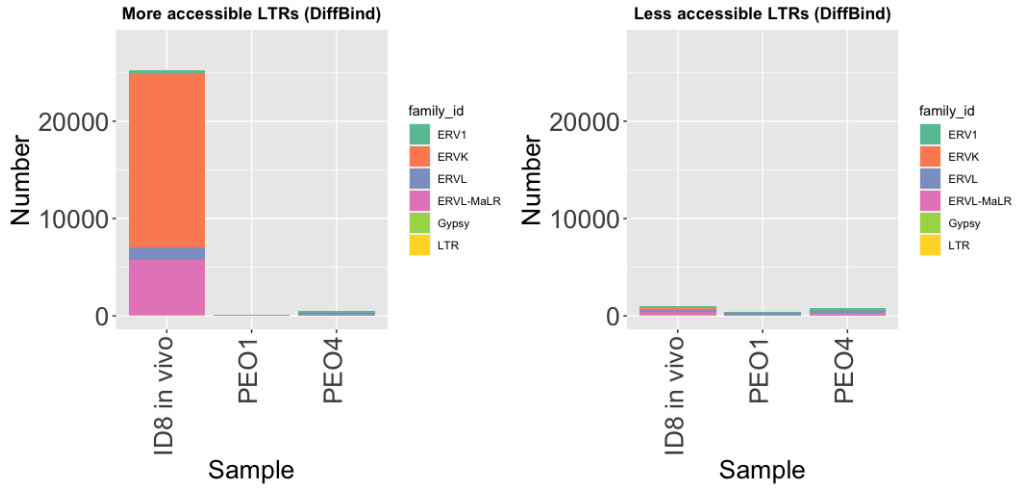


(a)

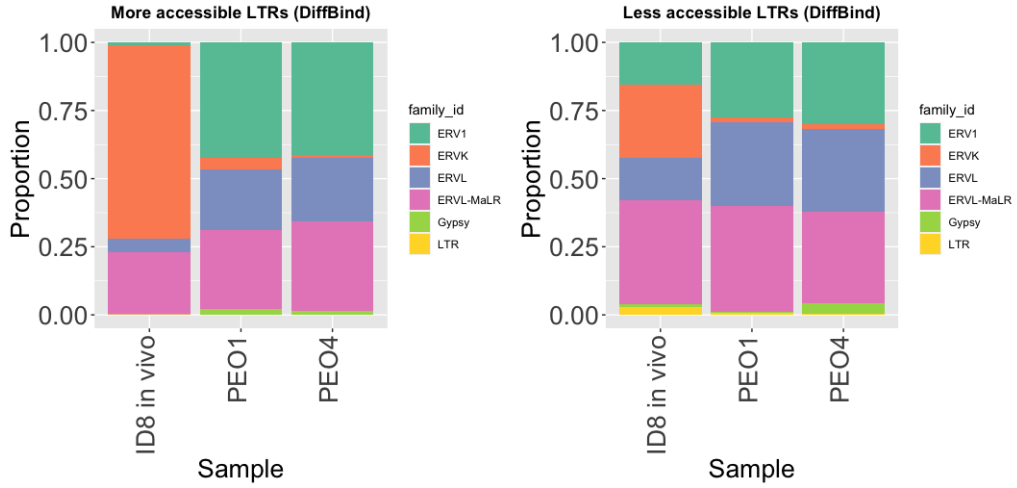


(b)

Figure 4.25 **Accessibility changes across TE superfamilies, in ID8 *in vitro* samples: DAC method.** Bar plots demonstrating chromatin accessibility changes in TEs in the *in vitro* BRCA1, BRCA2 and F3 ID8 samples, following HKMTi-1-005 treatment. (a) More (left) and less (right) accessible TEs, coloured by TE superfamily ("class_id"). (b) Same as in (a), but with TE superfamilies shown a proportion of total more (left) or less (right) accessible TEs. Threshold for differential accessibility is p-adj < 0.05.

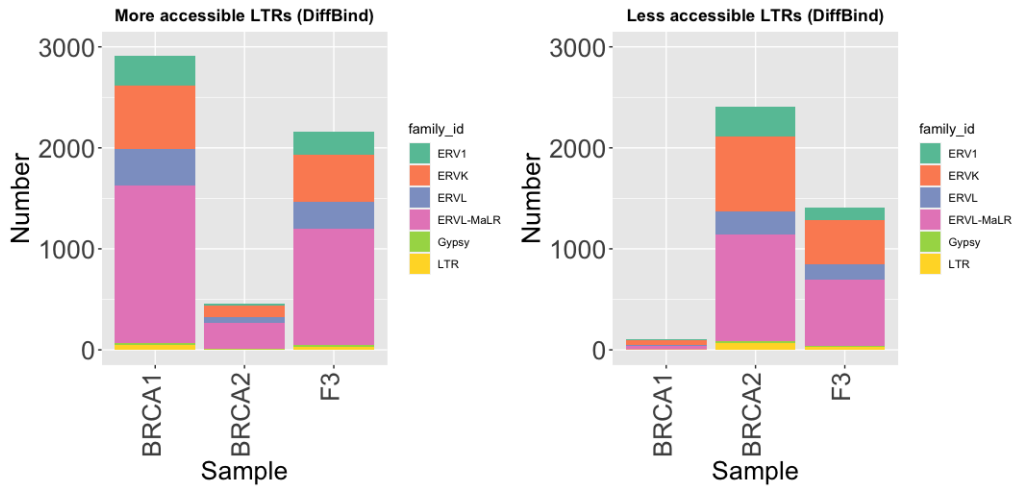


(a)

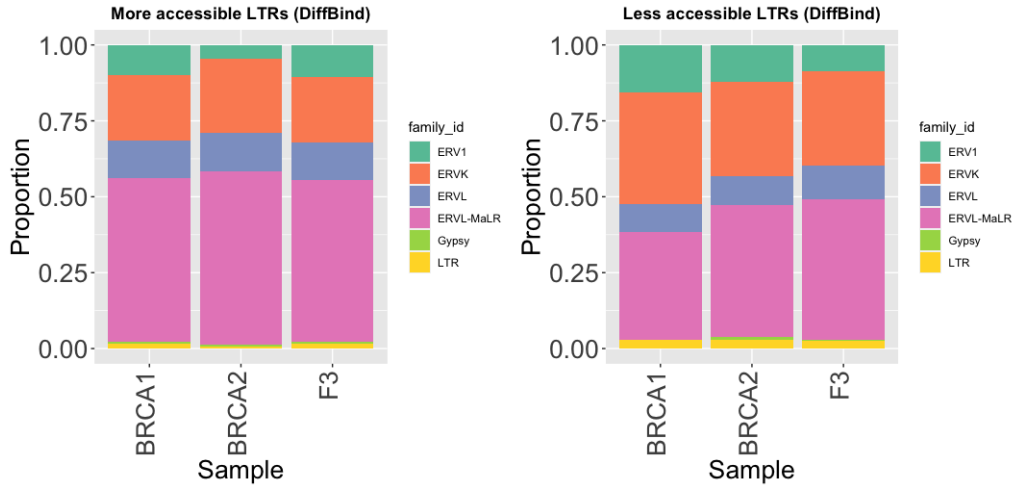


(b)

Figure 4.26 **Accessibility changes across LTR/ERV families, in PEO1, PEO4 and ID8 *in vivo* samples: DAC method.** Bar plots demonstrating chromatin accessibility changes in LTRs/ERVs in the PEO1 and PEO4 cell lines and ID8 *in vivo* samples, following HKMTi-1-005 treatment. (a) More (left) and less (right) accessible ERVs, coloured by ERV family ("family_id"). (b) Same as in (a), but with ERV families shown as a proportion of total more (left) or less (right) accessible ERVs. Threshold for differential accessibility is p-adj < 0.05. For the ID8 *in vivo* sample, an additional absolute log2FC threshold of > 1 is applied.



(a)



(b)

Figure 4.27 **Accessibility changes across LTR/ERV families, in ID8 *in vitro* samples: DAC method.** Bar plots demonstrating chromatin accessibility changes in LTRs/ERVs in the *in vitro* BRCA1, BRCA2 and F3 ID8 samples, following HKMTi-1-005 treatment. (a) More (left) and less (right) accessible ERVs, coloured by ERV family ("family_id"). (b) Same as in (a), but with ERV families shown as a proportion of total more (left) or less (right) accessible ERVs. Threshold for differential accessibility is p-adj < 0.05.

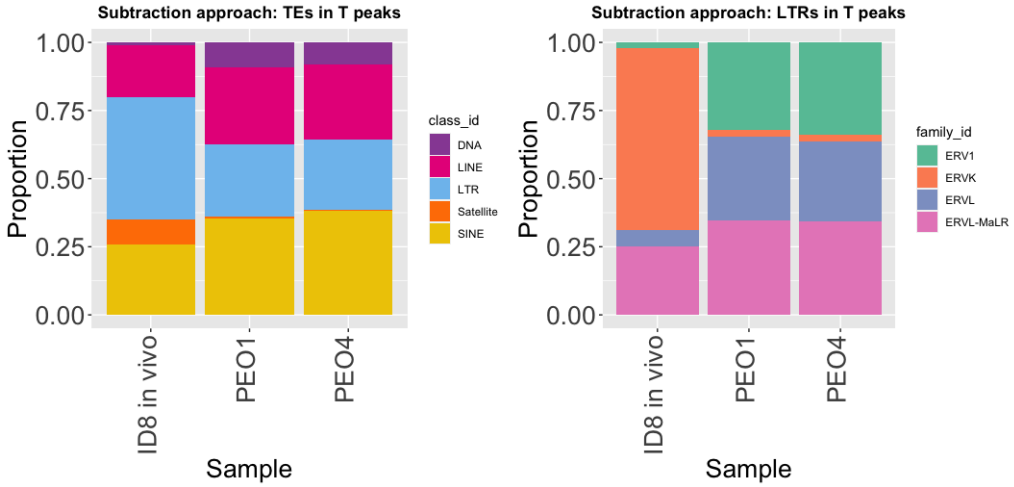


Figure 4.28 **Accessibility changes across TE superfamilies and LTR/ERV families, in PEO1, PEO4 and ID8 *in vivo* samples: subtraction method.** Bar plots demonstrating chromatin accessibility peaks unique to the HKMTi-1-005 treatment condition at TEs in the PEO1 and PEO4 cell lines and ID8 *in vivo* samples (subtraction approach). Left: TEs with accessibility peaks unique to the HKMTi-1-005 treatment condition, coloured by TE superfamily ("class_id"). Right: LTRs/ERVs with accessibility peaks unique to the HKMTi-1-005 treatment condition, coloured by ERV family ("family_id").

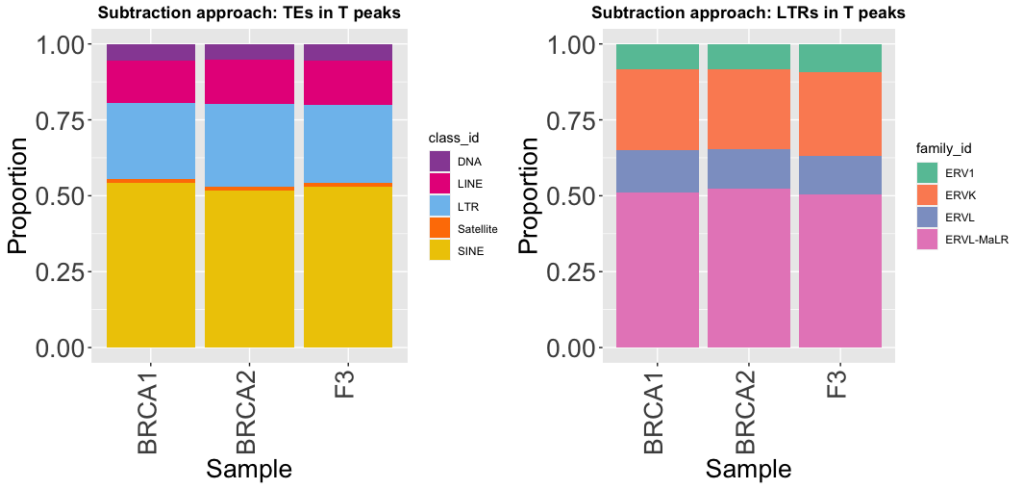


Figure 4.29 **Accessibility changes across TE superfamilies and LTR/ERV families, in ID8 *in vitro* samples: subtraction method.** Bar plots demonstrating chromatin accessibility peaks unique to the HKMTi-1-005 treatment condition at TEs in the BRCA1, BRCA2 and F3 ID8 cell lines (subtraction approach). Left: TEs with accessibility peaks unique to the HKMTi-1-005 treatment condition, coloured by TE superfamily ("class_id"). Right: LTRs/ERVs with accessibility peaks unique to the HKMTi-1-005 treatment condition, coloured by ERV family ("family_id").

4.4 Investigating the aberrant PEO4 response

4.4.1 A high basal immune signalling state

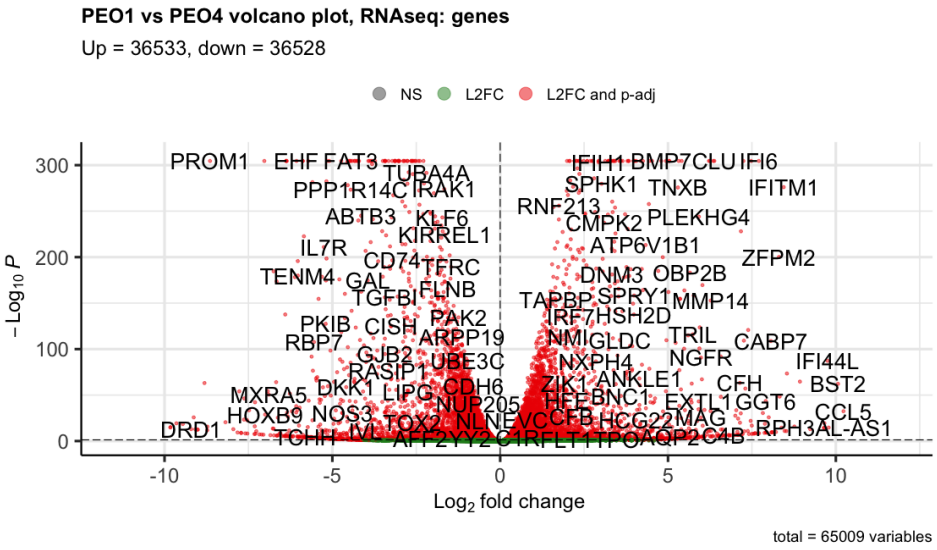
PEO1 and PEO4 lines were originally derived from the same patient, before and after resistance to platinum therapy had emerged [368]. Yet, their responses to HKMTi-1-005 treatment were highly dissimilar with regards to immune signalling (see 4.2.1). Hence, I performed a direct comparison between the transcriptional profiles of PEO1 and PEO4 cell lines before treatment, to determine whether these differences may provide insight into the failure of immune stimulation in PEO4 in response to HKMTi-1-005. In particular, I examined basal ERV/ISG status in these two cell lines.

There were extensive differences in gene expression between PEO1 and PEO4 (figure 4.30a). Many of the genes upregulated in the PEO4 line were immune-related, including almost every one of the VM ISGs described above (figure 4.30b), and every gene in the 7ISG set (figure 4.30c). This observation was corroborated using GSEA: the top 3 upregulated pathways in PEO4 were 'response to type I interferon', 'response to virus', and 'defense response to other organism' (table 4.2). Several other immune pathways including IFNA or IFNG response, NFKB signalling and regulation of innate immune response were also significantly upregulated in PEO4.

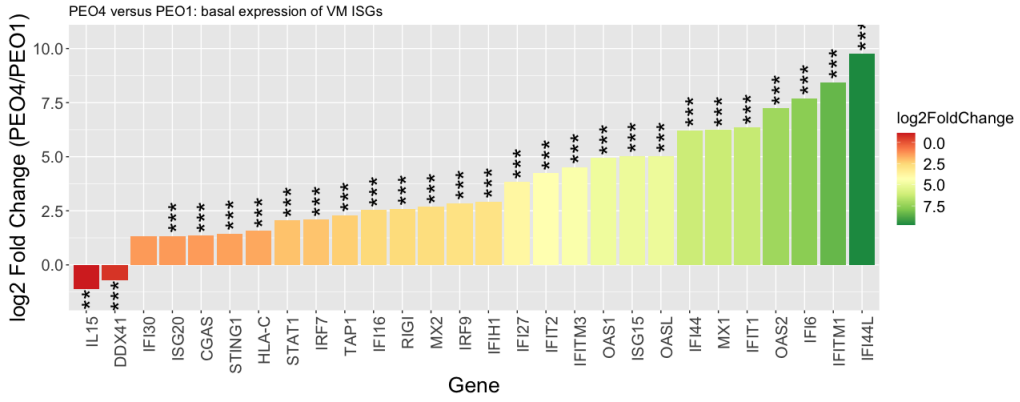
Gene Set	Description	Size	Leading Edge Number	ES	NES	P Value	FDR
GO:0034340	response to type I interferon	74	39	0.4538	4.5985	<2.2e-16	<2.2e-16
GO:0009615	response to virus	283	125	0.2237	4.3175	<2.2e-16	<2.2e-16
GO:0098542	defense response to other organism	376	117	0.1718	3.8015	<2.2e-16	<2.2e-16
GO:0006732	coenzyme metabolic process	341	155	0.176	3.754	<2.2e-16	<2.2e-16
GO:0009100	glycoprotein metabolic process	375	143	0.166	3.7022	<2.2e-16	<2.2e-16

Table 4.2 **PEO1 versus PEO4: top GSEA results.** The top 5 GOBP pathways upregulated in PEO4 as compared to PEO1, calculated using WebGestalt [402].

Subsequently, baseline expression of VM ISGs across all lines was assessed. Expression was determined using the median of ratios (MoR) normalisation method from DESeq2 [380]. This method accounts for differences in sequencing depth and RNA composition but not for differences in gene length, therefore allows comparison of expression for the same gene across samples, but not for comparison between different genes. The analysis demonstrated a striking difference in expression of almost all VM genes in the PEO4 sample at baseline: in many instances, expression of VM genes was several orders of magnitude higher in PEO4 compared to other cell lines (figure 4.31).



(a)



(b)



(c)

Figure 4.30 **Comparison of PEO1 and PEO4 at baseline: gene expression.** (a) Volcano plot demonstrating gene expression in PEO4 as compared to PEO1. Genes are labelled by colour, based on significance and fold-change: NS (grey) = not significant, L2FC (green) = absolute fold-change > 1 (absolute log2FC > 0), p-adj (blue) = adjusted p-value < 0.05, L2FC and p-adj (red) = both L2FC and p-adj criteria met. (b) Difference in expression of the ISGs defined in [439] between PEO1 and PEO4. (c) Difference in expression between PEO1 and PEO4 of the 7ISG subset of the 28 VM ISGs. For (b) and (c): * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001. In each case, PEO1 is the "control" sample.

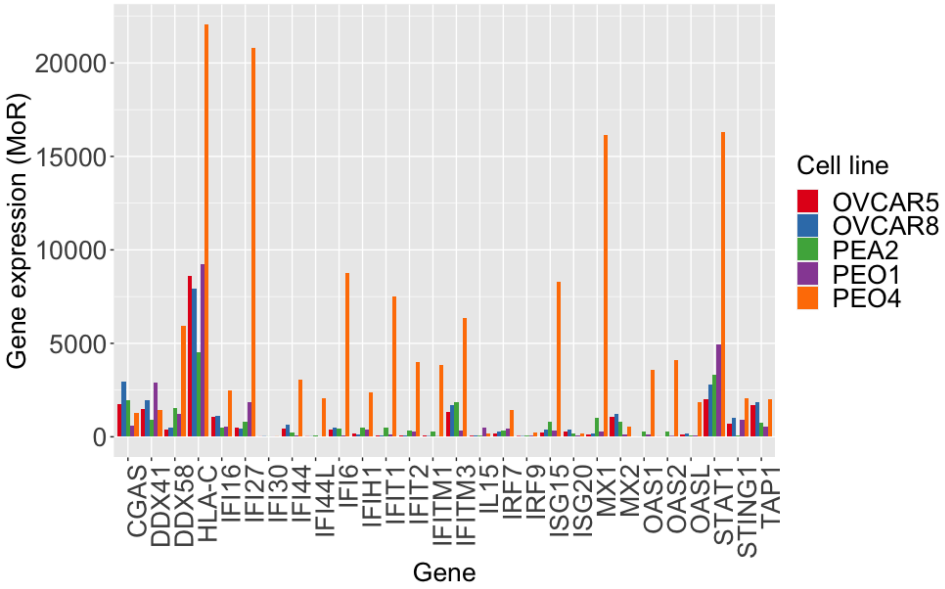


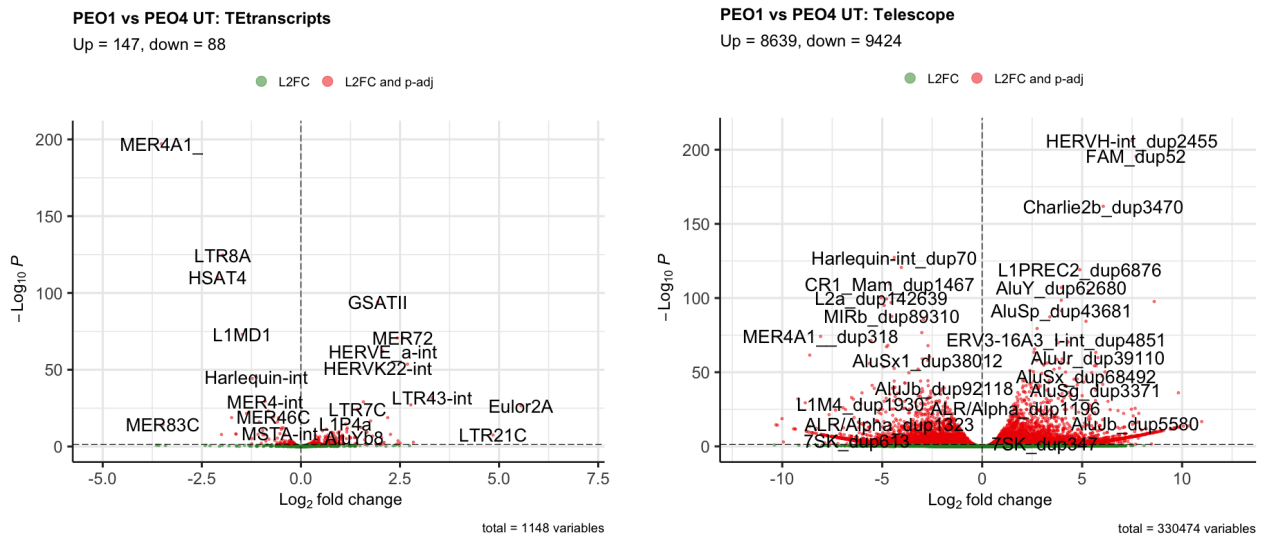
Figure 4.31 **Basal expression of VM ISGs across HGSC cell lines.** Basal expression of ISGs across the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 HGSC cell lines. ISGs defined in [439]. Gene expression calculated using the MoR normalisation method [380].

4.4.2 Heightened basal expression of HKMTi-1-005-induced TEs

HKMTi-1-005-induced immune stimulation is proposed to be due to a viral mimicry response. It is possible that heightened expression of TEs could contribute to the high baseline expression of VM ISGs. To investigate this hypothesis, I performed sub-family level and locus-specific level analyses of TE expression between untreated PEO1 and PEO4 cell lines.

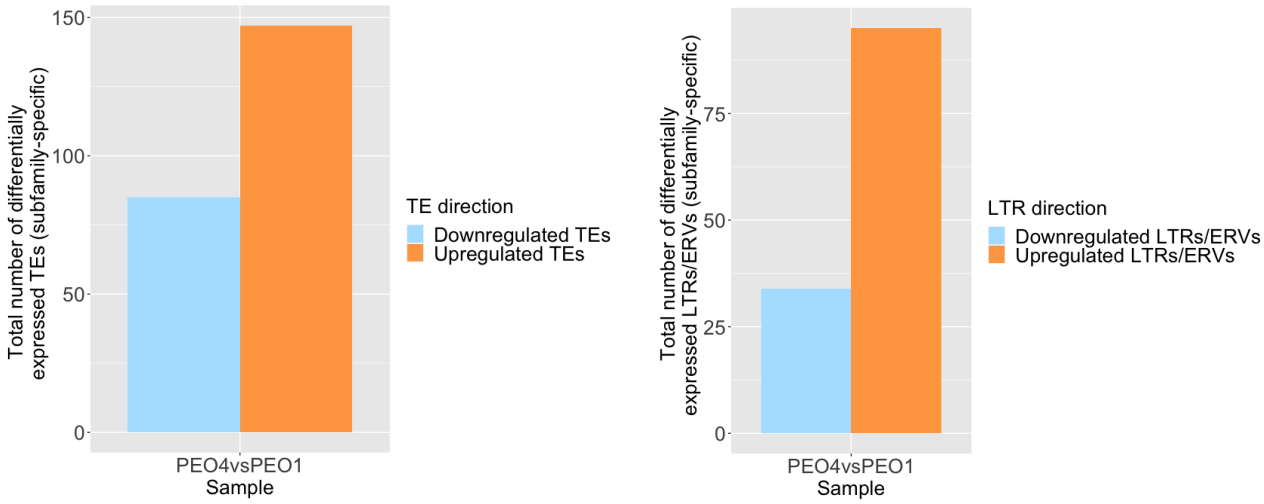
The transcriptional profile of TEs were vastly different between the PEO1 and PEO4 lines, at both the sub-family and locus-specific level (figure 4.32). Considering all TEs together, and LTRs/ERVs individually, PEO4 had a greater number of up- versus down-regulated elements at the sub-family level (figure 4.33). Locus-specific analyses revealed a greater total number of downregulated TE loci in PEO4; however, when LTRs/ERVs were considered separately, there were more upregulated loci in PEO4 (figure 4.34).

At the sub-family level, TEs upregulated in PEO4 were strongly skewed towards LTR/ERV superfamily TEs (figure 4.35). This contrasted with the downregulated TEs, where LTRs/ERVs made up a much smaller proportion. Within the LTRs/ERVs, upregulated elements were of the ERV1 family in almost 75% of cases, as compared to less than 50% of downregulated LTRs/ERVs (figure 4.36). These effects were less marked at the locus-specific level but were still evident (figures 4.37 and 4.38).



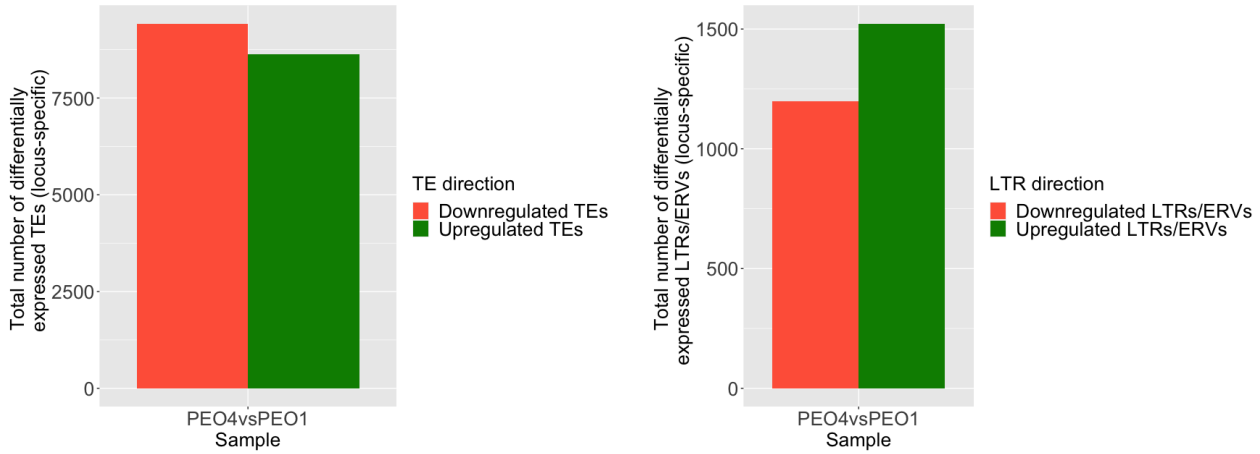
(a) (b)

Figure 4.32 **Comparison of PEO1 and PEO4 at baseline: TE expression.** Comparison of TE expression between PEO1 and PEO4 cell lines, at baseline (untreated). (a) TE expression quantified at the sub-family level, using the TEtranscripts tool [382, 383]. (b) TE expression quantified at the locus-specific level, using the Telescope tool [384]. TEs are labelled by colour, based on significance and fold-change: L2FC (green) = absolute fold-change >1 (absolute log2FC > 0), L2FC and p-adj (red) = both L2FC and p-adj < 0.05 criteria met. In each case, PEO1 is the "control" sample.



(a) (b)

Figure 4.33 **PEO1 versus PEO4: barplot of TE and LTR/ERV differences, sub-family level.** Bar plots showing total number of transcriptionally up- versus down-regulated (a) TEs or (b) LTRs/ERVs in PEO4 as compared to PEO1, untreated. TEs measured at the sub-family level, with a threshold of p-adj < 0.05 for differential expression. Blue = downregulated in PEO4, orange = upregulated in PEO4.



(a) (b)

Figure 4.34 **PEO1 versus PEO4: barplot of TE and LTR/ERV differences, locus-specific level.** Bar plots showing total number of transcriptionally up- versus down-regulated (a) TEs or (b) LTRs/ERVs in PEO4 as compared to PEO1, untreated. TEs measured at the locus-specific level, with a threshold of p-adj < 0.05 for differential expression. Red = downregulated in PEO4, green = upregulated in PEO4.

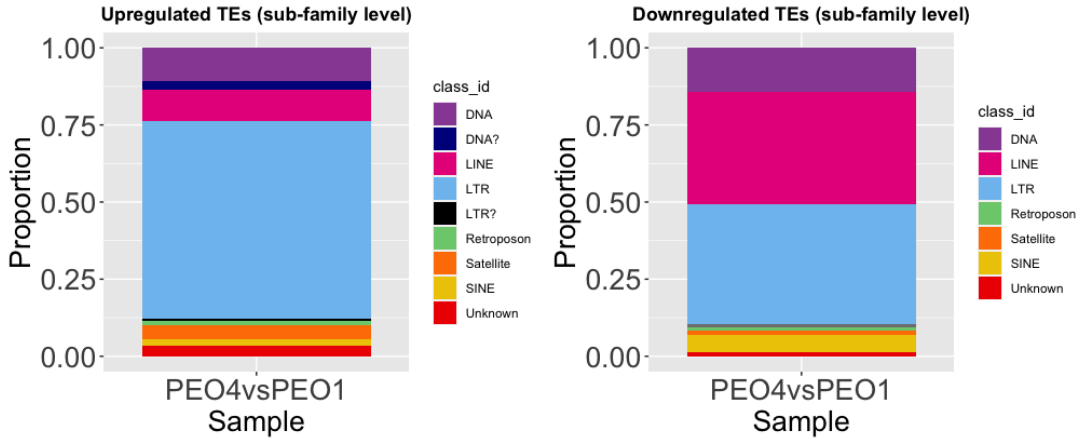


Figure 4.35 **PEO1 versus PEO4: proportion of TEs by superfamily, measured at sub-family level.** Bar plots demonstrating differences in TEs measured at the sub-family level in PEO4 line as compared to the PEO1 line, in control conditions. Significantly upregulated TEs (left) and downregulated TEs (right), coloured by TE superfamily ("class_id").

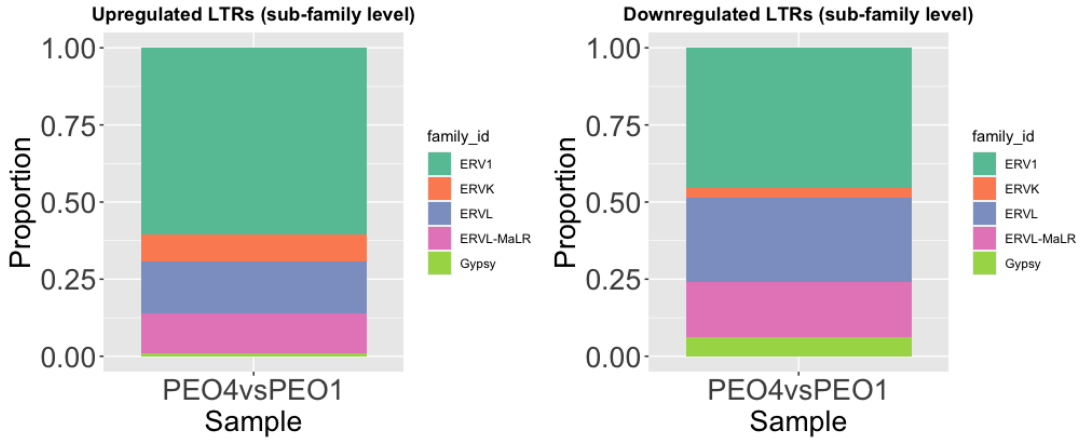


Figure 4.36 **PEO1 versus PEO4: proportion of LTRs/ERVs by family, measured at sub-family level.** Bar plots demonstrating differences in LTRs/ERVs measured at the sub-family level in PEO4 line as compared to the PEO1 line, in control conditions. Significantly upregulated LTRs/ERVs (left) and downregulated LTRs/ERVs (right), coloured by ERV family ("family_id").

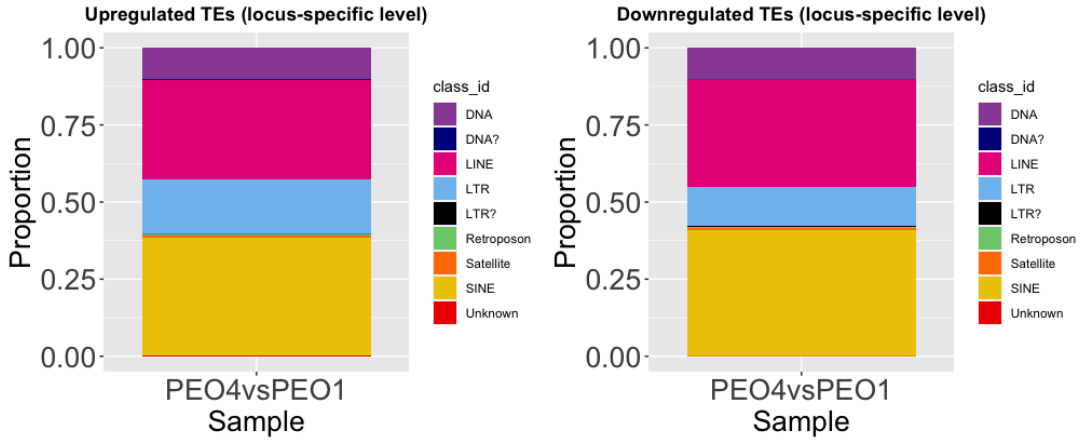


Figure 4.37 **PEO1 versus PEO4: proportion of TEs by superfamily, measured at locus-specific level.** Bar plots demonstrating transcriptional differences in TEs at the locus-specific level in the PEO4 line as compared to the PEO1 line, in control conditions. Significantly upregulated TEs (left) and downregulated TEs (right), coloured by TE superfamily ("class_id").

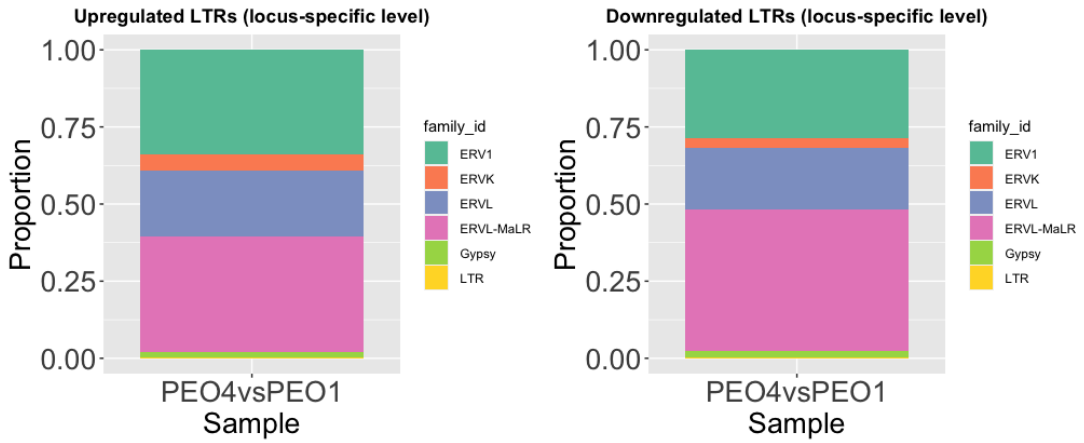
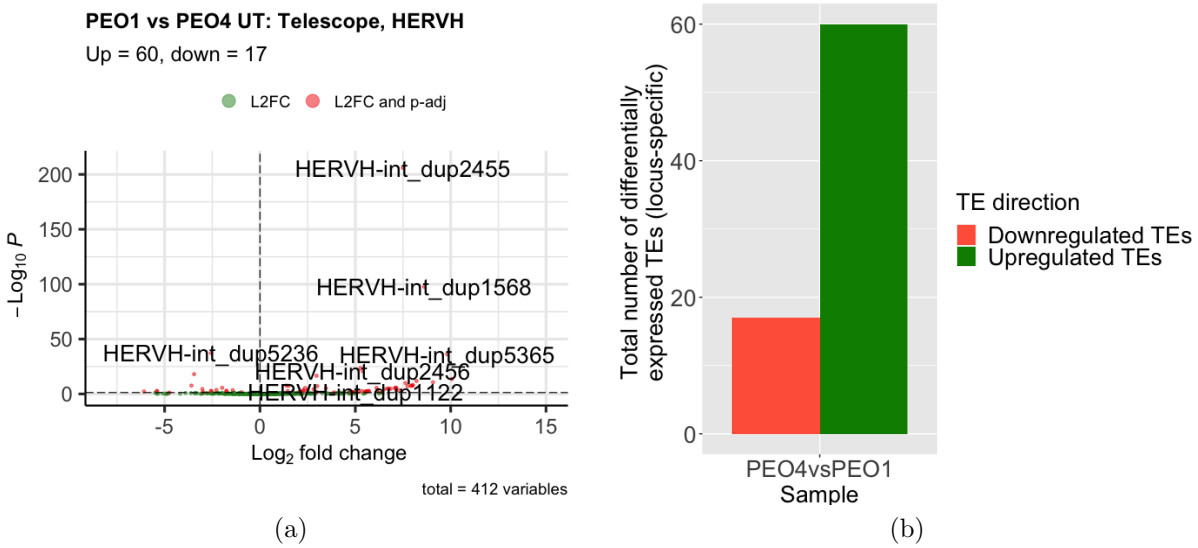


Figure 4.38 **PEO1 versus PEO4: proportion of LTRs/ERVs by family, measured at locus-specific level.** Bar plots demonstrating transcriptional differences in LTRs/ERVs at the locus-specific level in PEO4 line as compared to the PEO1 line, in control conditions. Significantly upregulated LTRs/ERVs (left) and downregulated LTRs/ERVs (right), coloured by ERV family ("family_id").

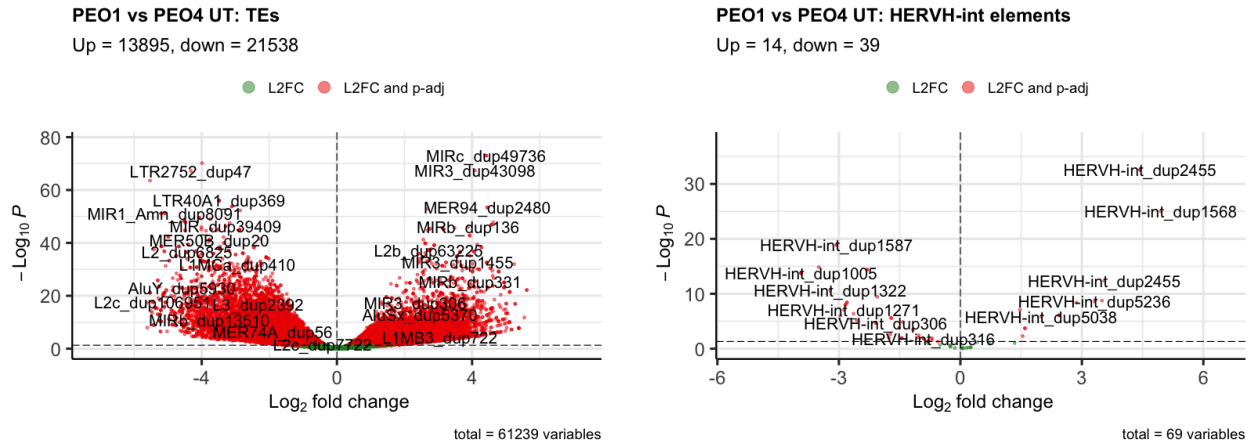
It was notable that many of the upregulated TEs appeared to be of the HERVH-int sub-family. This sub-family (of the ERV1 family) was the most consistently upregulated LTR/ERV sub-family following HKMTi-1-005 treatment of HGSC cell lines. Considering HERVH-int elements separately, there was a clear skew towards upregulation of these elements in the PEO4 line (figure 4.39). In combination, the data comparing PEO1 and PEO4 lines suggest that the high basal immune signalling in PEO4 – and lack of immune stimulation – may result from a high basal expression of TEs. In particular, PEO4 exhibits a high basal expression of the ERV1 family LTRs/ERVs and the HERVH-int sub-family elements that are the transcriptional targets of HKMTi-1-005 treatment, that may be responsible for induction of the VM response.

Accessibility profiles of TEs in untreated PEO1 and PEO4 cell lines were subsequently compared using the DAC approach. There was no clear skew towards increased accessibility of TEs in PEO4, including at HERVH-int sub-family elements (figure 4.40). Indeed, there were more TEs with decreased accessibility in PEO4 as compared to PEO1, a finding that held true when considering LTR/ERV elements only (figure 4.41). The TE superfamilies and ERV families that showed increased accessibility were very similar to the reference distribution, with only a slightly higher proportion of LTR/ERV superfamily and ERV1 family elements than expected based on the reference (figures 4.42 and 4.43, reference figures 4.12c and 4.14c).



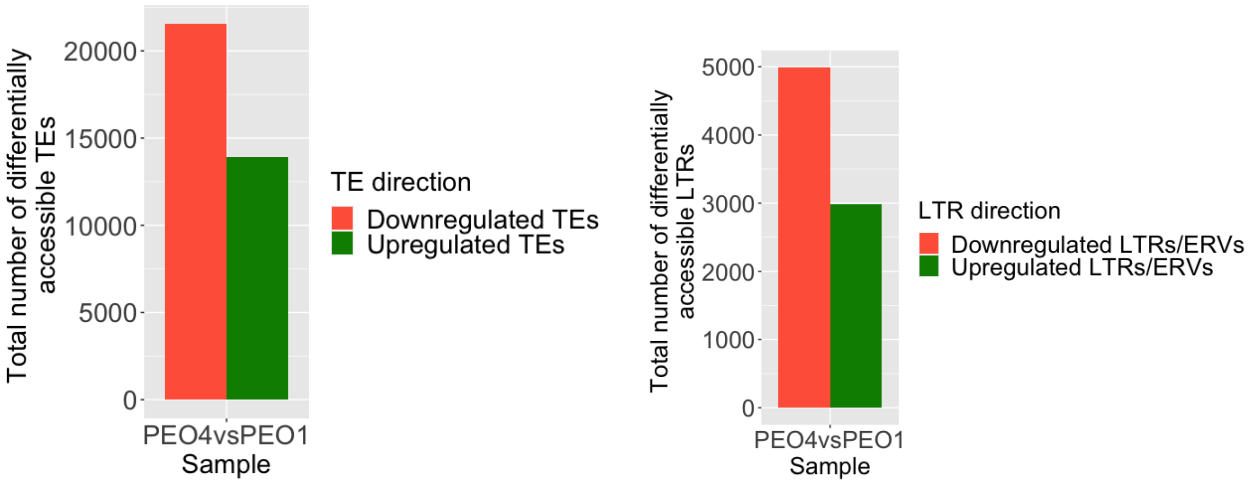
(a) (b)

Figure 4.39 **PEO1 versus PEO4: differences in HERVH expression at locus-specific level.** Comparison of HERVH-int sub-family element expression in PEO4 as compared to PEO1. (a) Volcano plot showing HERVH-int sub-family elements at the locus-specific level. Points are labelled by colour, based on significance and fold-change: L2FC (green) = absolute fold-change >1 (absolute log2FC > 0), L2FC and p-adj (red) = both L2FC and p-adj < 0.05 criteria met. (b) Total number of up- (green) versus down-regulated (red) HERVH-int sub-family elements in PEO4 as compared to PEO1, at the locus-specific level, with a threshold of p-adj < 0.05 for differential expression. In each case, PEO1 is the "control" sample.



(a) (b)

Figure 4.40 **PEO1 versus PEO4: differences in TE and HERVH accessibility.** Comparison of TE accessibility between untreated PEO1 and PEO4 cell lines, measured using DiffBind [390]. (a) Volcano plot showing accessibility differences across all TEs. (b) Volcano plot showing differences in HERVH-int element accessibility. DACs are labelled by colour, based on significance and fold-change: L2FC (green) = absolute fold-change >1 (absolute log2FC > 0), L2FC and p-adj (red) = both L2FC and p-adj < 0.05 criteria met. In each case, PEO1 is the "control" sample.



(a) (b)

Figure 4.41 **PEO1 versus PEO4: up- versus down-regulated DACs at TEs and LTR/ERVs.** Bar plots comparing differential accessibility at (a) all TEs and (b) LTRs/ERVs, in PEO4 as compared to PEO1. Upregulated TEs/ERVs in PEO4 are shown in green, whilst downregulated TEs/ERVs are in red.

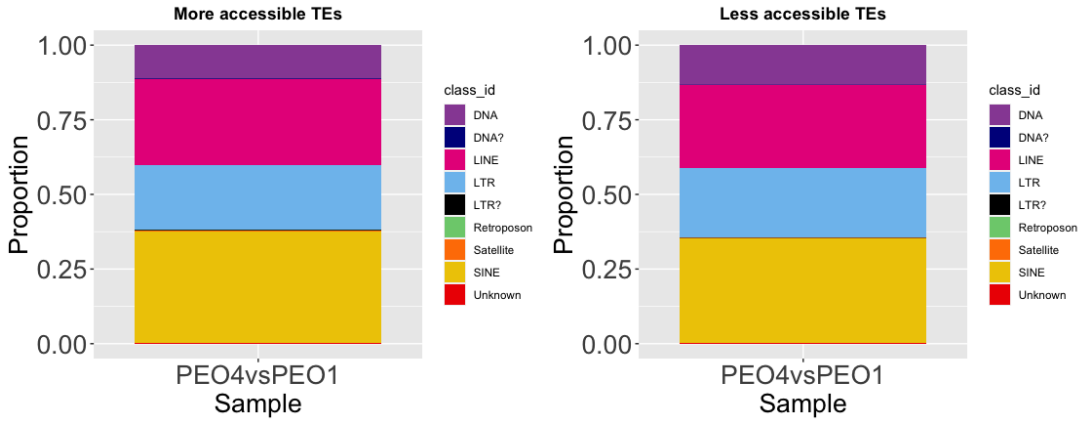


Figure 4.42 **PEO1 versus PEO4: differences in TE accessibility, proportion by superfamily.** Bar plots demonstrating accessibility differences in TEs measured in the PEO4 line as compared to the PEO1 line, in control conditions. Significantly more accessible (left) and less accessible (right) TEs, coloured by TE superfamily ("class_id").

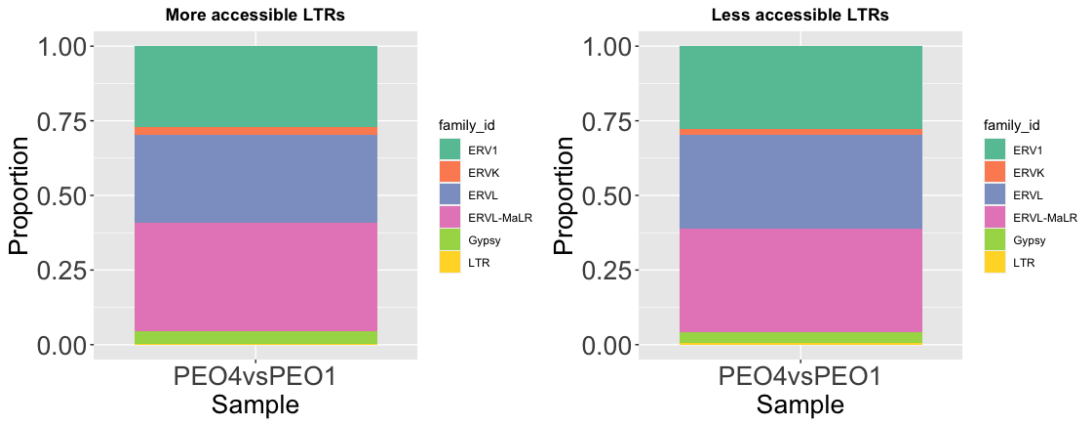


Figure 4.43 **PEO1 versus PEO4: differences in LTR/ERV accessibility, proportion by family.** Bar plots demonstrating accessibility differences in LTRs/ERVs measured in the PEO4 line as compared to the PEO1 line, in control conditions. Significantly more accessible (left) and less accessible (right) ERVs, coloured by ERV family ("family_id").

4.5 Discussion: conclusions, limitations and future work

The findings from this chapter demonstrate induction of a VM response following HKMTi-1-005 treatment of HGSC models and characterise the nature of this immune response. The data support de-repression of TEs, particularly LTRs/ERVs. Further, they support the hypothesis that the type of TEs induced may reflect the histone methylation targeting effect of HKMTi-1-005.

Analysis of the transcriptional changes in response to HKMTi-1-005 treatment illustrated up-regulation of immune pathways in all samples except the PEO4 cell line (see 4.2.1). Consistent effects were observed on the IFNG and innate immune response pathways, with variable results on IFNA/IFNB response pathways. A subset of the IFNG/innate immune response genes induced comprised ISGs, considered to be crucial components of the VM response [251]. To identify a reduced signature of ISGs describing the HKMTi-1-005-induced VM, strongly and consistently upregulated genes from the ISG set defined in [439] were chosen. This resulted in selection of 7 ISGs – referred to as the 7ISG – that characterise the immune signalling response induced by the drug in HGSC lines (figure 4.6). Upregulation of the 7ISG was subsequently confirmed in the ID8 *in vivo* model (figure 4.7). Importantly, evaluation of VM effects described in the literature demonstrated that there was no consistent induction of these 7ISGs by single HDACi or single DNMTi treatment of four OC cell lines [439], suggesting specificity of these genes in the HKMTi-1-005-induced VM response (figures 4.44 and 4.45).

There were several notable features of the HKMTi-1-005-induced immune response in the HGSC models. The first was the amplified response observed in the ID8 *in vivo* sample as compared to the cell lines (figures 4.1 and 4.2). The difference in HKMTi-1-005 treatment effects between HGSC cell lines and the ID8 sample are likely to be partially explained by differences in the models and treatment regimes. However, it is also likely that cancer cell-intrinsic immune changes induced by the compound are amplified *in vivo* due to effects on non-tumour cells; cancer cell-intrinsic events are known to heavily influence immune cells in the TME and subsequently patient outcome, via bi-directional immunomodulation [440]. Additionally, there may be direct effects of HKMTi-1-005 on non-tumour cells that would be captured in this bulk sequencing; inflammatory genes are demonstrated to be targets of H3K9 methylation in immune cells [142], hence treatment could induce immune signalling directly in these cells whilst also amplifying paracrine effects. Therefore, both direct and indirect effects of treatment on cells in the TME likely contribute to the heightened immune response in the ID8 *in vivo* sample.

A surprising feature of the immune response induced by the compound was the induction of innate immune and IFNG responses, but not IFNA/IFNB responses (figure 4.1). The IFNG cytokine is produced by cells in the TME, primarily T and NK immune cells [441]. No non-tumour cells can

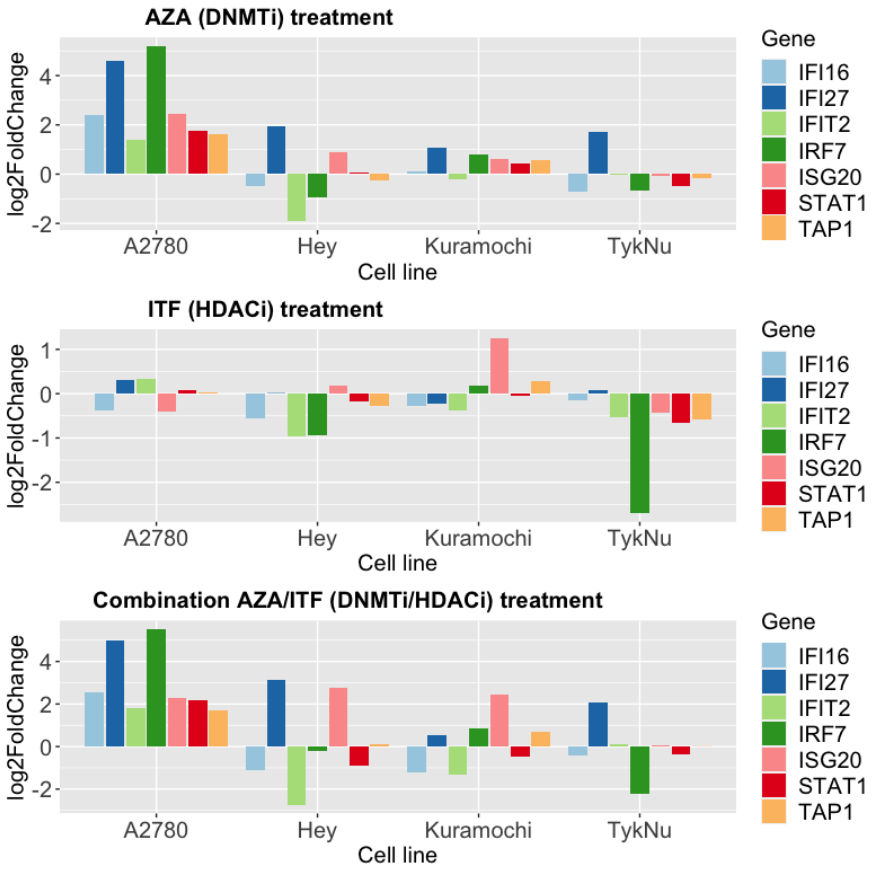


Figure 4.44 **Comparison of 7ISG changes in response to DNMTi, HDACi or combination in 4 ovarian cancer cell lines.** Change in 7ISGs following AZA (DNMTi, top), ITF (HDACi, middle) or combination (bottom) treatment in ovarian cancer cell lines A2780, Hey, Kuramochi or TykNu. Reproduced based on data from [439] (subset of data plotted in supplementary figure S1 of [439]). Data are available in GEO, accession GSE182430. Data represent single replicates for each cell line, hence no p-adjusted values are plotted (all p-adj > 0.9).

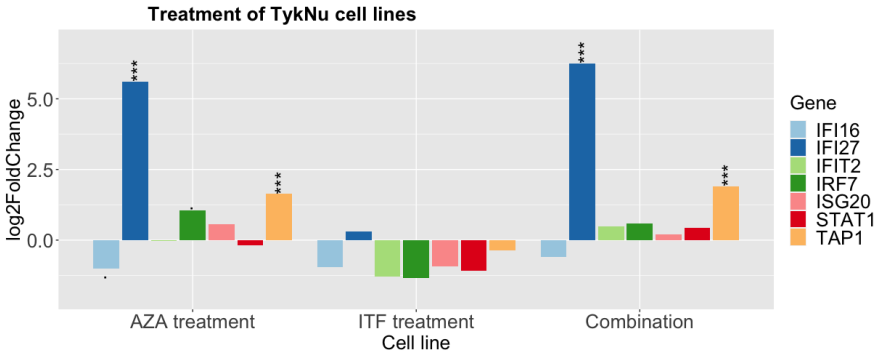


Figure 4.45 **Comparison of 7ISG changes in response to DNMTi, HDACi or combination in the TykNu ovarian cancer cell line.** Change in 7ISGs following AZA (DNMTi, left), ITF (HDACi, middle) or combination (right) treatment in the TykNu ovarian cancer cell line. Reproduced based on data from [439]. Data are available in GEO, accession GSE182430. Data represent triplicate RNAseq runs. . = p-adj < 0.1, * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001.

be accounted for in the HGSC cell line panel RNAseq assay. In line with this fact, IFNG itself was not expressed at a measurable level in any of the cell lines. These data suggest that either IFNG response genes are directly de-repressed by the compound, or the strong overlap in immune signalling pathways results in enrichment of the IFNG response pathway due to activation of an alternative immune pathway with convergence on IFNG effector genes. Unlike IFNG, IFNA/IFNB cytokines are expressed by tumour cells [441]. However, across the HGSC cell lines, IFNA/IFNB cytokines were not detectable. This was reflected in the lack of consistent induction of IFNA/IFNB response pathways using fGSEA. Whilst it is possible that the compound does not induce these signalling pathways, data from the ID8 *in vivo* sample and from ID8 F3 cell line analyses performed in our lab have demonstrated induction of IFNB (Nayana Iyer, personal communication). Further, the lab has demonstrated that IFNB responses spike at 24h and have fallen by 72h, whilst IFNG responses spike at 72h (Nayana Iyer, personal communication).

These findings could explain the immune signalling responses observed in HGSC cell line models without IFNA/IFNB responses: the HGSC cell line RNAseq was performed at 72h, by which point the dynamic and volatile type I IFNA/IFNB responses may have declined, whilst the secondary wave of downstream immune signalling typically associated with IFNG may still be detectable. Additional RNAseq or qRT-PCR analyses performed on HGSC cell lines at 24h after HKMTi-1-005 treatment, with a focus on IFNs, could help to confirm these observations. In parallel, western blot analyses of phosphorylation of key signalling proteins such as JAKs/STATs/IRFs in these cell lines could be beneficial. Many IFN signalling proteins are activated by phosphorylation rather than being dependent on upregulated gene expression, and data from our lab support consistent phosphorylation of key IFN signalling proteins such as TANK-Binding Kinase 1 (TBK1), STAT1 and Interferon Regulatory Factor 3 (IRF3) following treatment in ID8 F3 cell lines (Nayana Iyer, personal communication).

Having established transcriptional induction of immune pathways – particularly innate immune and IFNG response signalling, including ISGs – TE induction was assessed. Transcription of TEs can be assessed at various levels: TEs can be grouped by sub-family and expression summarised at all loci within this group, or can instead be measured at a locus-specific level. Sub-family level analysis of TE transcription in HGSC cell lines indicated a consistent up- rather than down-regulation of TEs (see 4.3.1). Additionally, there was a marked enrichment for the ERV superfamily in upregulated TEs, an effect that was not evident in downregulated TEs. An unexpected result from this sub-family analysis was the limited upregulation of ERVs in the ID8 *in vivo* sample, despite its strong immune signalling induction. Indeed, a larger number of ERV sub-families were down- instead of upregulated in this sample. However, subsequent analyses of TEs at a locus-specific level provided a potential explanation for this effect.

Locus-specific measurement of TEs demonstrated many differentially expressed loci in all samples (see 4.3.2). A large proportion of these changes were in ERV superfamily elements, and in every sample, there was a greater up- rather than down-regulation of ERVs following HKMTi-1-005 treatment. In contrast to the sub-family level analyses, this was also true of the ID8 *in vivo* sample. In fact, upregulated TEs in the ID8 *in vivo* sample were an order of magnitude larger than those measured in most HGSC cell lines. These findings suggest that although treatment with HKMTi-1-005 induces a vast number of TEs in the ID8 *in vivo* sample, few changes are concordant across sub-families. Potentially, this effect may once again be explained by the bulk nature of sequencing in the ID8 *in vivo* sample. Changes in cellularity in control versus HKMTi-1-005-treated samples were demonstrated by flow cytometry in the initial analysis of HKMTi-1-005 treatment of ID8 *in vivo* samples, including increased effector T and NK cell infiltration [155]. These cellularity changes are likely to have a strong effect on which TEs are expressed, as many TEs display cell type-specific patterns of expression at baseline [442]. Further complexity arises from the potential for HKMTi-1-005-induced TE expression to influence immune infiltration, thereby affecting cellularity. Additionally, HKMTi-1-005 treatment of different cell types may induce different TE loci as TE de-repression has been shown to be cell type-specific [153], which may obscure measurements at the sub-family level.

Another notable difference in sub-family versus locus-specific analysis of TEs was the extent of ERV superfamily overrepresentation in upregulated TEs. This was much more evident at the sub-family level, whilst in locus-specific analyses there was only a slight skew towards ERVs over other TE superfamilies, with many SINEs and LINEs also demonstrating transcriptional upregulation. These data indicate that, despite extensive changes in TEs across multiple superfamilies, there was more concordance in the ERV transcriptional changes across sub-families than within other TE superfamilies following HKMTi-1-005 treatment, supporting a consistent drug effect at loci within the ERV superfamily.

An important aspect of the TE investigations following HKMTi-1-005 treatment was determining whether there was support for a distinct TE effect when targeting histone methyltransferases, rather than DNMTs or alternative histone modifiers. Ohtani et al. first proposed this distinction [153], with a focus on ERVs. They hypothesised that older ERVs (majority ERV1/ERVL, or class I/III) are silenced by histone methylation, in contrast to younger ERVs (majority ERVK, or class II), which depend more heavily on DNA methylation (figure 4.46 and table 4.3). The group demonstrated that H3K9me2/3 marks at LTRs correlate with evolutionary age, and that the use of a DNMTi induces distinct LTRs as compared to knockdown of histone methyltransferases: LTRs induced by histone methyltransferase (HMT) knockdown had greater divergence and lower CpG density than those induced by DNMTi, indicating an older evolutionary age [153]. Importantly, Ohtani et al. suggested that histone methylation targeting may not induce significant VM in comparison with

DNMTi (table 4.3); they proposed that knockdown of certain histone methyltransferases (including G9a) induces older/intermediate age LTRs but that these LTRs not capable of inducing immune genes, and additionally that H3K27me3 does not directly silence LTRs and hence EZH2 knockdown does not stimulate VM. Finally, the group suggest that the majority of the oldest ERVs, such as the ERVL elements, are no longer transcriptionally competent and cannot be induced by DNMTi or knockdown of histone methyltransferases. To aid in assessment of HKMTi-1-005-induced VM versus alternative epigenetic treatments, the most direct comparison is the treatment of four ovarian cancer lines by McDonald et al. [439] with DNMTi, HDACi or the combination, followed by TE analysis at the sub-family and locus-specific levels using the same tools employed here.

HKMTi-1-005 treatment effects supported several of the hypotheses described. Both TEtranscripts and Telescope results demonstrated a strong skew towards the ERV1 family within the upregulated ERVs (figures 4.14b and 4.20c). ATACseq data supported RNAseq findings, showing a skew towards older/intermediate age ERV1 accessibility following HKMTi-1-005 treatment (see 4.3.3). This ERV1 (class I) family is an older/intermediate age family, in contrast to the evolutionarily young ERVK family and the oldest ERVL/ERVL-MaLR family (see figure 4.46) [153]. Moreover, the most consistently induced ERV sub-family following HKMTi-1-005 treatment was HERVH-int, upregulated in all cell lines. HERVH-int elements demonstrate a range of evolutionary ages [165], however are highly degenerated with a high average LTR divergence [165] and are described to have integrated at least 40 mya [167]. Notably, certain HERVH elements are evidenced to act as cancer epitopes with potential to be recognised by cytotoxic T cells [224]. In contrast, DNMTi/HDACi combination treatment in ovarian cancer lines consistently induced ERV9 sub-family elements [439] with an evolutionary age of 15 mya [167]. No ERV sub-families were induced in all four ovarian lines by single DNMTi treatment [439], but two sub-families were induced by DNMTi in at least 3 lines: HERVK11D-int and LTR10B1. The former is of the evolutionarily young ERVK family. The latter is of the older ERV1 family, however some ERV1 elements are younger than ERVK elements [153] and literature suggests this LTR10B1 sub-family is younger than most ERV1 family elements [443].

Although the oldest elements were hypothesised to be transcriptionally incompetent, the data did illustrate upregulation of several ERVL family elements, albeit not to the same extent as ERV1. Indeed, some of the most consistently induced ERVs at the locus-specific level were those of the MLT sub-family from the ERVL-MaLR family; these are exceptionally old, over 100 mya [167]. MLT elements are not described as induced by the DNMTi/HDACi treatment individually or in combination [439]. Taken together, the data support the hypothesis that older/intermediate age TEs are more heavily controlled by histone methylation than young ERVK elements, whilst DNA methylation regulates younger ERV silencing. Additionally, they support weaker transcriptional competence of the oldest ERVL elements yet demonstrate that a subset of these is inducible by

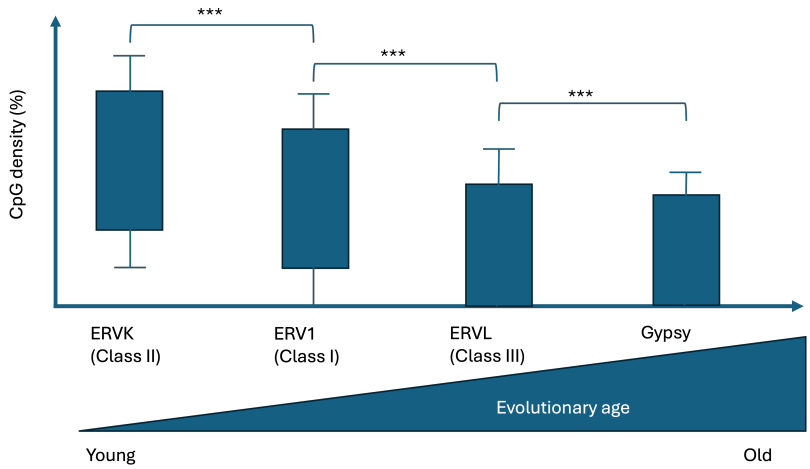


Figure 4.46 **ERV classes versus CpG density, from [153].** Reproduced based on [153], schematic illustrating differences in CpG density associated with LTRs from ERV families of different ages. *** = $p < 0.001$ by Mann-Whitney U test.

Treatments	**Activation**			
	Evolutionary age of LTRs			Viral defence genes
	Young	Intermediate	Old	
5-Aza-Cdr	+	-	-	+
HMT knockdowns	-	+	-	-
Dual depletion	++	++	-	++

Table 4.3 **LTR induction by epigenetic therapy, from [153].** Table reproduced from [153], demonstrating the differences in LTR ages induced by DNMTi, HMT knockdown or the combination. The extent of induction of viral defense genes also shown (defined as type I IFN response genes or defense response to virus). Cell lines used are HCT116 (colorectal carcinoma), HL-60 (acute myeloid leukaemia), MCF-7 (breast carcinoma), and HepG2 (hepatocellular carcinoma). 50 LTR families with highest CpG density were defined as young (CpG density > 1.34%), 50 with lowest CpG density were defined as old (CpG density < 0.54%). The remainders were defined as intermediate age.

histone methylation targeting.

Moreover, our data demonstrate consistent induction of ISGs in contrast to suggestions that histone targeting and/or older ERVs are not capable of doing so. The data presented by McDonald et al. [439] demonstrate that ISG induction by HKMTi-1-005 was very similar to that observed by combination DNMTi and HDACi treatment, and appeared even to be greater than ISG induction by either inhibitor individually (figure 4.44 and 4.45 and appendix figures A.17 and A.18). Given the differences in ERV ages induced by HKMTi-1-005 versus this combination, the data suggest that older/intermediate age TEs may indeed be capable of inducing a VM response to a similar extent to younger TEs. Whether the oldest ERVL elements as well as the intermediate age ERV1 elements contribute to the VM ISG upregulation observed cannot be determined from the present data, but represents an interesting mechanistic question for future investigations.

It is important to note that the knockdown efficiency of HMTs by Ohtani et al. was a concern acknowledged by the group (50-80% efficiency), with suggestion that only a subset of transcriptionally competent ERVs may have been identified [153]. Additionally, only LTRs rather than full-length

ERVs were considered and only those in intergenic regions, hence a substantial proportion of HMT-regulated ERVs may have been excluded. This is especially true given that the older, HMT-regulated ERVs are more likely to have been co-opted to become regulatory elements that may fall within genic regions [185]. Further, the ISGs examined by Ohtani et al. were from the type I IFN and defence response pathways, without inclusion of the IFNG or general innate immune response signalling pathways that were most consistently induced by HKMTi-1-005 treatment. Possibly, induction of ISGs from these latter pathways was overlooked in the analysis of HMT knockdown effects, leading to an underestimation of VM induction in this context.

A final comparative observation between the HKMTi-1-005-induced VM and the DNMTi/HDACi-induced VM described by McDonald et al. is the transcriptional and accessibility effects at SINE loci. McDonald et al. noted that many TEs upregulated by DNMTi/HDACi treatment were SINEs [439], although they were not robustly upregulated by DNMTi treatment alone. SINEs are known to be capable of stimulating a VM immune response, via formation of inverted repeats that are recognised by cytoplasmic RNA sensors [234]. Hence, it is possible that these elements contribute to the HKMTi-1-005-induced immune response in a similar manner to DNMTi/HDACi combination, especially given the known role of H3K9 methylation in silencing SINEs [28]. HKMTi-1-005 induced a large increase in SINE accessibility, with strong transcriptional upregulation at the locus-specific level, despite almost no concordance in transcriptional changes at the sub-family level. McDonald et al. similarly noted that increased chromatin accessibility at SINEs was not strongly correlated with transcriptional upregulation at the sub-family level [439]. Hence, the data support older/intermediate age ERVs as relevant TE targets of HKMTi-1-005 treatment but suggest that other TE superfamilies such as SINEs may contribute to HKMTi-1-005-induced VM, elements that are not induced by DNMTi alone. Additionally, they demonstrate that HMT targeting and DNMTi have distinct – and possibly complementary – VM effects.

Although comparison to the literature is of value in identifying general similarities and differences between VM induced by DNMTi/HDACi treatment versus histone methylation targeting, it is important to note that the cell line-specificity of VM effects prevent a conclusive comparison. Additionally, no comparison of DNMTi/HDACi VM effects is available in a mouse model. In contrast to the ERV1 family that dominated upregulated TEs in the HGSC cell lines, the ERVK family were markedly overrepresented in the ID8 *in vivo* mouse sample. The literature is much less comprehensive regarding the ages of mouse TEs than those in humans. However, the genomic distribution and abundance of ERVs are highly dissimilar between the mouse and human genomes [162, 163]. Importantly, ERV1 and ERVL families make up 2.92% and 1.55% of the human genome respectively, but only 0.68% and 0.58% of the mouse genome. In contrast, ERVK constitute 0.3% of the human genome but 3.14% of the mouse [163]. Additionally, there is conflicting evidence regarding the ERV targets of histone

methylation silencing in the mouse genome; whilst *G9a/GLP* knockout was shown to induce class II ERVK but not class III ERVL elements by one group [198], ERVL elements were demonstrated to be the primary target of the same knockout by another [199]. Hence, several explanations exist for ERVK overrepresentation in the ID8 sample. Potentially, the epigenetic silencing mechanisms differ between mouse and human genomes, and the evolutionary ages of subsets of ERVK elements are more similar to the older elements in humans. It is also possible that the bulk sequencing skews measurement of TE expression, for the reasons described previously.

Of note, whilst the transcriptional analyses of TEs were performed using specialised tools, equivalent tools to analyse chromatin accessibility changes at these repetitive loci are not yet available. A tool created by the same group as the TEtranscripts tool (named "TE peaks") is under development for use with ATACseq data, which could prove useful for future experiments (description at `https://github.com/mhammell-laboratory/TEpeaks`). The lack of such tools at present may explain the limited overlap between accessibility and transcriptional changes at TE sites. For example, when considering DiffBind accessibility analyses the choice of peak width did not materially affect the results, however the choice of background peaks used for normalisation considerably affected the output (see 2.2.4, 2.3.4, 2.4.3). The existence of DACs in both directions at single TE loci further complicates accessibility analyses. Additionally, some changes in TE accessibility are likely to reflect changes in TEs acting as regulatory elements without accompanying transcription. ERV1 elements in particular are overrepresented as enhancer elements [185], and certain ERV elements are demonstrated to act as enhancers for IFNG response signalling [184]. Despite these variables, there was significant overlap in the PEO1 and PEO4 TE accessibility and transcription results that were not confounded by cellularity changes, by both ATACseq methods. The use of TEs as tissue-specific enhancers and promoters [442] render accessibility versus expression analyses in the bulk ID8 *in vivo* sample particularly difficult to compare, however the ATACseq subtraction approach still demonstrated significant overlap between TEs with increased accessibility and those with increased expression.

Whilst there was strong evidence for a VM immune response induced by HKMTi-1-005 across most HGSC models, the PEO4 cell line demonstrated a strikingly divergent response (figures 4.1 and 4.2). TEs were still upregulated in this cell line, however the IFNG and innate immune response genesets were transcriptionally downregulated following treatment. This observation was particularly interesting in the PEO4 line, derived from the same the patient as PEO1 but after the development of platinum resistance; literature suggests that resistance to chemotherapy can result in a vulnerability to VM induced by EZH2i [195], in opposition to PEO4 observations. Given literature relating baseline ERV/ISG expression to the extent of the VM response [214, 220], these features were compared between the PEO4 and PEO1 cell lines.

The PEO4 cell line had a markedly higher expression of immune genes at baseline than PEO1 (see 4.4). This included immune genes that were relevant to the HKMTi-1-005-induced VM response, such as the 7ISGs that were most consistently induced by treatment. Considering PEO4 as compared to all other HGSC lines tested, the same effect was observed: ISGs were highly upregulated in untreated PEO4 cells compared to the other lines.

PEO4 also demonstrated a high basal TE expression compared to PEO1. ERVs - especially the ERV1 sub-family - were overrepresented in the TEs that were upregulated in PEO4 at baseline. Given that the HERVH-int sub-family elements were the most consistently induced by HKMTi-1-005, transcriptional changes in these elements were assessed; many HERVH-int loci showed high basal expression in untreated PEO4 cell lines. This did not appear to be highly correlated with increased accessibility of these elements, suggesting alternative mechanisms may drive their expression.

It was notable that despite the lack of immune induction in PEO4, TEs were upregulated by HKMTi-1-005 treatment to a similar extent to other cell lines, including upregulation of the HERVH-int element. Chronic immune signalling can lead to IFN insensitivity, for example via inhibition of IFN signalling components such as STATs and IRFs [441]. Taken together, these data suggest that a high basal TE expression – especially, high basal ERV expression including HERVH-int – could drive a chronically immune-stimulated state in the PEO4 cell line to prevent further immune induction by HKMTi-1-005, despite transcriptional upregulation of TEs. The data lend support to the hypothesis that if high basal ERV expression is accompanied by high basal ISG expression, this may relate to insensitivity to VM induction [214] rather than a "primed" state [220]. However, verification in a broader panel of cell lines is required to increase confidence in this association.

Finally, it must be acknowledged that induction of VM genes cannot be causally linked to the recognition of TEs by receptors. Both TE expression and immune stimulation are observed, but there is complexity in the potential connection between these two axes. For example, TEs can act as regulatory elements for IFNG signalling [184]. This may explain the upregulation of the IFNG response pathway in HGSC cell lines following HKMTi-1-005 treatment: although the IFNG cytokine itself may not be induced, IFNG response genes could be transcriptionally upregulated due to activation of associated TE enhancer elements. Additionally, whilst canonical VM is due to TE recognition with downstream IFN induction, the inverse can occur: IFN signalling can induce TE expression [184]. Other pathways such as DNA damage are also known to stimulate immune responses [238] and EZH2/G9a are directly involved in repression of certain immune genes [127, 128, 249]. Hence, there is the possibility that immune induction is achieved by direct de-repression of immune genes or downstream of alternative events, either instead of or in addition to TE detection. One potential mechanism for a direct induction of immune genes by HKMTi-1-005 is discussed in chapter 5.

Chapter 5

Results Section 3: Investigating the relationship between *MYC*-mediated immune repression and HKMTi-1-005

5.1 Introduction

Chapters 3 and 4 demonstrated a strong induction of many immune pathways following HKMTi-1-005 treatment of HGSC models. In addition to the TE expression with viral mimicry described in chapter 4, it is possible that the immune stimulation observed is achieved via direct de-repression of immune genes. Both EZH2 and G9a have been implicated in the repression of immune genes, in a cell-intrinsic manner, hence HKMTi-1-005 treatment may act directly at these immune targets [127, 128, 249, 250].

Additionally, EZH2 and G9a may achieve further repression in complex with other proteins. Recent literature has described considerable interplay between EZH2/G9a and cMYC (see 1.8.3 and 1.8.4), a protein that is increasingly understood to have a non-canonical role in repression of immune genes [269, 272–274, 277, 278]. Particularly, the literature demonstrates a role for cMYC in the repression of many innate immune signalling components, including JAK/STAT, IRF, STING and type I/II IFN signalling [269, 272–275, 277, 278]. Additional immune repression effects by cMYC are achieved via inhibition of antigen presentation machinery, immune checkpoint ligands and production of inflammatory cytokines [270, 272, 276, 278, 280].

Preclinical data suggest that these effects may be antagonised by inhibition of EZH2 or G9a, with evidence that targeting of either enzyme can reduce cMYC activity [137, 139, 267, 283, 287–289]. Both enzymes are capable of direct and indirect regulation of cMYC at the protein and mRNA level and each exists in a complex regulatory feedback loop with cMYC (see 1.8.3 and 1.8.4).

Interdependence is further demonstrated by the significant co-occurrence of *MYC* amplification with *G9a* alteration that has been demonstrated in ovarian cancer amongst other cancer types [268]. Indeed, alteration of *G9a* and *EZH2* both significantly co-occur with *MYC* alteration across the TCGA cohorts (see appendix table A.2).

Deregulation of *MYC* is frequently found in cancer [258]. Literature demonstrates that cMYC hyperactivation or perturbation can reduce the threshold for IFN signalling and VM [235, 311], implying that *MYC*-deregulated cancers may be particularly sensitive to VM induction. This is potentially due to increased accumulation of dsRNA in these cells – downstream of cMYC effects on transcription, splicing and DNA damage [235, 258, 279, 312, 444] – and a reduced threshold for cell death [279]. Especially, *MYC*-deregulated tumours may be sensitive to VM induced by dual targeting of EZH2 and G9a, due to the extensive cMYC-G9a/EZH2 interplay described with potential for direct relief of the immune targets of cMYC repression. Aside from VM effects, EZH2 and G9a targeting may disrupt canonical transactivating cMYC activity owing to the role of these enzymes in regulating *MYC* mRNA and cMYC protein.

HGSC tumours are frequently *MYC*-amplified [313]. Supporting the literature, treatment of HGSC models with HKMTi-1-005 induced many of the immune targets of cMYC repression (see chapters 3 and 4). Of the hallmark pathways included in the MSigDB, the IFNG and inflammatory response pathways that were upregulated by HKMTi-1-005 have been demonstrated as key targets of cMYC repression that are consistently upregulated upon knockdown of *MYC* [272]. Additionally, HKMTi-1-005 treatment induced downregulation of hallmark cMYC and the E2F target pathways, once again mimicking cMYC knockdown effects (see chapter 3 and appendix figure A.4) [272]. Furthermore, several of the most sensitive tumour types to HKMTi-1-005 in the GDSC2 panel were those comprising many *MYC*-amplified cell lines, such as SCLC cell lines (see figure 3.40). Hence, I hypothesised that VM immune stimulation by HKMTi-1-005 may be amplified by direct de-repression of immune genes, in part via disruption of cMYC-mediated repression and with heightened effect in *MYC*-deregulated cancers.

This hypothesis proposed several lines of investigation. First, I assessed the relationship between cMYC-G9a/EZH2 repression and immune signalling; whilst cMYC repression of immune genes is well characterised, as is cMYC-G9a/EZH2 interaction, literature demonstrating the convergence of these two axes is in its infancy [250, 285]. Only one other dual inhibitor has been described at the time of writing, and this compound has been suggested to downregulate the cMYC signalling axis in non-small-cell lung cancer [445]. Of note, enzymatic targeting of EZH2 has thus far been insufficient for targeting components of cMYC-EZH2 interaction including immune inhibition [250, 292, 300, 301]. Hence, I next assessed the effect of HKMTi-1-005 treatment on cMYC activity in HGSC models. In particular, I examined whether transcriptional effects of treatment are distinct

between cMYC activation and repression targets; although induction of cMYC-repressed genes is desirable, induction of the largely pro-oncogenic cMYC-activated genes is not.

The heightened sensitivity of *MYC*-deregulated cancers to VM may provide insight into markers of sensitivity to HKMTi-1-005, proposing another line of investigation, namely, whether *MYC* deregulation relates to HKMTi-1-005 sensitivity. If inflammatory cell death induced by the compound is partly due to de-repression of cMYC-G9a/EZH2-repressed immune genes, it is also possible that basal expression of immune genes may predict sensitivity to HKMTi-1-005. This effect may be greater in *MYC*-deregulated tumours where there is a lower threshold for VM, and possibly a greater reliance on cMYC-G9a/EZH2 repression. Hence, subsequent analyses assessed the relationship between HKMTi-1-005 sensitivity, *MYC* deregulation and basal immune signalling. Given that the 7ISG constitutes a signature geneset describing HKMTi-1-005-induced immune stimulation, these investigations included measurement of basal 7ISG scores amongst other immune genesets.

At this stage in the development of the compound, it is not possible to assess the sensitivity of patients to HKMTi-1-005 based on *MYC* or immune statuses, nor efficacy of the drug in patients. However, publicly available survival and transcriptomic data were interrogated to assess preliminary evidence for anti-tumour HKMTi-1-005 efficacy *in vivo* and for prediction of potentially sensitive patient groups. Particularly, the relationship between 7ISG scores and survival were examined to determine whether basal 7ISG scores in tumour samples are associated with better prognosis, and whether any prognostic effects relate to *MYC* status.

Finally, further testing of the hypothesis included investigation into an additional *MYC*-deregulated model to study HKMTi-1-005-mediated immune effects. Lung adenocarcinoma (LuAd) tumours have frequent dysregulation of *MYC* [359], especially the large subset of LuAd with a mutation in *KRAS*; in these tumours, cMYC drives an immunosuppressive microenvironment and has been shown to be essential for tumourigenesis [309, 362]. Furthermore, IFN signalling specifically has been shown to be repressed by cMYC in *KRAS*-mutated LuAd [363].

The KMA model (genotype: $LSL\text{-}\boldsymbol{K}Ras^{G12D};Rosa26^{DM.lsl-\boldsymbol{M}YC/DS.lsl-\boldsymbol{A}POBEC3B}$) is a murine LuAd model that combines CRE-dependent alleles of endogenously expressed $KRas^{G12D}$ with *MYC* overexpression and *APOBEC3B* co-expression. APOBEC enzymes are cytidine deaminases that are involved in innate immune responses to viral infection and in the suppression of TEs [446], but are also implicated in tumourigenesis [447]. Hence, *APOBEC3B* deregulation in tumours could result in aberrant TE expression and responses [448], potentially contributing to a lower threshold for VM especially when found in combination with *MYC* alteration. Initial experiments performed by Dr Sarah Laing at the CRUK Scotland Institute demonstrated that HKMTi-1-005 treatment could induce immune stimulation and suppress proliferation in the KMA model, including upregulation of at least 4 of the 7 genes in the 7ISG signature [449].

These findings suggested *MYC*-deregulated LuAd may be sensitive to HKMTi-1-005-induced immune stimulation with anti-tumour efficacy. However, further investigations were required to validate that treatment efficacy was not dependent on APOBEC3B deregulation. The KM^2 model (genotype: $LSL\text{-}\boldsymbol{K}Ras^{G12D};Rosa26^{DM.lsl-\boldsymbol{M}YC/lsl-\boldsymbol{M}YC}$) overexpresses *MYC* alongside a $KRAS^{G12D}$ mutation to drive LuAd [450]. Hence, this highly *MYC*-dependent model provided a suitable framework to examine HKMTi-1-005 anti-tumour efficacy in an additional *MYC*-deregulated tumour type. LuAd investigations were performed by Dr Sarah Laing and Professor Daniel J Murphy at the CRUK Scotland Institute.

In summary, the aims and objectives of this chapter are to address the following questions:

- Does cell line evidence support cell-intrinsic repression of immune genes by cMYC, via formation of repressive complexes with EZH2 and G9a?
- Are these effects evident in patient-derived bulk tumour samples from *MYC*-deregulated tumour types?
- Do *MYC*-deregulated cell lines or tumours display an amplified level of cMYC-G9a/EZH2-dependent repression of immune genes?
- Can HKMTi-1-005 alleviate transcriptional repression by cMYC without promoting its oncogenic transcriptional activation effects?
- Given cell-intrinsic immune repression by cMYC, does sensitivity to HKMTi-1-005 relate to *MYC* status and basal immune status (particularly, the 7ISG)?
- Does the HKMTi-1-005-induced immune signature (7ISG) have prognostic significance in patient cohorts, and is this dependent on the *MYC* status of patients?
- Does the therapeutic benefit and immune stimulation provided by HKMTi-1-005 validate in LuAd, another *MYC*-deregulated tumour type?

5.2 cMYC-G9a and EZH2 repression anticorrelates with immune signalling scores: *in vitro* studies

5.2.1 cMYC-G9a/EZH2 repression inversely correlates with immune signalling in the GDSC2 cell line panel

The GDSC2 cell line panel was used to investigate repression by cMYC-G9a or EZH2, as well as the relationship between repression and immune signalling. Cell lines derived from solid tumour types with gene expression data were matched to include only those with sensitivity data for HKMTi-1-005 and single EZH2i/G9ai, as described in chapter 3 (n=565). Each line was scored for four immune genesets (the 7ISG, IFNA response, IFNG response and innate immune response), as well as for genesets reflecting cMYC-G9a repression (MYC_G9A_REPRESSION_TARGETS) [268] or EZH2 repression (EZH2_REPRESSION_TARGETS) [109]. No cMYC-EZH2 repression geneset was available in the literature at the time of writing. Therefore, in place of choosing one of the individual EZH2 repression genesets described in section 3.2.5, the EZH2 repression geneset used here was based on a meta-analysis of consistent gene expression changes from 18 different studies assessing EZH2 knockdown effects (analysis performed by Curry et al. [109]). An additional geneset for canonical (also referred to as hallmark) cMYC signalling (HALLMARK_MYC_TARGETS_V1) was included to compare relationships between repressive versus activating cMYC effects. These genesets are listed in table 2.6. Scoring was performed using the singscore package, a single-sample scoring method that allows for analysis of gene expression profiles without introducing bias due to differences in sample composition [418].

First, the relationships between cMYC-G9a repression, EZH2 repression and canonical hallmark cMYC signalling scores were assessed. Note that a higher score for the cMYC-G9a or EZH2 repression genesets indicates higher expression of the genes in these sets, reflecting less repression. Scores for cMYC-G9a and EZH2 repression highly correlated with each other across cell lines (figure 5.1), suggesting the two repressive environments often co-exist. Additionally, higher cMYC-G9a/EZH2 scores anticorrelated with canonical cMYC signalling (figure 5.1), suggesting high cMYC activity in a cell results in both greater repression and activation of the respective cMYC targets. However, the relationship between cMYC-G9a repression and EZH2 repression was significantly stronger than either relationship with hallmark cMYC signalling (cMYC-G9a repression versus EZH2 repression 95% CI [0.69,0.77], cMYC-G9a repression versus canonical cMYC signalling 95% CI [-0.39,-0.24], EZH2 repression versus canonical cMYC signalling 95% CI [-0.62,-0.51]).

Next, cMYC-G9a/EZH2 repression and canonical cMYC signalling were compared to immune scores. The correlation between cMYC-G9a/EZH2 repression and each of the four immune genesets

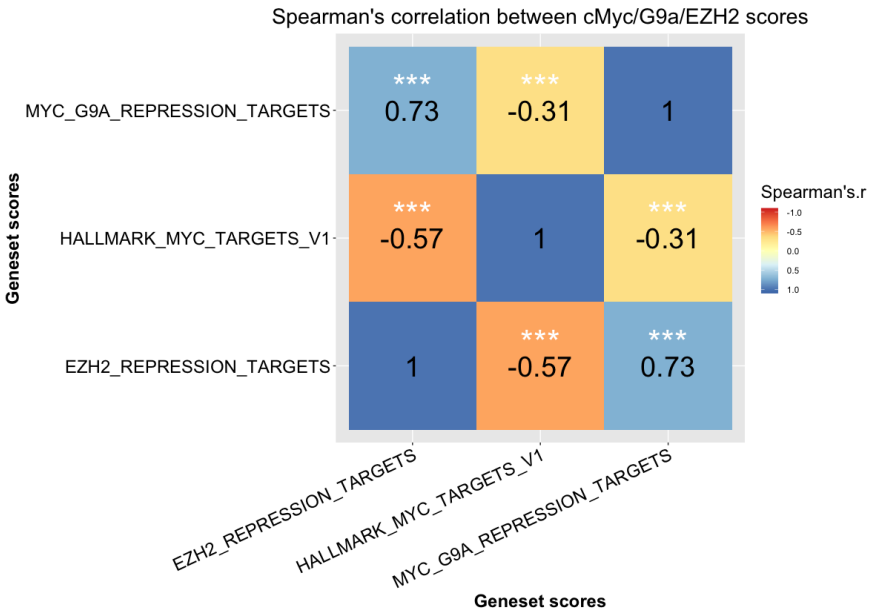


Figure 5.1 **Correlation between cMYC-G9a/EZH2 repression scores and canonical cMYC signalling in the GDSC2 cell line panel.** Spearman's correlation between cMYC-G9a repression scores, EZH2 repression scores and canonical hallmark cMYC signalling scores across solid, matched cell lines from the GDSC2 cell line panel with gene expression data (n=565). Strength of correlation depicted by the tile colour, see legend. MYC_G9A_REPRESSION_TARGETS = cMYC-G9a repression, EZH2_REPRESSION_TARGETS = EZH2 repression, HALLMARK_MYC_TARGETS_V1 = canonical cMYC signalling. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

demonstrated a highly significant relationship between high levels of repression – indicated by lower scores – and lower immune signalling (figure 5.2). In contrast, higher scores for hallmark cMYC signalling were highly anticorrelated with immune scores (figure 5.2). Considering *MYC*-amplified lines (n=65) separately from those without *MYC* amplification (n=500), it was notable that the relationship between EZH2 repression and lower immune signalling was significantly stronger in *MYC*-amplified lines (figure 5.3, see appendix table A.7 for CIs). The relationship between cMYC-G9a repression and immune signalling demonstrated the same trend, but with overlapping confidence intervals preventing statistical distinction between the two groups (see appendix table A.8 for CIs).

5.2.2 *MYC*-amplified cell lines have greater cMYC-G9a/EZH2 repression with lower immune scores

Having established that *MYC* amplification impacted the strength of the correlation between cMYC-G9a/EZH2 repression and immune signalling, I next determined whether amplification affected absolute levels of cMYC-G9a/EZH2 repression as inferred by geneset scores. Comparison of *MYC*-amplified to non-amplified lines revealed greater cMYC-G9a/EZH2 repression (i.e. lower expression scores) in amplified lines (figure 5.4), accompanied by lower scores for immune signalling and higher scores for canonical cMYC signalling (figure 5.5). Overall, these GDSC2 data support a cell-intrinsic anticorrelation between immune signalling and cMYC-G9a/EZH2 repression that is stronger

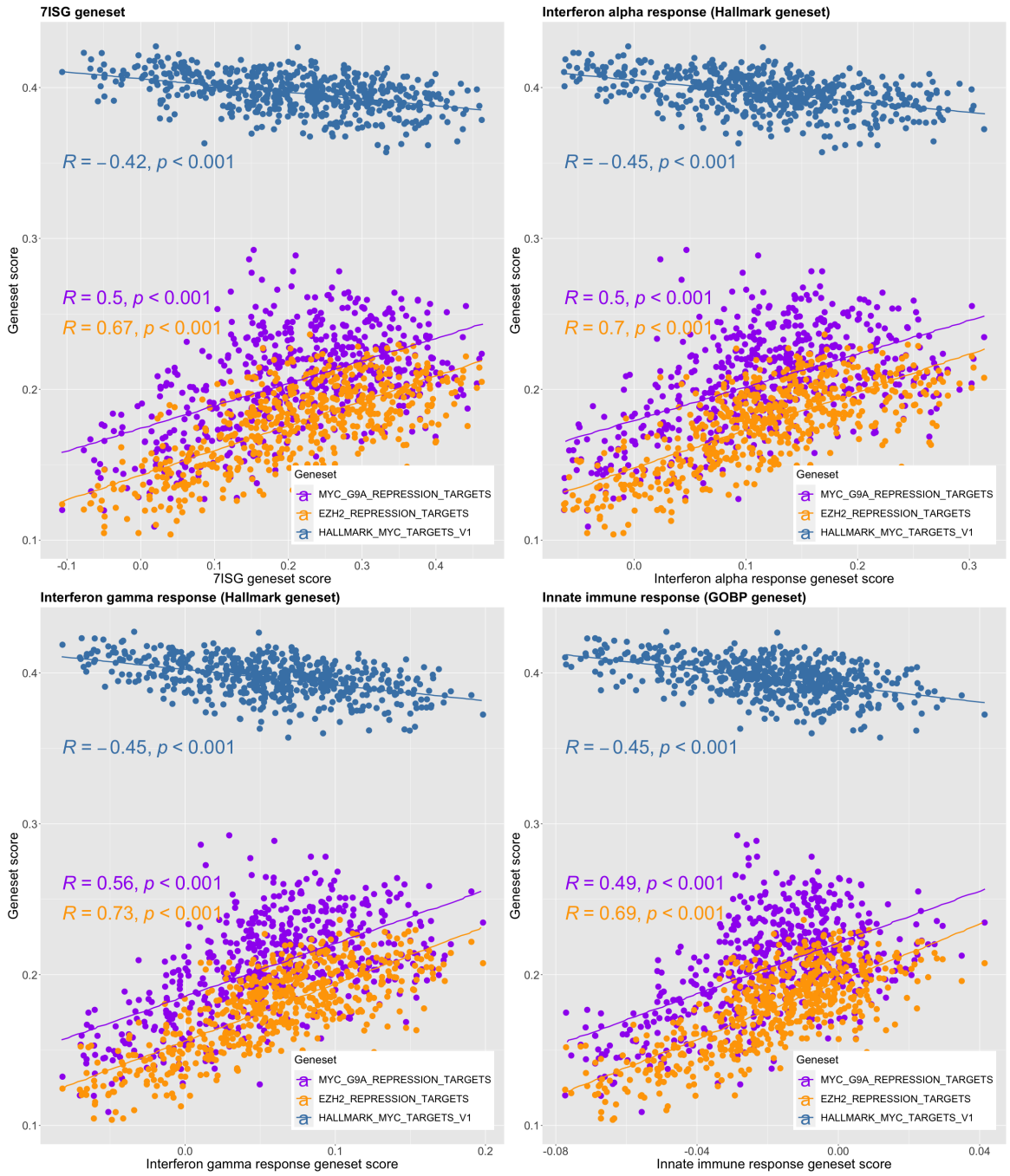


Figure 5.2 **Correlation between immune genesets and genesets relating to the extent of cMYC-G9a/EZH2 repression or the extent of hallmark cMYC activity in the GDSC2 cell line panel.** Spearman's correlation between immune genesets (x-axes) and genesets relating to the extent of cMYC-G9a repression, the extent of EZH2 repression, or the extent of hallmark cMYC activity (y-axes). Assessed across a panel of 565 matched cell lines derived from solid tumours, from the GDSC2 database. Immune genesets from left to right, top to bottom: the 7ISG geneset (derived in chapter 4), HALLMARK_INTERFERON_ALPHA_RESPONSE = IFNA signalling, HALLMARK_INTERFERON_GAMMA_RESPONSE = IFNG signalling, GOBP_INNATE_IMMUNE_RESPONSE = innate immune response. MYC_G9A_REPRESSION_TARGETS (purple) = cMYC-G9a repression, EZH2_REPRESSION_TARGETS (orange) = EZH2 repression, HALLMARK_MYC_TARGETS_V1 (blue) = canonical cMYC signalling.

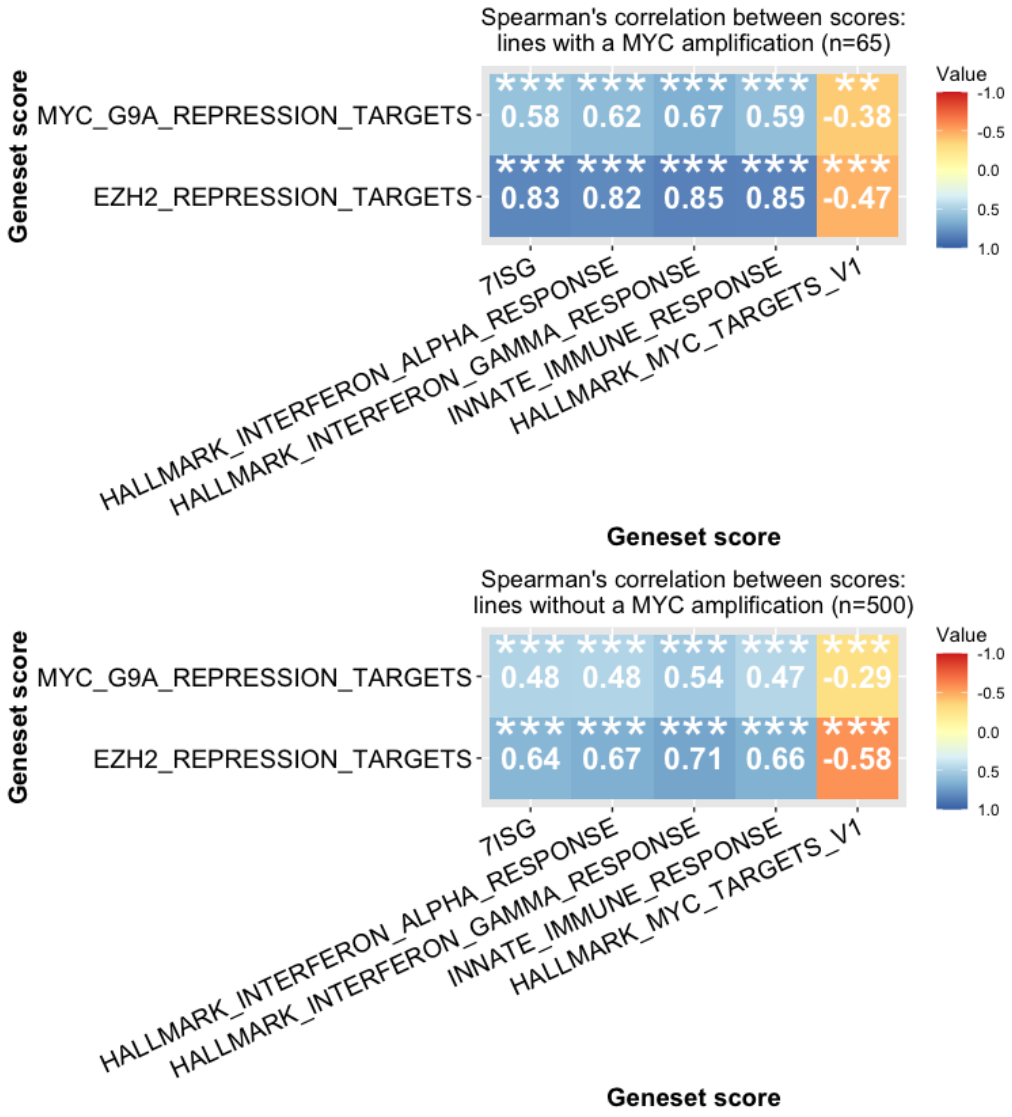


Figure 5.3 **Correlation between cMYC-G9a/EZH2 repression scores and canonical MYC signalling or immune signalling scores in the GDSC2 cell line panel: stratified by *MYC* status.** Spearman's correlation between immune genesets or the extent of hallmark cMYC activity (x-axes) and genesets relating to the extent of cMYC-G9a or EZH2 repression (y-axes). Cell lines are stratified by *MYC* amplification status: lines with *MYC* amplification (top, n=65) tend towards a stronger relationship between cMYC-G9a/EZH2 repression and immune scores than non-amplified lines (bottom, n=500), with the difference between these groups reaching statistical significance for EZH2 repression correlations. Confidence intervals listed in appendix tables A.7 and A.8. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

in *MYC*-amplified lines than non-amplified lines. In contrast, higher hallmark cMYC signalling positively correlated with lower immune scores. The absolute level of cell-intrinsic cMYC-G9a/EZH2 repression was higher in lines with *MYC* amplification, as was the extent of canonical hallmark cMYC signalling.

The GDSC2 panel contains cell lines from an array of tumour types. To mitigate any confounding effects due to tissue-specific results, I next looked to validate the relationships described within tumour types.

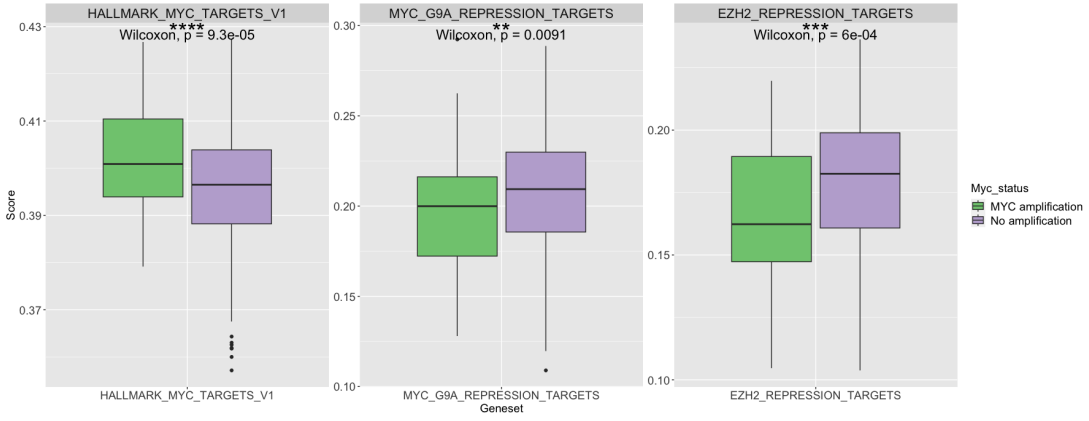


Figure 5.4 ***MYC* amplification versus cMYC-G9a/EZH2 scores in GDSC2 cell lines.** Box plots demonstrating the relationship between *MYC* amplification and scores for hallmark cMYC signalling and cMYC-G9a/EZH2 repression. Relationships assessed across a panel of 565 matched cell lines derived from solid tumours, from the GDSC2 panel. MYC_G9A_REPRESSION_TARGETS = cMYC-G9a repression, EZH2_REPRESSION_TARGETS = EZH2 repression, HALLMARK_MYC_TARGETS_V1 = canonical cMYC signalling. * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

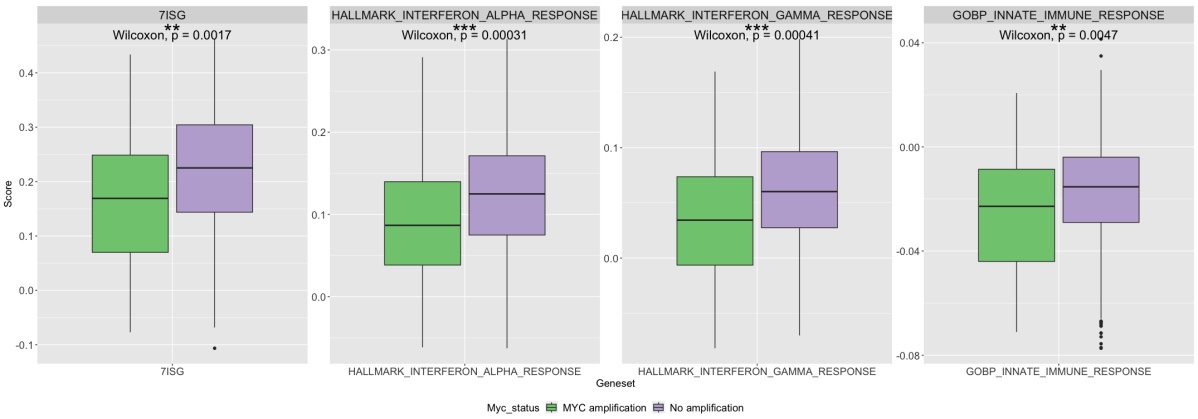


Figure 5.5 ***MYC* amplification versus immune scores in GDSC2 cell lines.** Box plots demonstrating the relationship between *MYC* amplification and scores for the 7ISG, IFNA response, IFNG response and innate immune response genesets. Relationships assessed across a panel of 565 matched cell lines derived from solid tumours, from the GDSC2 panel. 7ISG = HKMTi-1-005-induced immune signature, derived in chapter 4. HALLMARK_INTERFERON_ALPHA_RESPONSE = IFNA signalling, HALLMARK_INTERFERON_GAMMA_RESPONSE = IFNG signalling and GOBP_INNATE_IMMUNE_RESPONSE = innate immune response. * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

5.2.3 Pan-cancer cell line observations are corroborated in individual analyses of ovarian, breast and LuAd cell lines

To achieve maximum statistical power in subsequent analyses, cell lines included in the GDSC2 panel were first counted by tumour type to identify cancers with the greatest number of representative lines. The most abundant cell lines were those derived from LuAd (n=49), breast (n=41), large intestine (n=37), SCLC (n=34), glioma (n=32) and ovarian (n=32) cancers. Of these, those that exhibited the most frequent *MYC* amplification in the corresponding patient data from The Cancer Genome Atlas (TCGA) project were chosen, resulting in the selection of ovarian, breast and LuAd cell line cohorts [121–123].

Correlation analyses between cMYC-G9a/EZH2 repression and canonical cMYC signalling in these cohorts demonstrated strong support for findings across the pan-cancer GDSC2 panel (figure 5.6). Further, in almost every case, higher immune signalling scores correlated with scores reflecting lower cMYC-G9a/EZH2 repression; where this relationship did not reach significance, a strong trend in the same direction was observed ($p <= 0.1$ in all cases) (figure 5.7). The relationship between canonical cMYC signalling and immune scores was slightly more variable. In the ovarian and LuAd lines, more canonical cMYC signalling still correlated significantly with lower immune signalling in almost all cases (with $p <= 0.1$ in all cases) (figures 5.7a and 5.7c). However, in the breast cancer cell lines, there was no significant relationship between canonical cMYC signalling and immune scores (figure 5.7b). Note that too few *MYC*-amplified cell lines exist for each of these tumour types to assess *MYC* amplification effect on the strength of correlation or on absolute scores (n=2, n=7 and n=8 respectively).

Taken together, the data suggest that in ovarian, breast and LuAd cancer cell lines there is a cell-intrinsic relationship between the extent of cMYC-G9a/EZH2 repression and immune signalling. Further, they suggest that this relationship is specific to cMYC-G9a/EZH2 repression over canonical cMYC signalling.

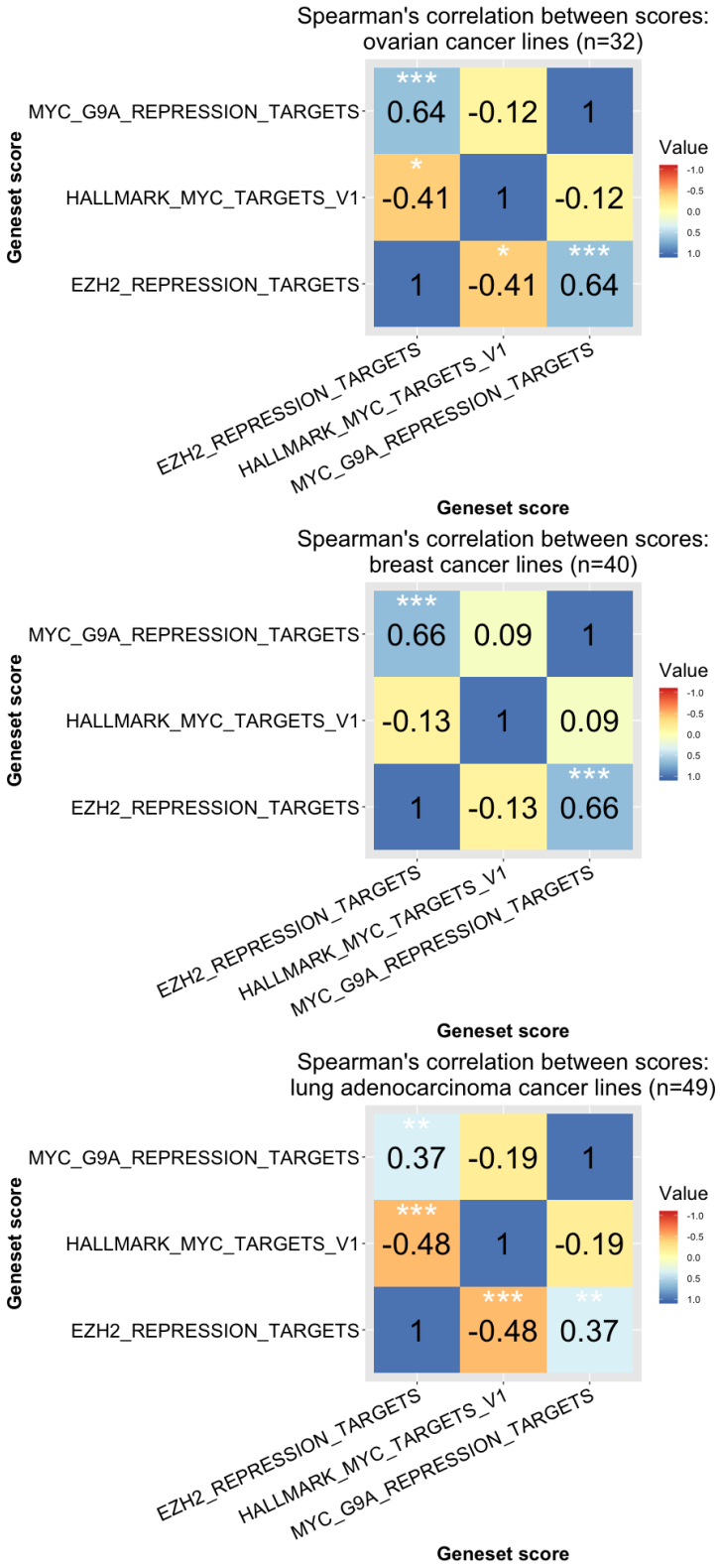
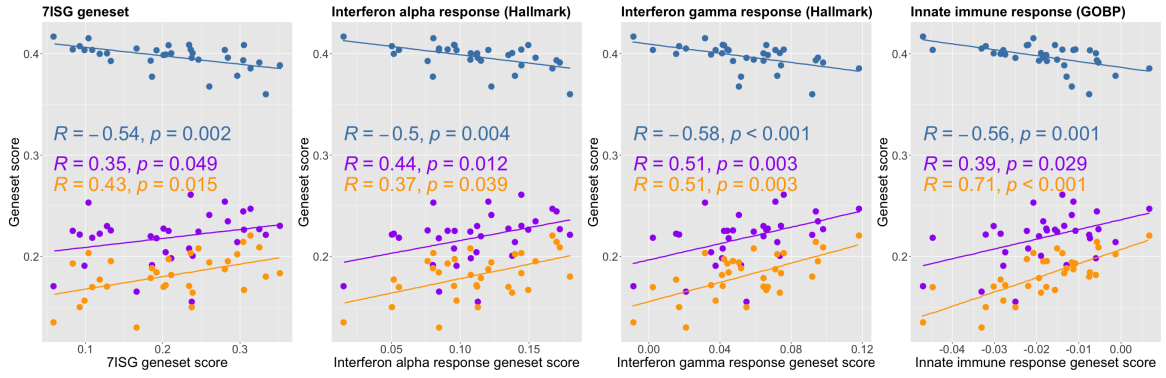
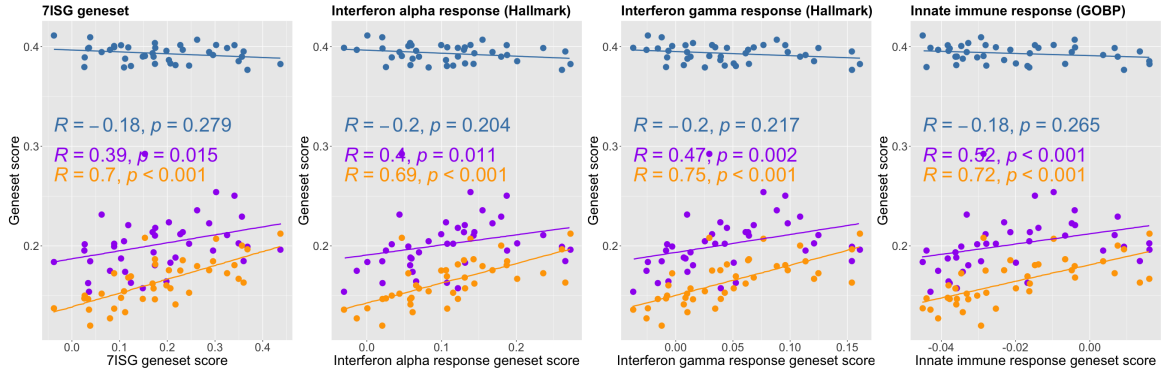


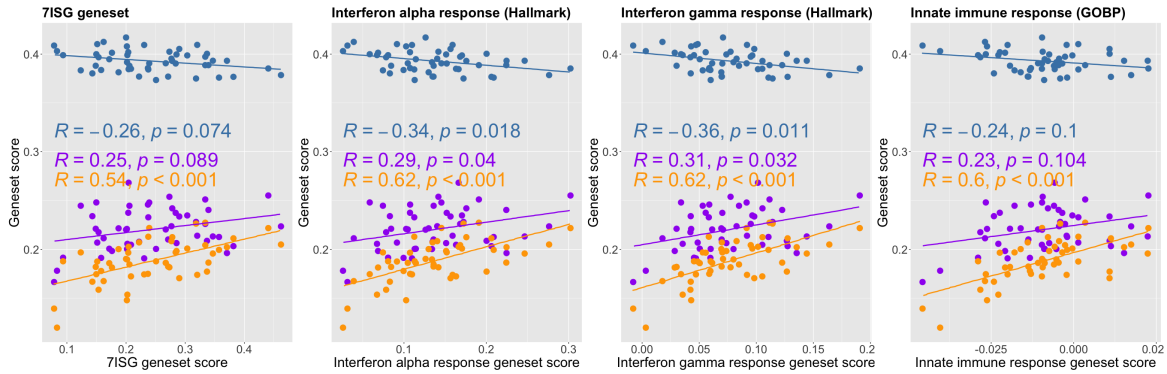
Figure 5.6 **Correlation between cMYC-G9a/EZH2 repression scores and canonical cMYC signalling in ovarian, breast and LuAd cell lines.** Spearman's correlation between cMYC-G9a repression scores, EZH2 repression scores and canonical hallmark cMYC signalling scores across ovarian cancer cell lines (top, n=32), breast cancer cell lines (middle, n=40; one breast line in GDSC2 panel lacks gene expression data) and lung adenocarcinoma cell lines (n=49, bottom). Strength of correlation depicted by the tile colour, see legend. MYC_G9A_REPRESSION_TARGETS = cMYC-G9a repression, EZH2_REPRESSION_TARGETS = EZH2 repression, HALLMARK_MYC_TARGETS_V1 = canonical cMYC signalling. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.



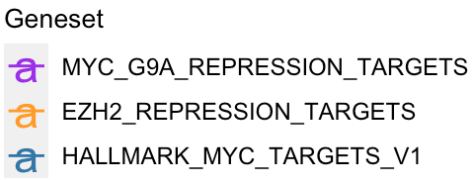
(a)



(b)



(c)



(d)

Figure 5.7 **Correlation between immune genesets and genesets relating to the extent of cMYC-G9a/EZH2 repression or the extent of hallmark cMYC activity in ovarian, breast and LuAd cell lines.** Spearman's correlations across (a) ovarian cancer cell lines (n=32), (b) breast cancer cell lines (n=40) and (c) LuAd cell lines (n=49). Immune genesets from left to right: the 7ISG geneset (derived in chapter 4), HALLMARK_INTERFERON_ALPHA_RESPONSE = IFNA signalling, HALLMARK_INTERFERON_GAMMA_RESPONSE = IFNG signalling, GOBP_INNATE_IMMUNE_RESPONSE = innate immune response. MYC_G9A_REPRESSION_TARGETS (purple) = cMYC-G9a repression, EZH2_REPRESSION_TARGETS (orange) = EZH2 repression, HALLMARK_MYC_TARGETS_V1 (blue) = canonical cMYC signalling. Legend shown in (d).

5.3 cMYC-G9a and EZH2 repression anticorrelates with immune signalling scores: *in vivo* studies

5.3.1 cMYC-G9a/EZH2 repression inversely correlates with immune signalling in ovarian, breast and LuAd patient cohorts

To complement the assessment of cMYC-G9a/EZH2-immune relationships within individual tumour types *in vitro*, TCGA data from the ovarian (TCGA-OV), breast (TCGA-BRCA) and LuAd (TCGA-LUAD) cohorts were analysed. An additional ovarian patient cohort hosted by the International Cancer Genome Consortium (ICGC) was also included.

These cohorts were scored for cMYC-G9a/EZH2 repression and hallmark cMYC signalling, as well as for the 7ISG, IFNA, IFNG and innate immune response genesets. The data confirmed the initial cell line observation of a strong positive correlation between cMYC-G9a and EZH2 repression scores (figure 5.8). Additionally, an anticorrelation between these repression genesets and canonical cMYC signalling was observed in all cases excepting the ICGC cohort, where no significant correlation with canonical cMYC signalling was observed (figure 5.8).

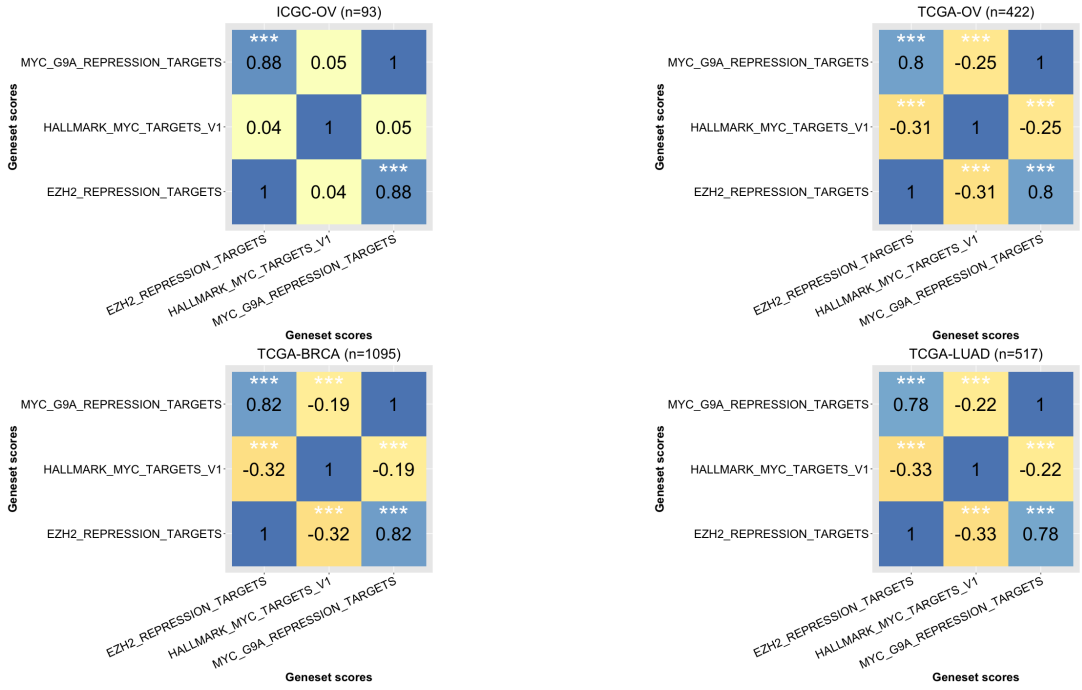
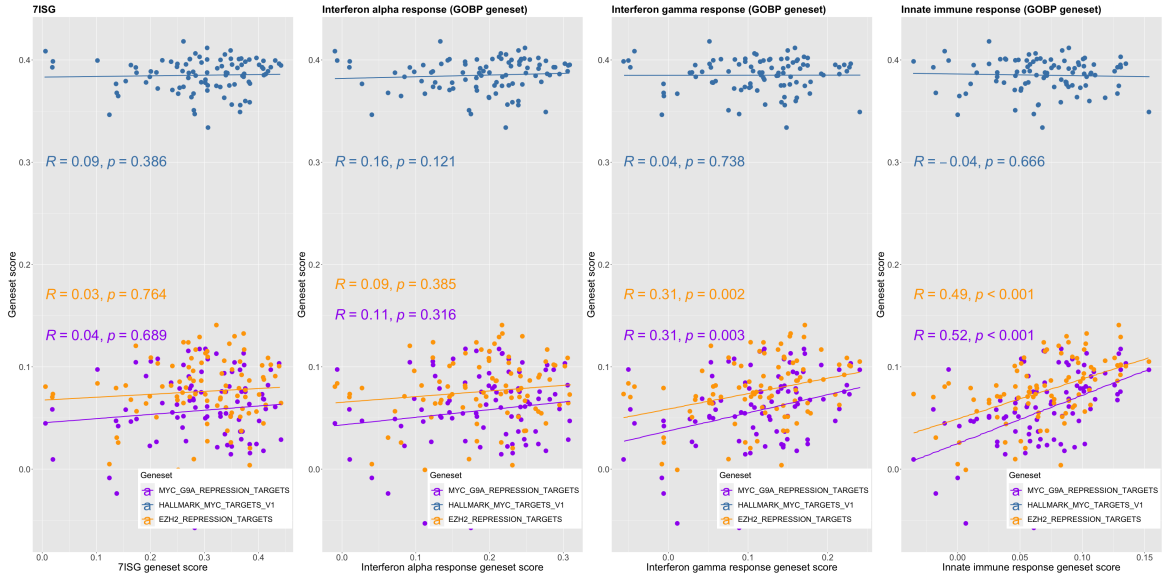
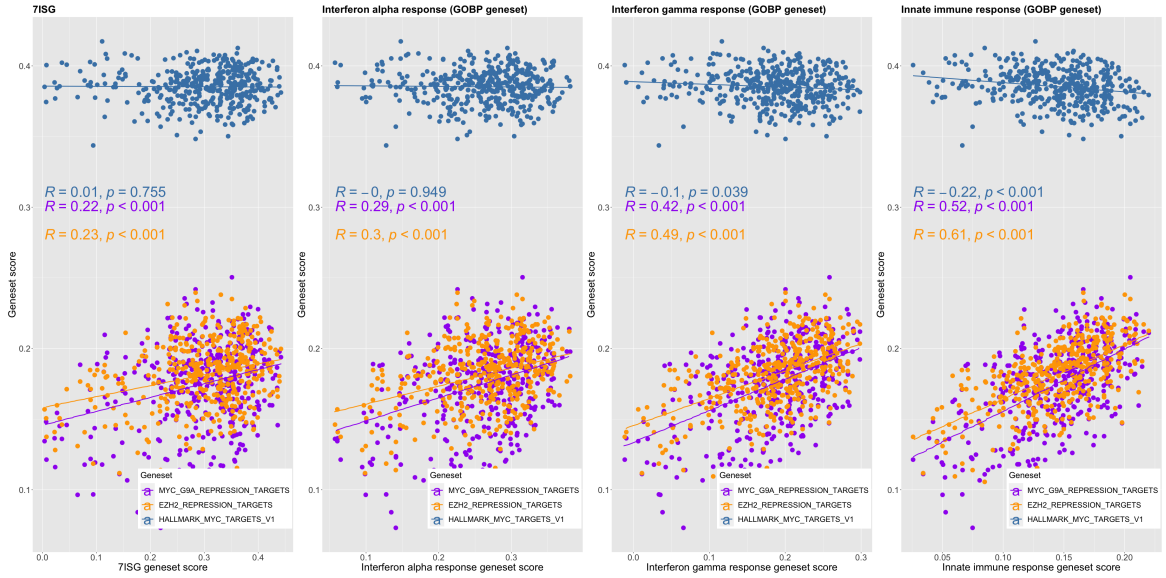


Figure 5.8 **Correlation between cMYC-G9a/EZH2 repression scores and canonical cMYC signalling in *in vivo* patient cohorts.** Spearman's correlation between cMYC-G9a repression scores, EZH2 repression scores and canonical hallmark cMYC signalling scores across: the ICGC-OV (n=93), TCGA-OV (n=422), TCGA-BRCA (n=1095) and TCGA-LUAD (n=517) patient cohorts (left to right, top to bottom). Strength and direction of correlation depicted by the tile colour, from negative (deeper yellow) to positive (deeper blue). MYC_G9A_REPRESSION_TARGETS = cMYC-G9a repression, EZH2_REPRESSION_TARGETS = EZH2 repression, HALLMARK_MYC_TARGETS_V1 = canonical cMYC signalling. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.



(a) ICGC-OV cohort



(b) TCGA-OV cohort

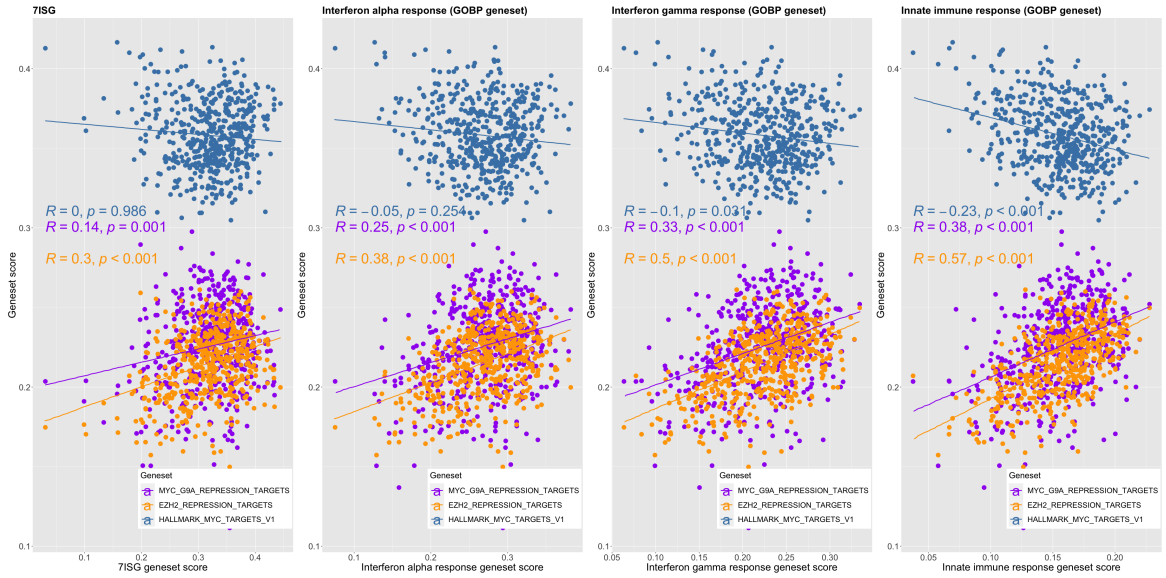
Figure 5.9 **Correlation between immune genesets and genesets relating to the extent of cMYC-G9a/EZH2 repression or the extent of hallmark cMYC activity in *in vivo* patient cohorts.** Continued overleaf.

Further, these data once again demonstrated a strong correlation between lower levels of repression by cMYC-G9a/EZH2 and higher levels of immune signalling (figure 5.9). In contrast, canonical hallmark cMYC signalling was either negatively correlated with immune signalling or exhibited a non-significant relationship (figure 5.9). The relationships were least evident in the smaller ICGC cohort of 93 patients where IFNA and 7ISG score correlations did not reach significance (figure 5.9a). However, IFNG and innate immune signalling scores still illustrated a strong and significant anticorrelation with cMYC-G9a/EZH2 repression.

The exception to these observations was in the TCGA-BRCA cohort, where there was a weak positive correlation between hallmark cMYC signalling and 7ISG/IFNA/IFNG signalling (figure 5.9c). However, hallmark cMYC signalling relationships with IFNG or innate immune response



(c) TCGA-BRCA cohort



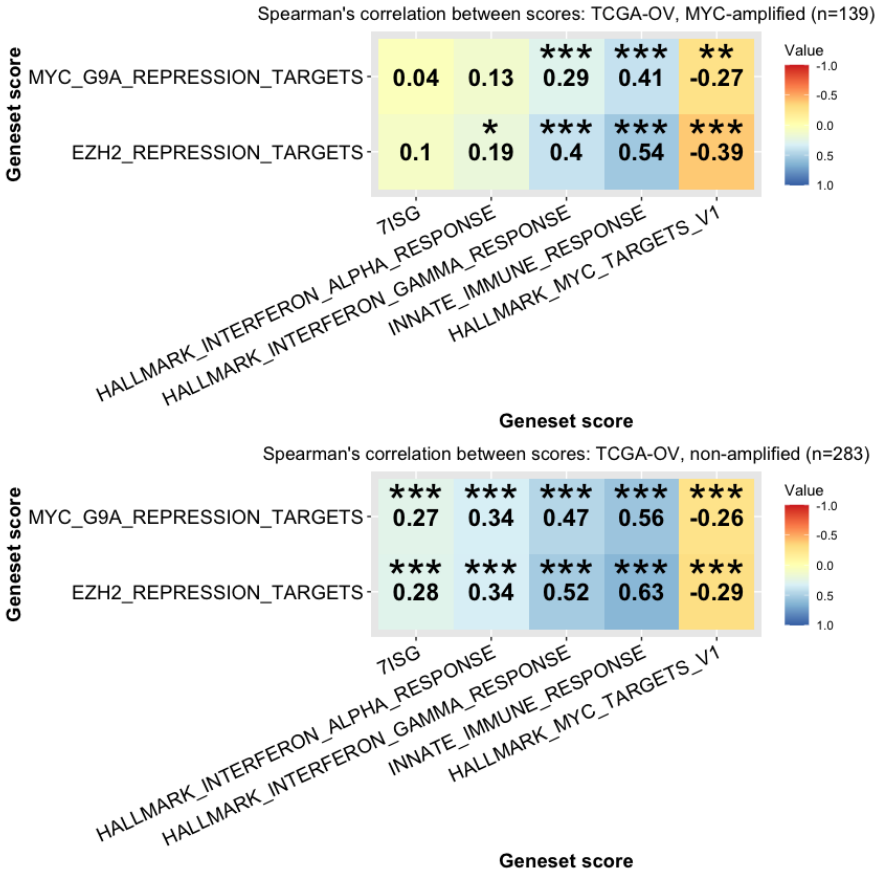
(d) TCGA-LUAD cohort

Figure 5.9 **Correlation between immune genesets and genesets relating to the extent of cMYC-G9a/EZH2 repression or the extent of hallmark cMYC activity in *in vivo* patient cohorts.** Spearman's correlation between immune genesets (x-axes) and genesets relating to the extent of cMYC-G9a repression, EZH2 repression or hallmark cMYC activity (y-axes). Correlations assessed across the (a) ICGC-OV cohort (n=93), (b) TCGA-OV cohort (n=422), (c) TCGA-BRCA cohort (n=1095) and (d) TCGA-LUAD cohort (n=517). Immune genesets from left to right: the 7ISG geneset (derived in chapter 4), HALLMARK_INTERFERON_ALPHA_RESPONSE = IFNA signalling, HALLMARK_INTERFERON_GAMMA_RESPONSE = IFNG signalling, GOBP_INNATE_IMMUNE_RESPONSE = innate immune response. MYC_G9A_REPRESSION_TARGETS (purple) = cMYC-G9a repression, EZH2_REPRESSION_TARGETS (orange) = EZH2 repression, HALLMARK_MYC_TARGETS_V1 (blue) = canonical cMYC signalling.

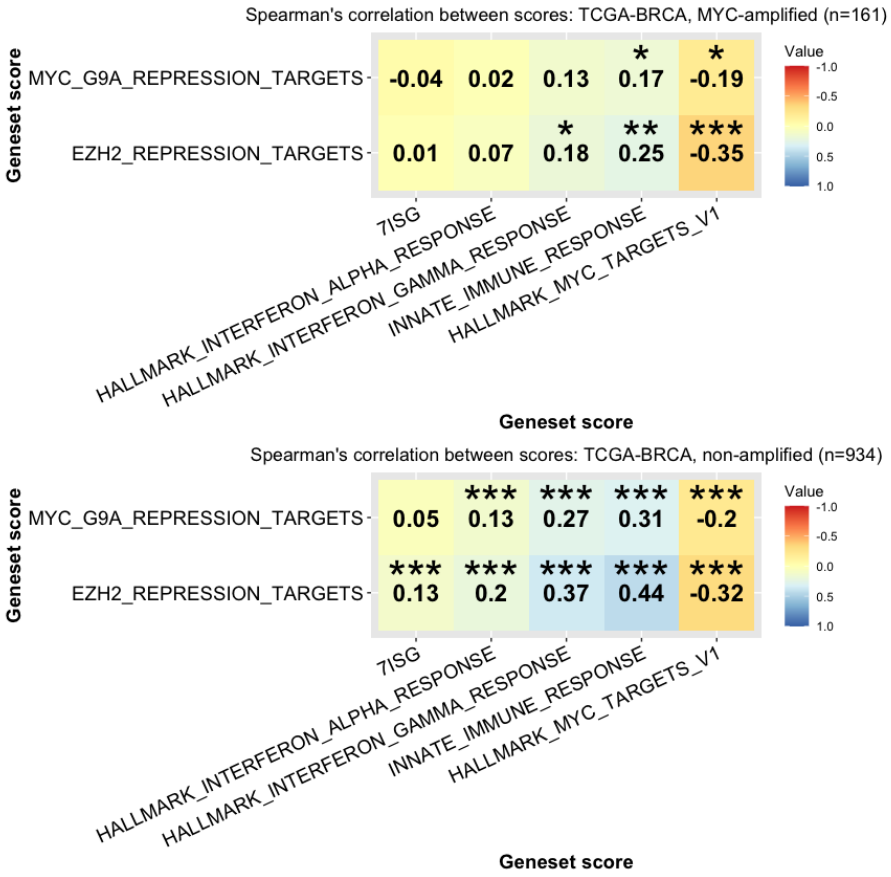
scores were significantly weaker than those of cMYC-G9a/EZH2 repression with these immune sets in the TCGA-BRCA cohort (see appendix table A.9 for CIs). Hence, the data support a strong and specific correlation between cMYC-G9a/EZH2 repression and immune signalling – especially IFNG and innate immune signalling – in the TCGA-OV, TCGA-BRCA and TCGA-LUAD cohorts.

The TCGA cohorts were subsequently annotated with *MYC* amplification status to examine any effects of *MYC* amplification on the strength of repression-immune score correlations. The ICGC cohort was not included in this analysis of *MYC*-dependence, as too few samples were annotated as *MYC*-amplified (the ICGC data portal includes copy number, ploidy and whole genome duplication data, but not for all samples. Using COSMIC criteria [407] on the annotated subset, only 7 samples were found to be *MYC*-amplified).

In contrast to the GDSC2 cell line panel, there was no significant relationship between *MYC* amplification and the strength of cMYC-G9a/EZH2 and immune score correlations (figure 5.10). Indeed, in the TCGA-OV and TCGA-BRCA groups there was a trend towards stronger repression-immune relationships in the non-amplified samples, whilst in the TCGA-LUAD cohort the opposite trend was apparent. These differences in correlation strength did not reach significance in any of the cohorts (overlapping CIs).

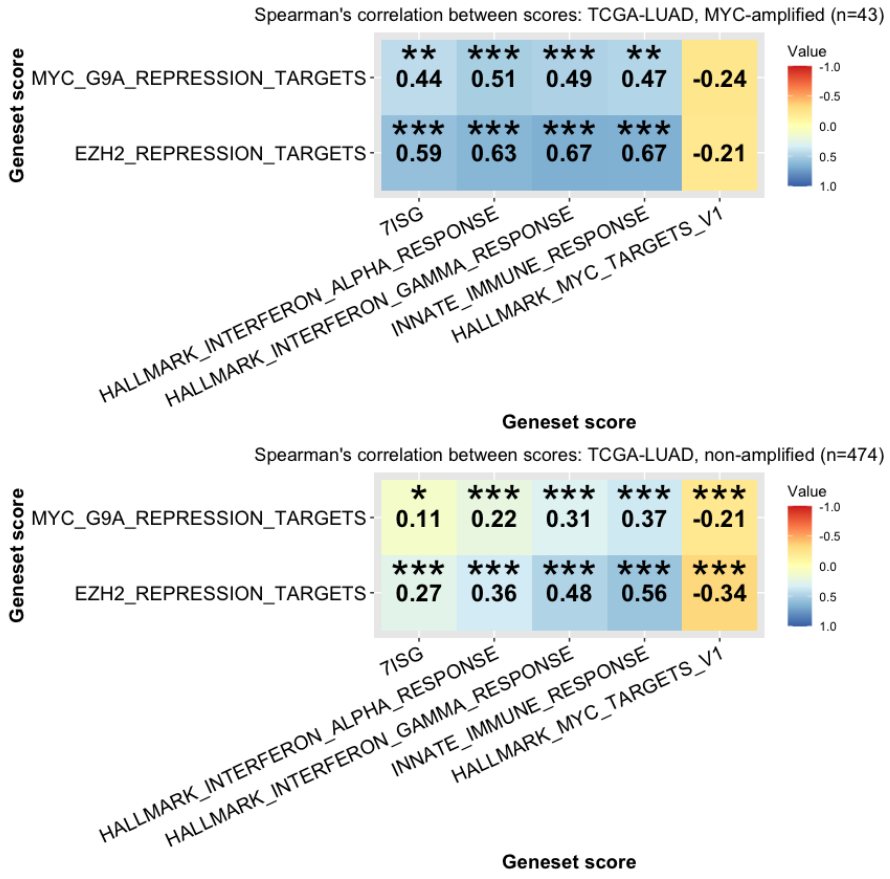


(a)



(b)

Figure 5.10 **Correlation between cMYC-G9a/EZH2 repression scores and canonical MYC signalling or immune signalling scores in *in vivo* patient cohorts: stratified by *MYC* status.** Continued overleaf.

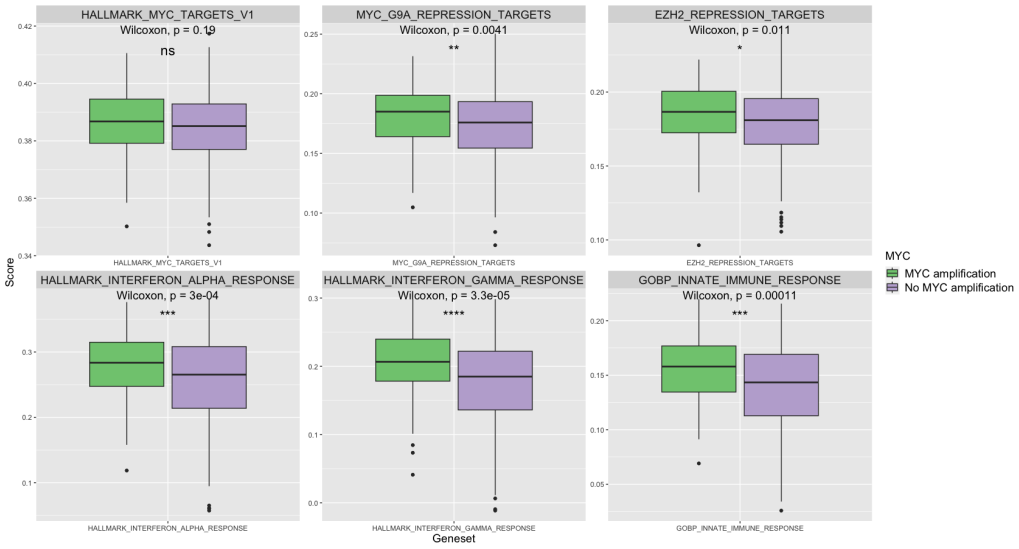


(c)

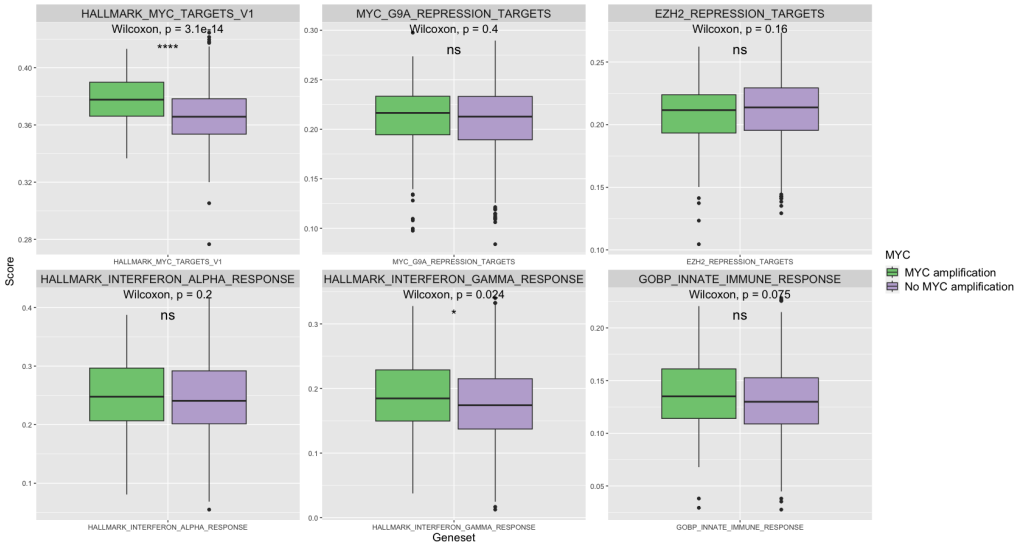
Figure 5.10 **Correlation between cMYC-G9a/EZH2 repression scores and canonical MYC signalling or immune signalling scores in *in vivo* patient cohorts: stratified by *MYC* status.** Spearman's correlation between immune genesets or genesets reflecting the extent of hallmark cMYC activity (x-axes) and genesets relating to the extent of cMYC-G9a or EZH2 repression, (y-axes) in patient cohorts, stratified by *MYC* amplification status. Correlations assessed across (a) the TCGA-OV cohort (n=422), (b) the TCGA-BRCA cohort (n=1095) and (c) the TCGA-LUAD cohort (n=517). In each case, *MYC*-amplified samples are depicted in the top graph, non-amplified in the bottom graph. Immune genesets: the 7ISG geneset (derived in chapter 4), HALLMARK_INTERFERON_ALPHA_RESPONSE = IFNA signalling, HALLMARK_INTERFERON_GAMMA_RESPONSE = IFNG signalling, GOBP_INNATE_IMMUNE_RESPONSE = innate immune response. MYC_G9A_REPRESSION_TARGETS = cMYC-G9a repression, EZH2_REPRESSION_TARGETS = EZH2 repression, HALLMARK_MYC_TARGETS_V1 = canonical cMYC signalling. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

5.3.2 *MYC*-amplified patient samples do not consistently show greater cMYC-G9a/EZH2 repression or lower immune scores

The effect of *MYC* amplification on absolute scores for cMYC-G9a/EZH2 repression and immune genesets was next examined. The results once again demonstrated a distinct effect to the GDSC2 cell line panel: *MYC* amplification was not consistently related to higher cMYC-G9a/EZH2 repression or lower immune signalling (figures 5.11 and 5.12). Instead, in the TCGA-OV cohort the scores suggested lower cMYC-G9a/EZH2 repression and higher immune signalling in *MYC*-amplified samples (figures 5.11a and 5.12), whilst in the TCGA-LUAD and TCGA-BRCA cohorts the majority of *MYC*-amplified sample scores were not significantly different to non-amplified scores (figures 5.11b, 5.11c and 5.12). The TCGA-BRCA cohort did demonstrate higher canonical cMYC signalling in *MYC*-amplified samples, but this group also showed higher IFNG response scores (figure 5.11b).

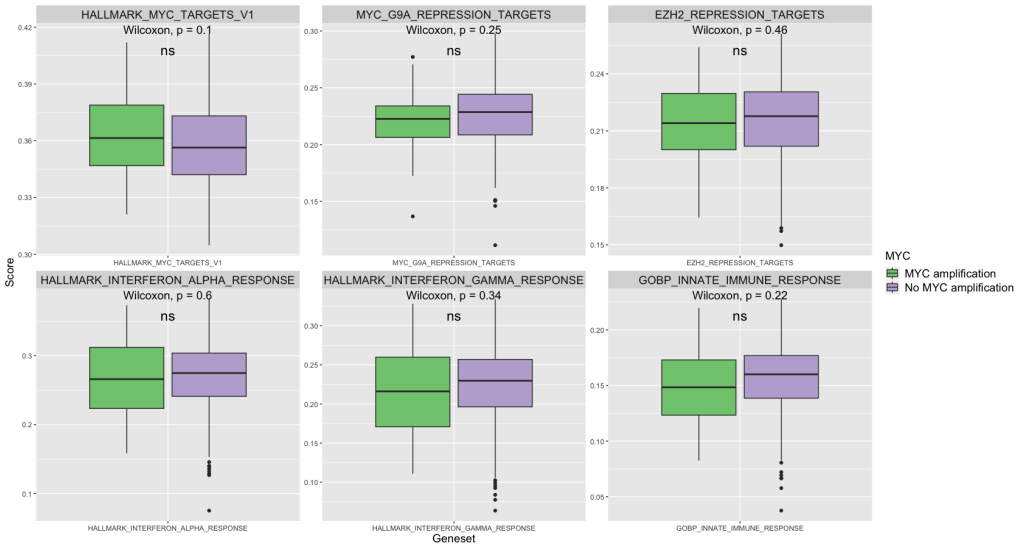


(a)



(b)

Figure 5.11 ***MYC* amplification versus cMYC-G9a/EZH2 scores or immune scores in *in vivo* patient cohorts.** Continued overleaf.



(c)

Figure 5.11 ***MYC* amplification versus cMYC-G9a/EZH2 scores or immune scores in *in vivo* patient cohorts.** Box plots demonstrating the relationship between *MYC* amplification and canonical cMYC signalling, cMYC-G9a/EZH2 repression or immune signalling scores. Relationships assessed across (a) the TCGA-OV cohort (n=422), (b) the TCGA-BRCA cohort (n=1095) and (c) the TCGA-LUAD cohort (n=517). MYC_G9A_REPRESSION_TARGETS = cMYC-G9a repression, EZH2_REPRESSION_TARGETS = EZH2 repression, HALLMARK_MYC_TARGETS_V1 = canonical cMYC signalling, HALLMARK_INTERFERON_ALPHA_RESPONSE = IFNA signalling, HALLMARK_INTERFERON_GAMMA_RESPONSE = IFNG signalling and GOBP_INNATE_IMMUNE_RESPONSE = innate immune response. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

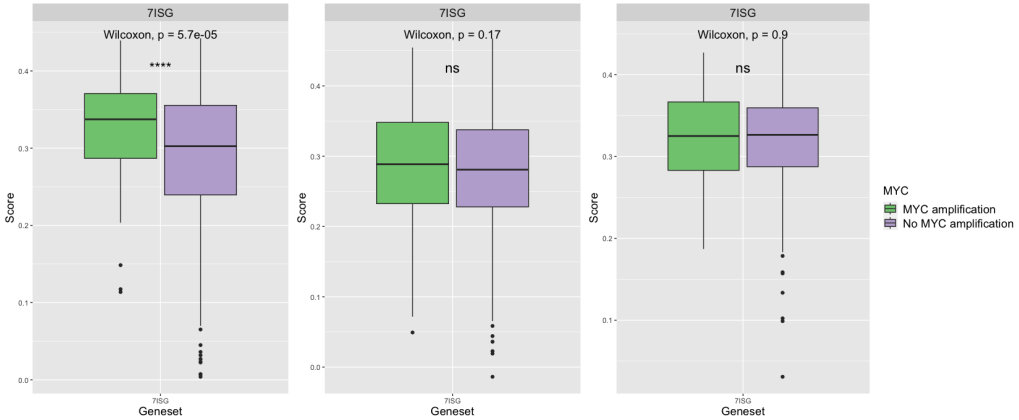


Figure 5.12 ***MYC* amplification versus 7ISG scores in *in vivo* patient cohorts.** Box plots demonstrating the relationship between *MYC* amplification and 7ISG scores. Relationships assessed across the TCGA-OV cohort (n=422), the TCGA-BRCA cohort (n=1095) and the TCGA-LUAD cohort (n=517) (left to right). MYC_G9A_REPRESSION_TARGETS = cMYC-G9a repression, EZH2_REPRESSION_TARGETS = EZH2 repression, HALLMARK_MYC_TARGETS_V1 = canonical cMYC signalling, HALLMARK_INTERFERON_ALPHA_RESPONSE = IFNA signalling, HALLMARK_INTERFERON_GAMMA_RESPONSE = IFNG signalling and GOBP_INNATE_IMMUNE_RESPONSE = innate immune response. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

5.3.3 *MYC* amplification status impacts immune infiltration in patient samples

The GDSC2 data provided strong support for the hypothesis that *MYC* amplification is accompanied by increased cMYC-G9a/EZH2 repression and decreased immune signalling in a cancer cell-intrinsic manner. Hence, I looked to determine whether non-tumour cells may be contributing to the discordant results in the patient data that are derived from bulk tumour samples. Particularly, significant literature demonstrates an effect of *MYC* status on the extent of immune cell infiltration [279, 451, 452]; differences in immune infiltration in *MYC*-amplified samples could result in changes to cMYC-G9a/EZH2 and immune scores that are not reflective of cancer cell-intrinsic effects.

Vast numbers of different immune inference methods exist for use on bulk sequencing data [411, 453]. CIBERSORTx is a deconvolution approach that uses gene expression to infer abundance of immune cell types in bulk tissue, based on signature matrices for each cell type [454]. The CIBERSORTx platform was used in absolute mode to infer immune composition in the TCGA samples and determine the effect of *MYC* amplification status on levels of immune infiltration. In parallel, relative immune infiltration was calculated using the Consensus^{TME} (cTME) package [411]. This approach compiles cell type-specific gene expression defined by seven published methods, with benchmarking against TCGA datasets. Additionally, it allows for specification of the tumour type being investigated to account for cell type-specific gene expression and hence is a robust method for estimation of relative cell type infiltration in bulk sequencing data. An example of CIBERSORTx and cTME results are provided based on scoring of the ICGC-OV samples (figures 5.13 and 5.14).

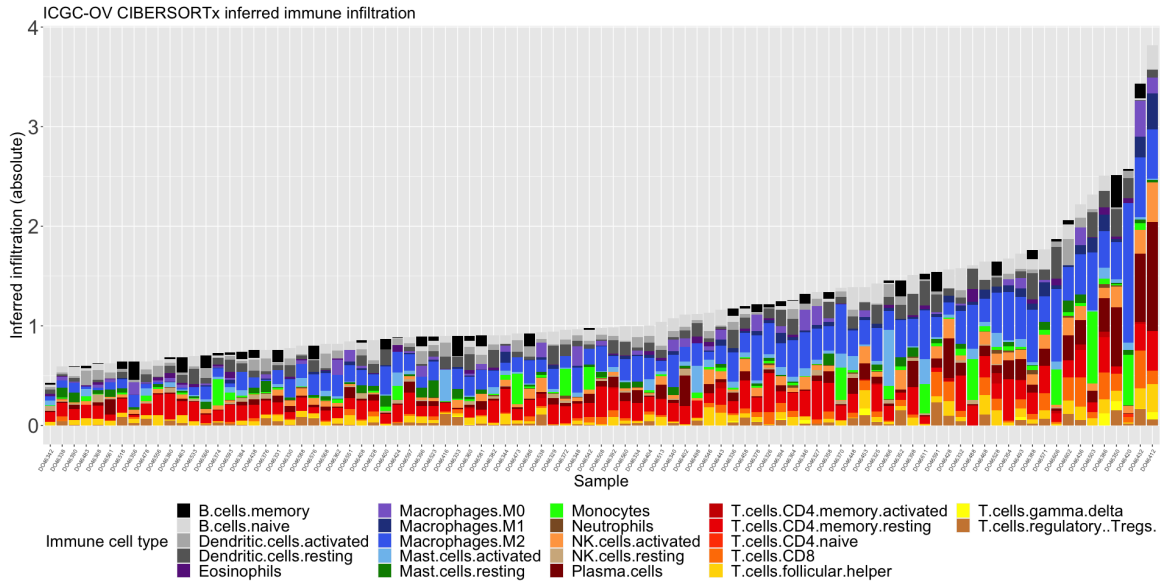


Figure 5.13 **CIBERSORTx results in the ICGC-OV cohort.** Example of CIBERSORTx results, for the ICGC-OV cohort (n=93). Bar plot visualising the results of CIBERSORTx inferred immune infiltration by cell type in the ICGC-OV cohort, sorted from smallest to largest total infiltration of all combined cell types (left to right).

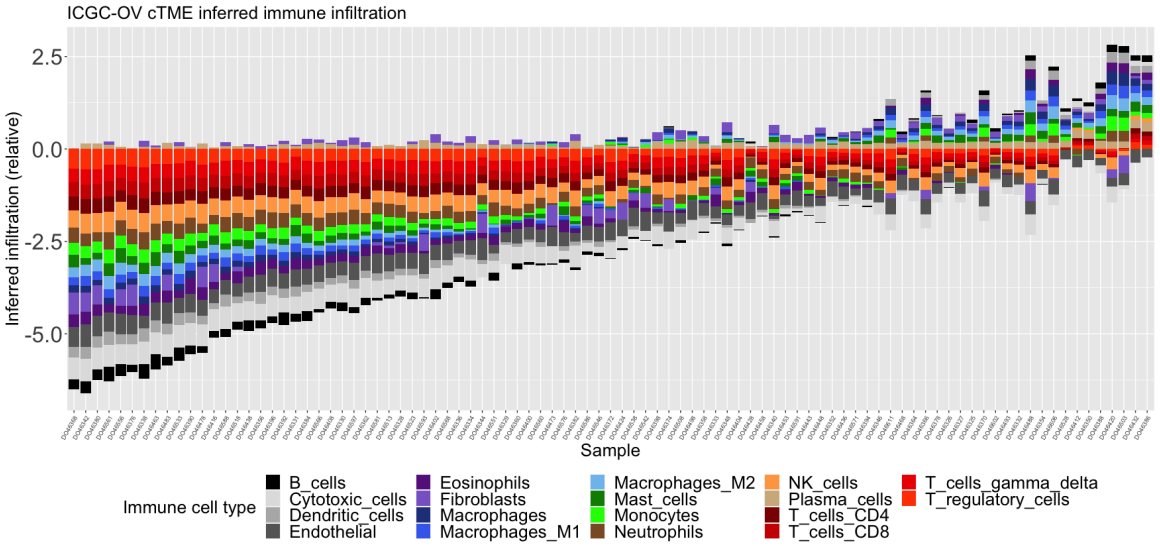


Figure 5.14 **cTME results in the ICGC-OV cohort.** Example of cTME results, for the ICGC-OV cohort (n=93). Bar plot visualising the results of cTME relative inferred immune infiltration by cell type in the ICGC-OV cohort, sorted from least to most relative infiltration of all combined cell types (left to right).

In the TCGA-OV dataset, *MYC*-amplified samples exhibited a markedly different immune infiltration profile to non-amplified samples (figures 5.15 and 5.16). Indeed, using the cTME approach, there was greater relative infiltration of every individual immune cell in the *MYC*-amplified subset as well as a higher total "immune score" (figure 5.15), a score generated by combining gene sets for each immune cell into one larger set for each TCGA cancer type. CIBERSORTx results also include an overall infiltration score, which exhibited the same result of greater immune infiltration in *MYC*-amplified TCGA-OV samples (figure 5.16). Fewer individual cell types reached significance between groups by CIBERSORTx, but M1 macrophages and T follicular helper (Tfh) cells both demonstrated increased absolute infiltration levels in the *MYC*-amplified TCGA-OV group.

In contrast to TCGA-OV data, the TCGA-LUAD dataset demonstrated no significant difference between relative or absolute levels of immune cell types based on *MYC* amplification status, excepting a higher memory B cell count in non-amplified samples by CIBERSORTx (figure 5.17). The TCGA-LUAD cohort had demonstrated a trend towards the expected effect of *MYC* amplification (figures 5.11c and 5.12), although without reaching significance: scores suggested higher cMYC-G9a/EZH2 repression levels and lower immune signalling in *MYC*-amplified samples, as observed in the GDSC2 cell line data (figure 5.5). Overall immune scores as calculated by either cTME or CIBERSORTx platforms were not significantly different between *MYC*-amplified and non-amplified samples for the TCGA-LUAD nor the TCGA-BRCA cohorts. However, the TCGA-BRCA cohort did demonstrate several differences in immune infiltration of individual cell types by both cTME and CIBERSORTx analysis (figures 5.18 and 5.19).

All significant differences in immune cells between *MYC*-amplified and non-amplified TCGA-

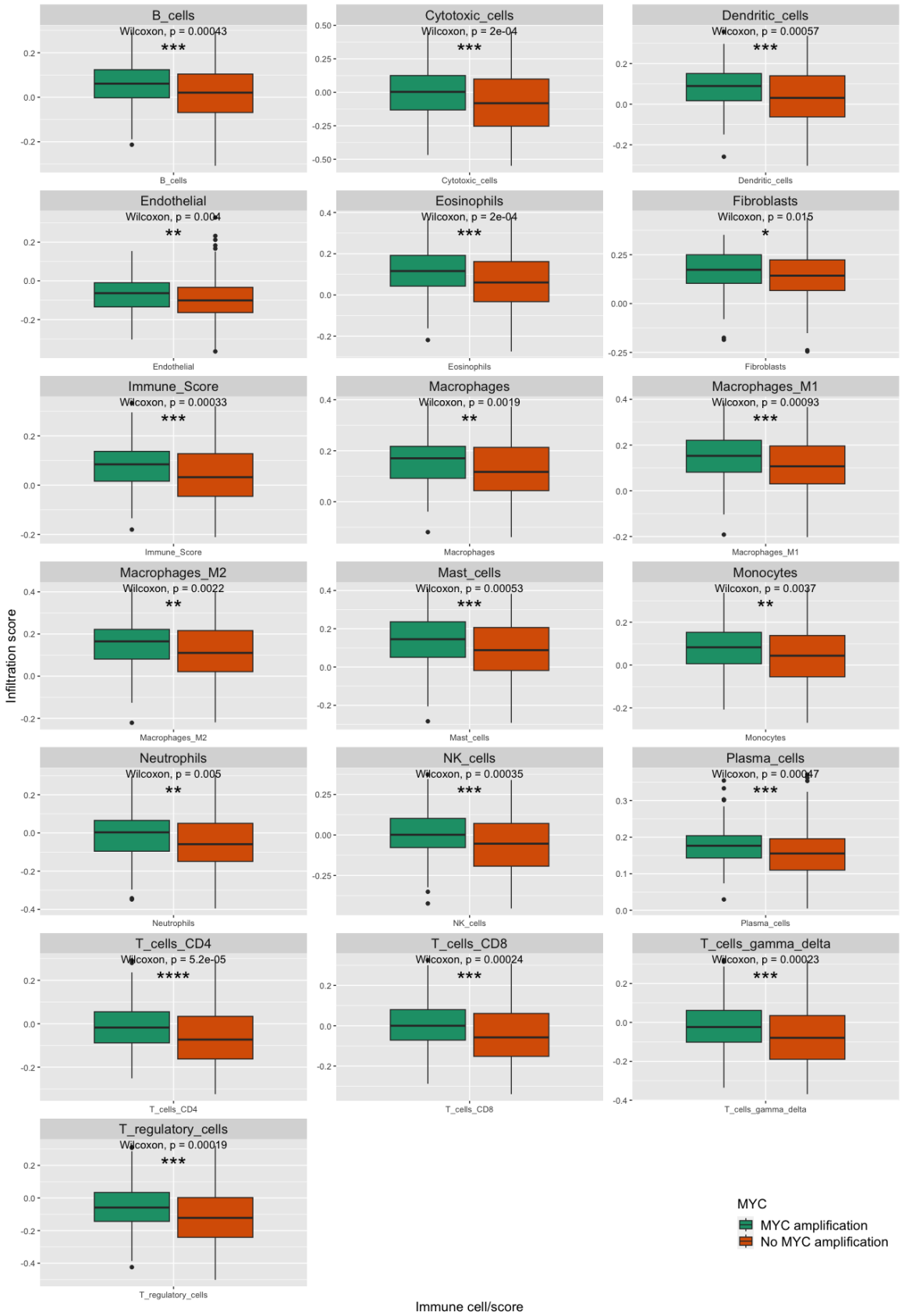


Figure 5.15 **cTME results in the TCGA-OV cohort, stratified by *MYC* status.** Comparison of cTME-inferred immune infiltration in the TCGA-OV cohort (n=422) between *MYC*-amplified and non-amplified patients. Only significant differences are plotted ($p < 0.05$). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

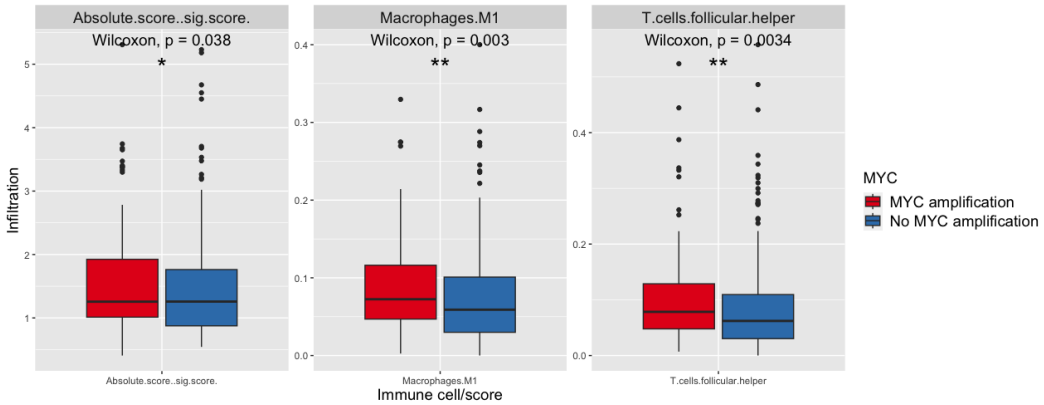


Figure 5.16 **CIBERSORTx results in the TCGA-OV cohort, stratified by *MYC* status.** Comparison of CIBERSORTx-inferred immune infiltration in the TCGA-OV cohort (n=422) between *MYC*-amplified and non-amplified patients. Only significant differences are plotted ($p < 0.05$). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

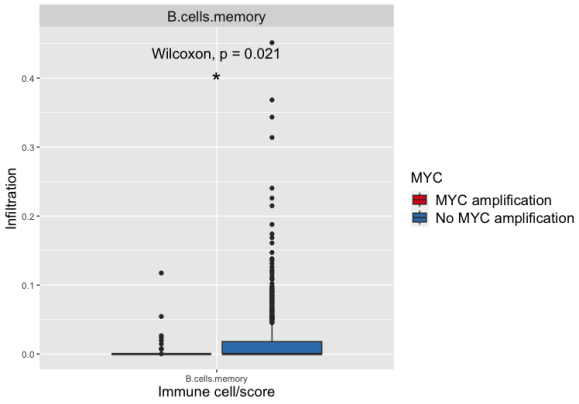


Figure 5.17 **CIBERSORTx results in the TCGA-LUAD cohort, stratified by *MYC* status.** Comparison of CIBERSORTx-inferred immune infiltration in the TCGA-LUAD cohort (n=517) between *MYC*-amplified and non-amplified patients. Only significant differences are plotted ($p < 0.05$). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

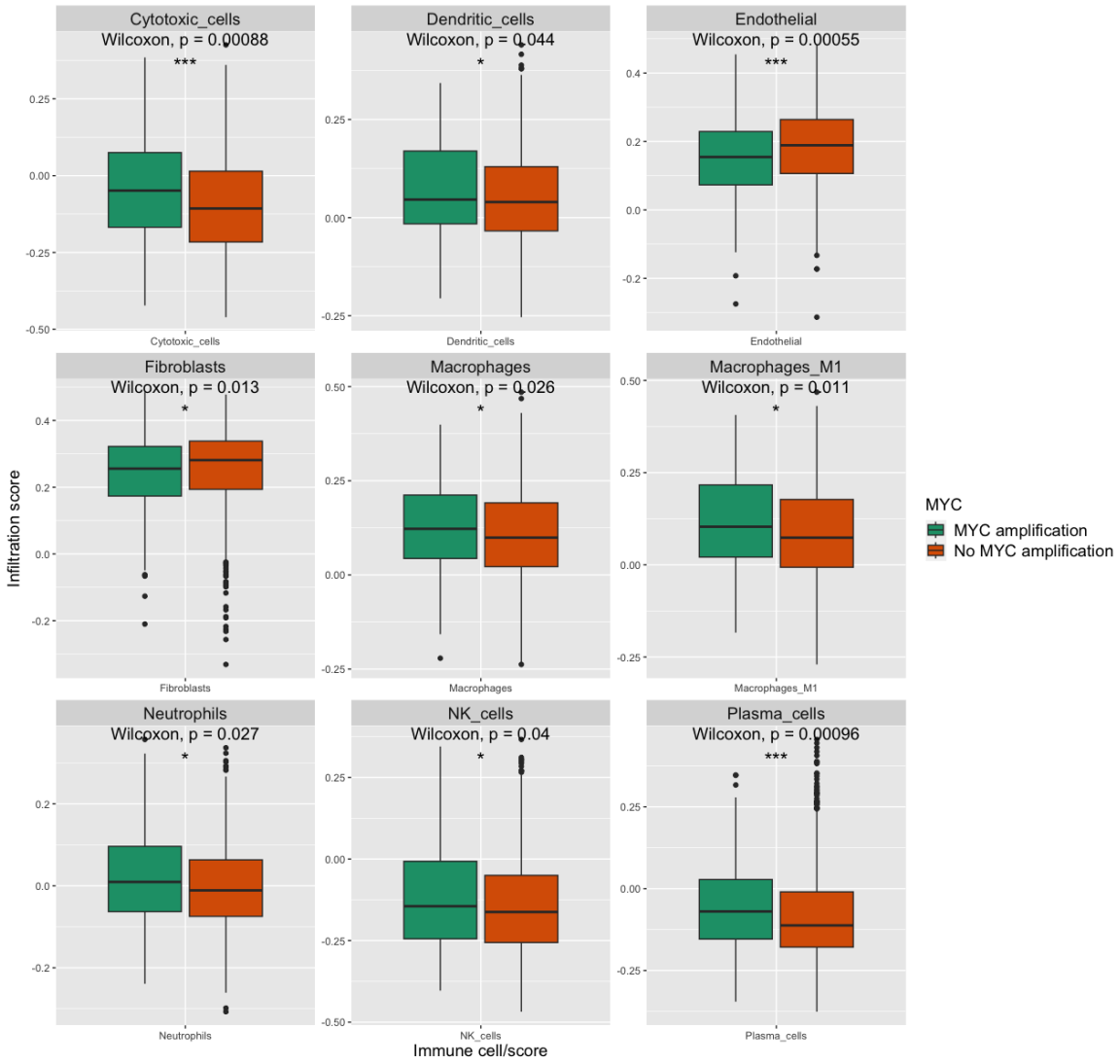


Figure 5.18 **cTME results in the TCGA-BRCA cohort, stratified by *MYC* status.** Comparison of cTME-inferred immune infiltration in the TCGA-BRCA cohort (n=1095) between *MYC*-amplified and non-amplified patients. Only significant differences are plotted ($p < 0.05$). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

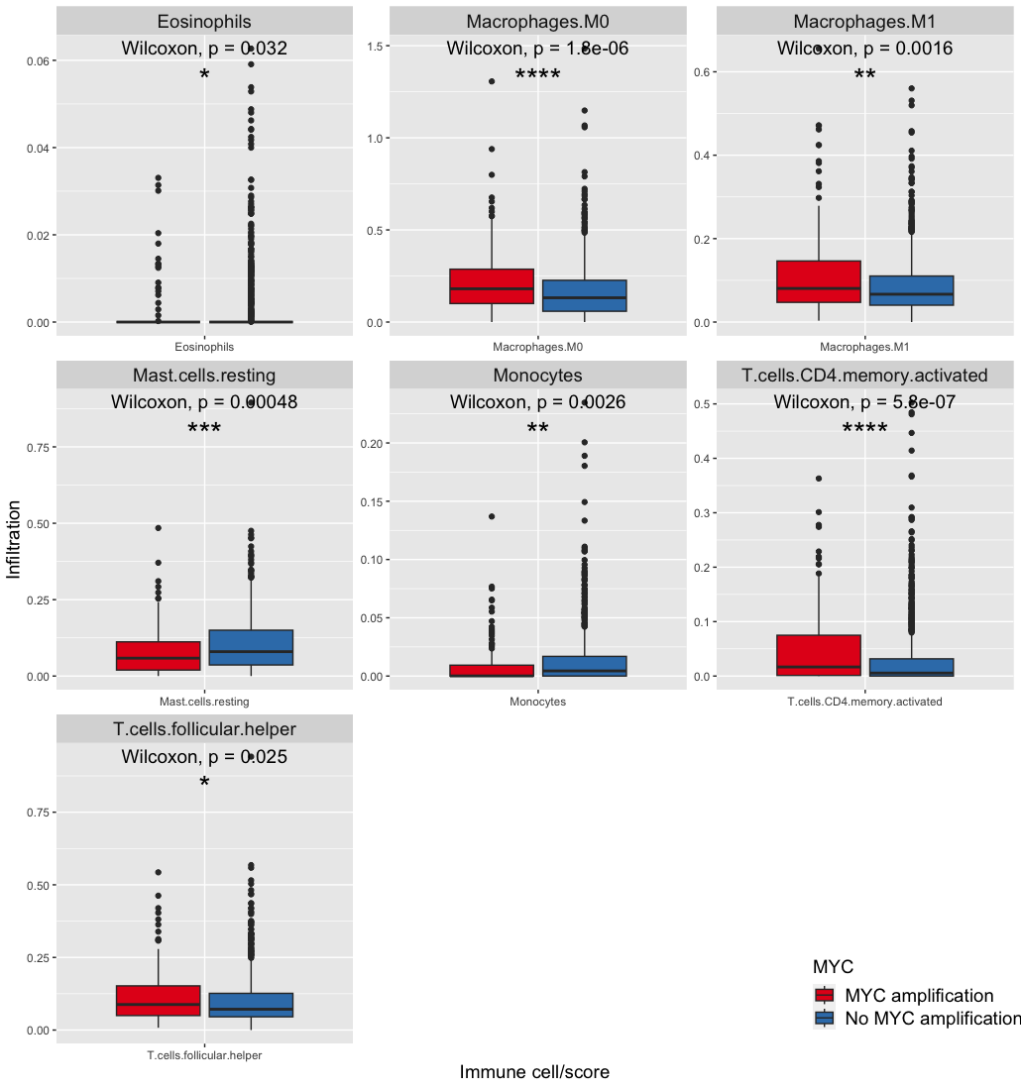


Figure 5.19 **CIBERSORTx results in the TCGA-BRCA cohort, stratified by *MYC* status.** Comparison of CIBERSORTx-inferred immune infiltration in the TCGA-BRCA cohort (n=1095) between *MYC*-amplified and non-amplified patients. Only significant differences are plotted ($p < 0.05$). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

BRCA groups by cTME analysis illustrated a higher infiltration in the *MYC*-amplified group, with a reduction only in relative levels of endothelial and fibroblast cells (figure 5.18). Higher relative infiltration was observed for cytotoxic cells, dendritic cells, macrophages, neutrophils, NK cells, plasma cells and gamma-delta T cells ($\gamma\delta$T) cells in *MYC*-amplified TCGA-BRCA patients. CIBERSORTx demonstrated differences in several T cell types and myeloid cells, again demonstrating greater infiltration in *MYC*-amplified tumours, except for eosinophils, monocytes and mast cells which had lower infiltration scores in this group (figure 5.19).

The data demonstrate that *MYC* amplification is associated with significant changes in immune infiltration in the TCGA-OV and TCGA-BRCA datasets, two cohorts that had both displayed unexpected relationships between *MYC* amplification and cMYC-G9a/EZH2 or immune scores as compared to cell line observations. Importantly, subsequent analysis of the correlation between inferred immune cell infiltration and cMYC-G9a/EZH2/immune scores suggested a strong impact of immune composition on scores: higher inferred immune composition was correlated with higher scores for both cMYC-G9a/EZH2 repression and for immune signalling. To illustrate this at a general level, the overall infiltration of immune cells inferred by CIBERSORTx was first compared to immune scores and to cMYC-G9a/EZH2 scores in the TCGA-OV and TCGA-BRCA cohorts (figures 5.20 and 5.21), showing the strong positive correlation for all genesets except canonical cMYC signalling.

The TCGA-OV cohort was next investigated more specifically. 7ISG scores had been shown to be significantly higher in the *MYC*-amplified versus non-amplified TCGA-OV samples (see figure

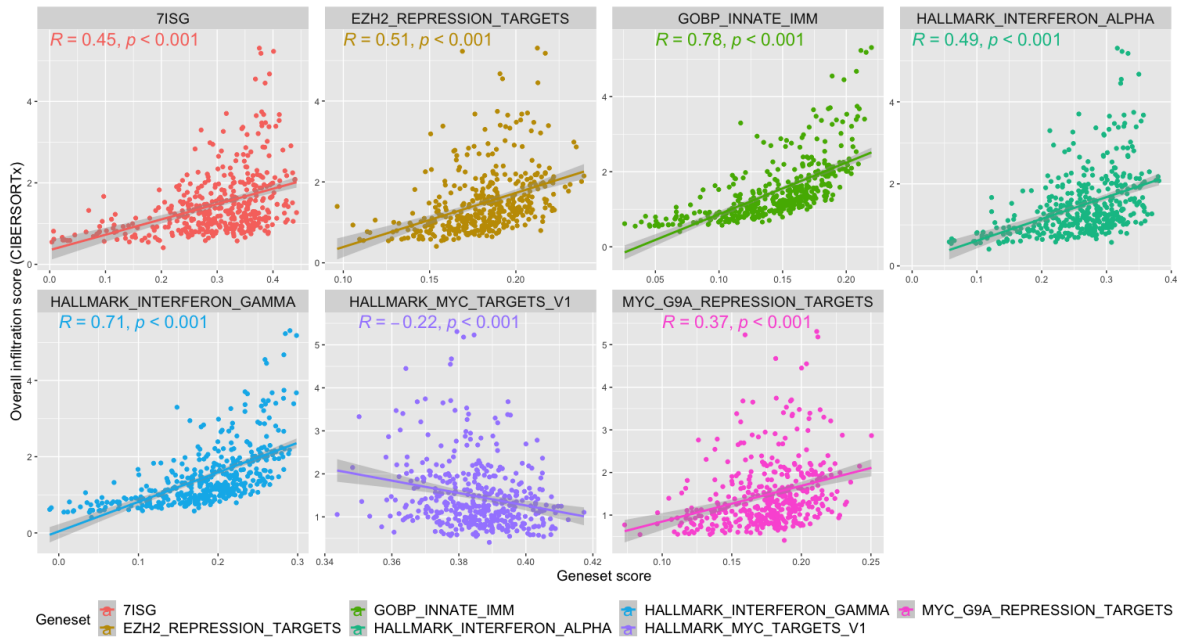


Figure 5.20 **Inferred immune infiltration in the TCGA-OV cohort: correlation with geneset scores.** Spearman's correlation between scores for overall immune infiltration (y-axis, inferred by CIBERSORTx) versus scores for cMYC-G9a/EZH2 repression, canonical cMYC signalling and immune signalling (x-axis), in the TCGA-OV cohort (n=422). Genesets listed in legend. "GOBP_INNATE_IMM" = GOBP_INNATE_IMMUNE_RESPONSE geneset.

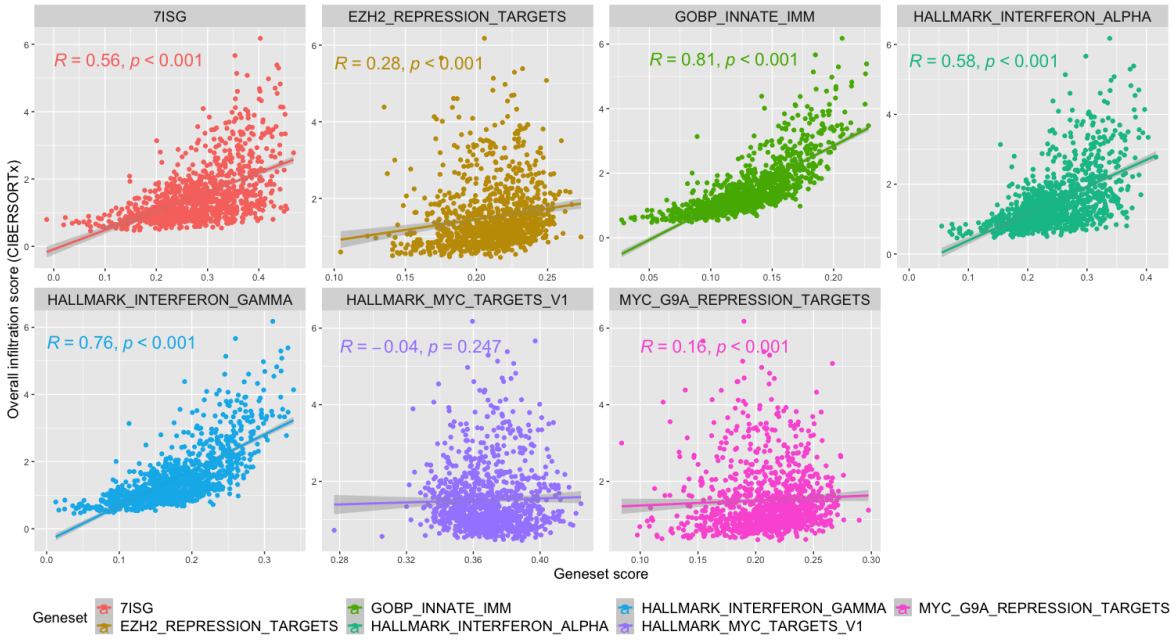
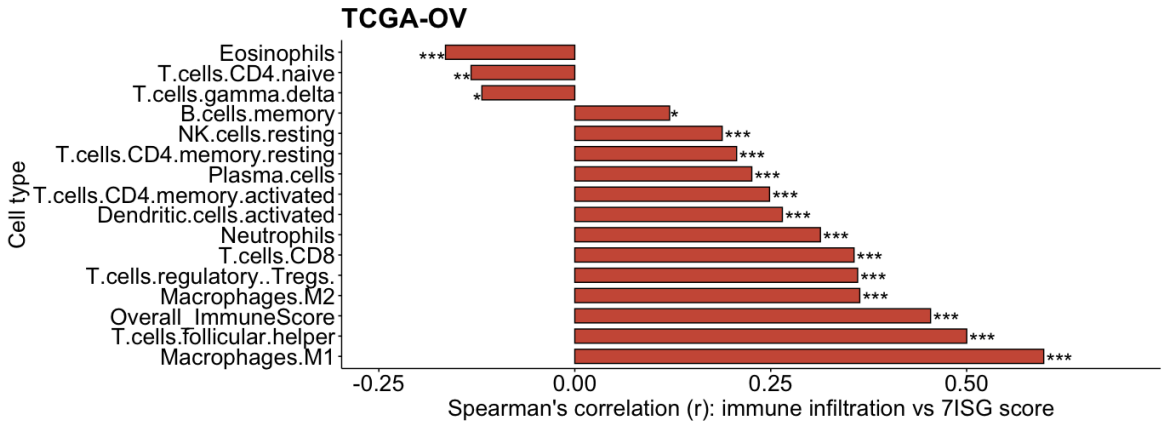


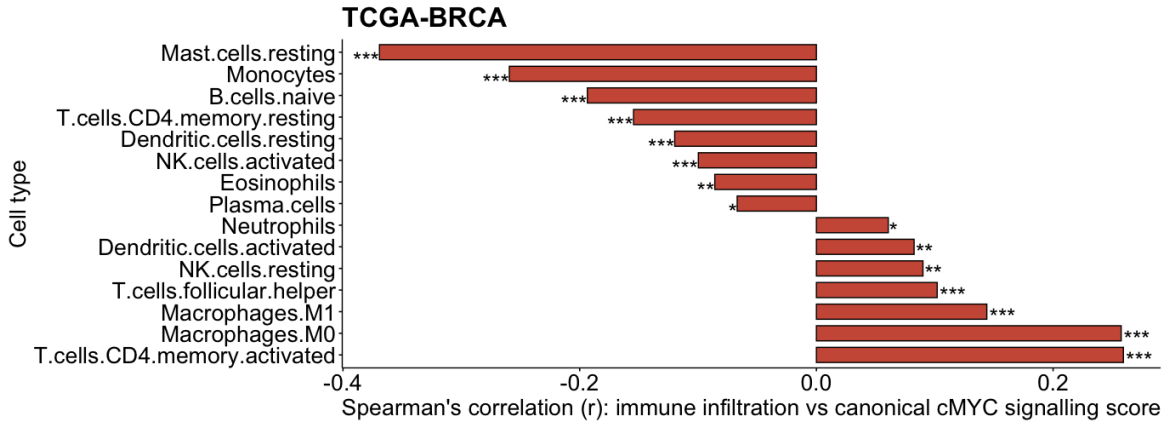
Figure 5.21 **Inferred immune infiltration in the TCGA-BRCA cohort: correlation with geneset scores.** Spearman's correlation between scores for overall immune infiltration (y-axis, inferred by CIBERSORTx) versus scores for cMYC-G9a/EZH2 repression, canonical cMYC signalling and immune signalling (x-axis), in the TCGA-BRCA cohort (n=1095). Genesets listed in legend. "GOBP_INNATE_IMM" = GOBP_INNATE_IMMUNE_RESPONSE geneset.

5.12), hence the top immune cell correlates with the 7ISG were investigated in this cohort. The top 3 immune correlates with 7ISG scores in the TCGA-OV cohort were M1 macrophages, Tfh cells and the overall immune score (figure 5.22a): these were also the three immune sets that were significantly higher in the *MYC*-amplified TCGA-OV group compared to the non-amplified group (see figure 5.16). Whilst they cannot provide conclusive evidence, these correlative data support the hypothesis that immune cells may contribute to the higher cMYC-G9a/EZH2 repression and immune scores observed in *MYC*-amplified TCGA-OV patients, where cell line data had indicated the opposite relationship.

A similar effect was observed in the TCGA-BRCA cohort, where scores for the canonical cMYC geneset had been shown to be significantly greater in *MYC*-amplified samples (see figure 5.11b). Canonical cMYC scores most strongly correlated with memory activated CD4+ T cells, M0/M1 macrophages and Tfh cells, whilst anticorrelating most strongly with monocytes and mast cells (figure 5.22b). The former cells were those that demonstrated enriched infiltration in *MYC*-amplified samples, whilst the latter were depleted (see figure 5.19). Overall, the data suggest that differences in immune composition between *MYC*-amplified and non-amplified patient samples could heavily confound comparison of scores between these two groups, suggesting cell line or single cell data are more reliable for these investigations into what is hypothesised to be a cancer cell-intrinsic effect.



(a)



(b)

Figure 5.22 **CIBERSORTx results in the TCGA-OV and TCGA-BRCA cohorts: correlation with 7ISG and canonical cMYC scores.** Barplots showing the strength of Spearman's correlation between scores for absolute infiltration of immune cell types as inferred by CIBERSORTx versus (a) 7ISG scores in the TCGA-OV cohort (n=422), or (b) canonical cMYC signalling scores in the TCGA-BRCA cohort (n=1095).

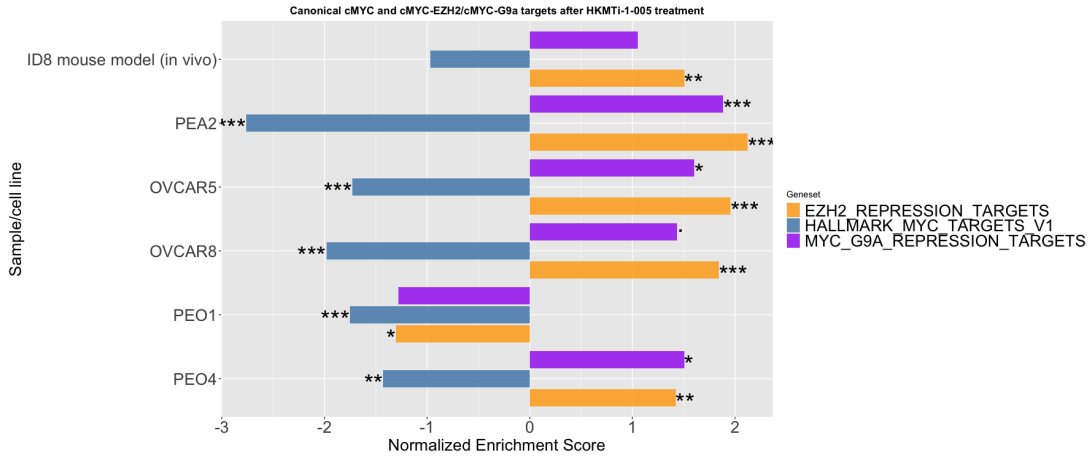
5.4 HKMTi-1-005 upregulates cMYC-G9a/EZH2 targets

5.4.1 HKMTi-1-005 treatment induces de-repression of cMYC-G9a and EZH2 targets, with downregulation of canonical cMYC signalling

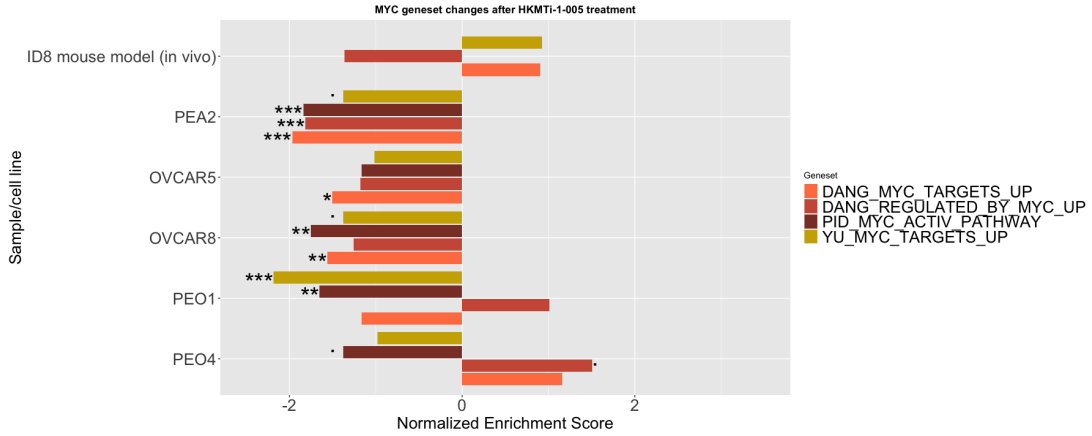
Stimulation of an immune response by HKMTi-1-005 was demonstrated in chapters 3 and 4, as was de-repression of EZH2 targets using paired genesets from MSigDB. IFNG and innate immune signalling were induced and a signature 7ISG geneset was defined to characterise the core VM response.

The HGSC cell lines and ID8 *in vivo* model were next analysed to assess treatment effect on cMYC-G9a targets and the broad EZH2 repression geneset, alongside hallmark cMYC gene targets. fGSEA performed on the PEA2, OVCAR5, OVCAR8, PEO1 and PEO4 cell lines and ID8 *in vivo* sample demonstrated that the immune stimulation described is accompanied by a consistent upregulation of cMYC-G9a- or EZH2-repressed genes, in contrast to a downregulation of hallmark cMYC signalling (figure 5.23a). Notably, the data demonstrated downregulation of the *MYC* transcript itself in only 2 of the 6 samples (appendix table A.10; one additional sample showed downregulated *MYC* with $p = 0.05$.).

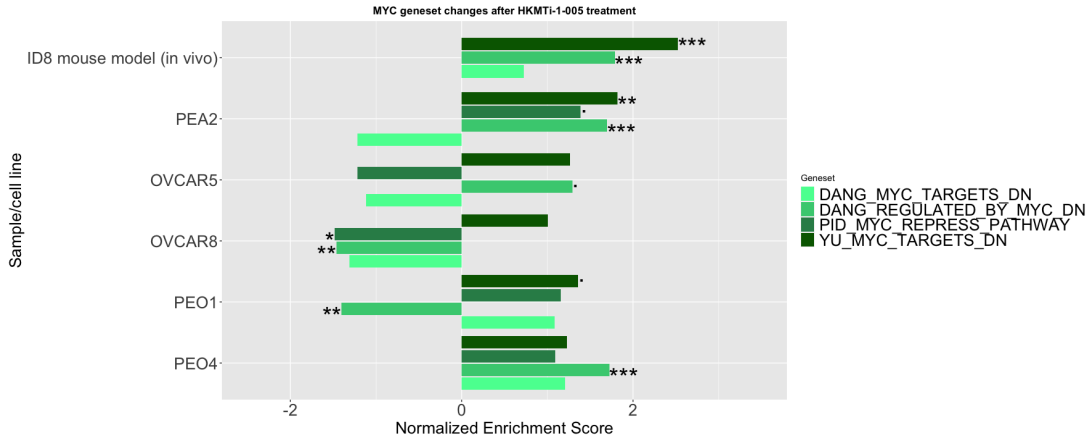
To validate the divergent effect of HKMTi-1-005 treatment on cMYC-repressed versus cMYC-activated genes, MSigDB was mined for paired genesets relating to cMYC activity. fGSEA results provided strong support for initial findings: cMYC-upregulated genes were consistently downregulated by treatment (figure 5.23b), whilst cMYC-repressed genes showed a strong trend towards upregulation (figure 5.23c).



(a)



(b)



(c)

Figure 5.23 **fGSEA results for cMYC-G9a/EZH2 and canonical cMYC genesets, HGSC samples.** fGSEA results for PEA2, OVCAR5, OVCAR8 PEO1 and PEO4 cell lines and the ID8 *in vivo* sample, following treatment with HKMTi-1-005. (a) fGSEA results for cMYC-G9a repression, EZH2 repression and canonical cMYC signalling genesets. EZH2_REPRESSION_TARGETS = EZH2 repression, HALLMARK_MYC_TARGETS_V1 = canonical cMYC signalling, MYC_G9A_REPRESSION_TARGETS = cMYC-G9a repression. (b) fGSEA results for cMYC "UP" genesets from MSigDB, comprising genes that are cMYC-upregulated. (c) fGSEA results for cMYC "DN" genesets from MSigDB, comprising genes that are cMYC-downregulated. . = p-adj < 0.1, * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001.

5.5 Sensitivity to HKMTi-1-005 relates to *MYC* status and the 7ISG

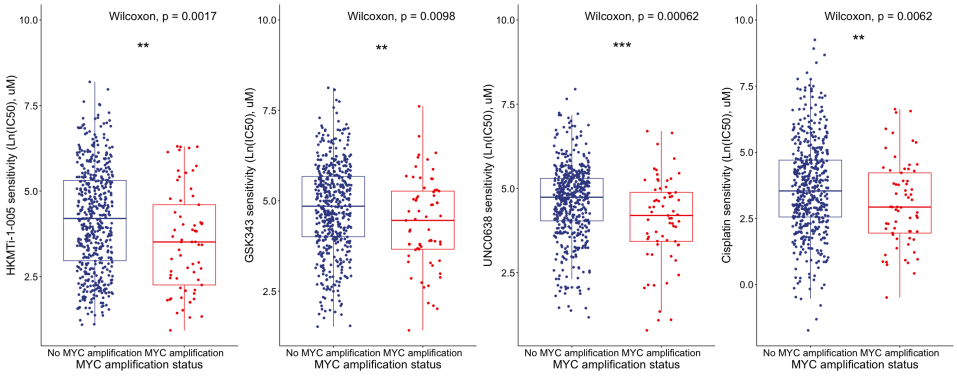
5.5.1 *MYC* deregulation is associated with increased HKMTi-1-005 sensitivity

Following analysis of HKMTi-1-005 treatment effect on cMYC target genes in HGSC models, the solid, matched GDSC2 cell line panel was investigated for associations between HKMTi-1-005 sensitivity and *MYC* deregulation. First, *MYC*-amplified lines were compared to lines without *MYC* amplification. Alongside HKMTi-1-005, sensitivity to the single EZH2i GSK343 and to the single G9ai UNC0638 were compared between these cell line groups, as well as to the first-line HGSC chemotherapy cisplatin.

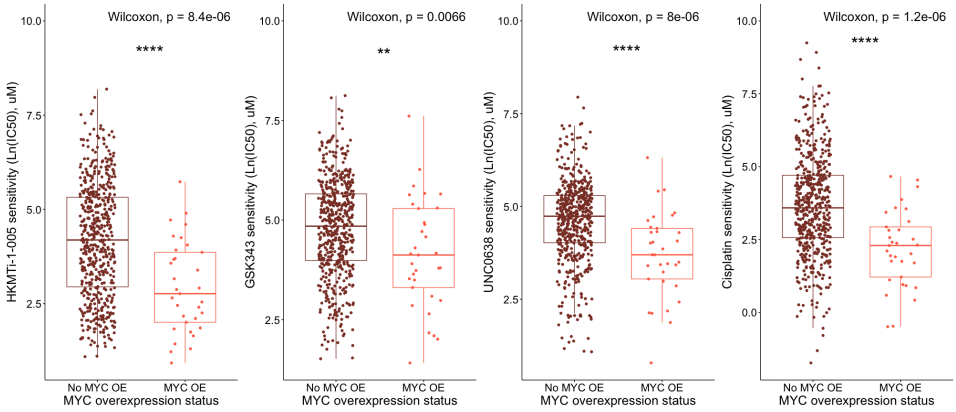
The data demonstrated a significantly increased sensitivity in *MYC*-amplified cell lines (figure 5.24a). Importantly, *MYC*-amplified lines did not have a significantly faster doubling time than non-amplified lines (appendix figure A.19), suggesting that heightened sensitivity is not simply due to greater proliferative rate. However, the sensitivity effect of *MYC* amplification was not specific to HKMTi-1-005; sensitivity to each of GSK343, UNC0638 and cisplatin was also increased in the *MYC*-amplified cohort (figure 5.24a).

Other forms of *MYC* deregulation were assessed in parallel. *MYC* overexpression was associated with greater sensitivity to all four compounds (figure 5.24b). *MYC* mutation was not significantly associated with greater sensitivity to any of the compounds, but very few of the lines were defined as having a *MYC* mutation according to the COSMIC database [407] (figure 5.24c). However, a trend towards greater HKMTi-1-005 and EZH2i sensitivity was still evident in *MYC*-mutated lines.

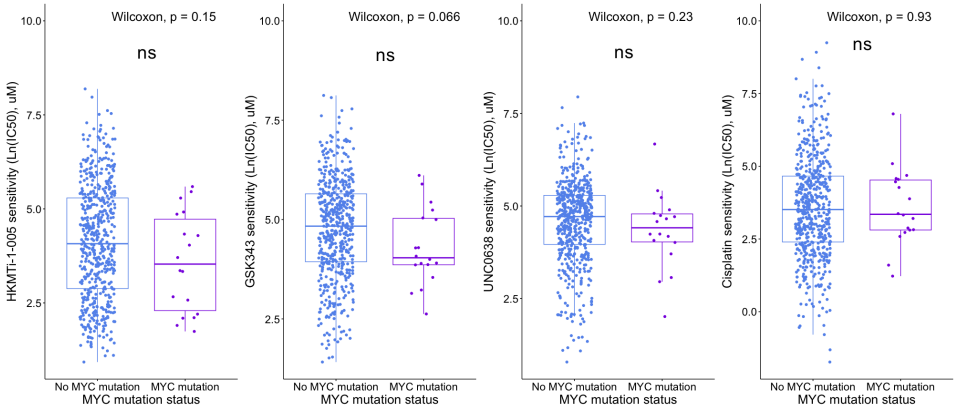
The data support increased HKMTi-1-005 sensitivity of *MYC*-deregulated cell lines, yet demonstrate that this effect is not specific to the compound over single EZH2i/G9ai or cisplatin. *MYC* amplification is used as a proxy for *MYC* deregulation going forward; *MYC* amplification is more clinically relevant than overexpression and a very significant overlap between the two alterations exists in the cell line panel (appendix figure A.20), with the small number of *MYC*-mutated lines limiting robust analyses using this subgroup.



(a)



(b)



(c)

Figure 5.24 ***MYC* deregulation versus sensitivity in the GDSC2 cell line panel.** Relationship between (a) *MYC* amplification, (b) *MYC* overexpression or (c) *MYC* mutation and sensitivity to HKMTi-1-005, GSK343 (EZH2i), UNC0638 (G9ai) and cisplatin (left to right). Cell lines included are those from the GDSC2 panel with sensitivity data for each of the four compounds, derived from any solid tumour type (n=567 for amplification, n = 565 for OE, n = 564 for mutation). N=65 with *MYC* amplification, n=502 without. N=33 with *MYC* OE, n=532 without. N=18 with *MYC* mutation, n=546 without. Cell lines with multiple different mutation statuses were excluded (n=3), with mutations defined using the COSMIC database [407]. . * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$. $Ln(IC_{50})$ = natural log of half-maximal inhibitory concentration.

5.5.2 Low basal 7ISG scores correlate with greater HKMTi-1-005 sensitivity, independent of canonical cMYC activity

I next looked to determine whether basal immune status was associated with sensitivity to HKMTi-1-005. Particularly, I investigated whether the 7ISG immune signature might relate to compound sensitivity, and whether this effect was *MYC*-dependent.

A panel of compounds with different modalities was included alongside HKMTi-1-005 to assess the specificity of any observations. In addition to the single EZH2i (GSK343) and single G9ai (UNC0638), the compounds were selected to reflect drugs used in HGSC treatment (cisplatin, paclitaxel, olaparib (PARPi)), additional classic chemotherapy agents (gemcitabine, methotrexate, topotecan) and epigenetic drugs (5-azacytidine (DNMTi), GSK-LSD1-2HCl (LSD1i), vorinostat (HDACi)). Solid cell lines from the GDSC2 panel were matched to ensure only those with sensitivity to each of the twelve compounds were included (n=515). Scores for the immune genesets described – the 7ISG, IFNA, IFNG and innate immune response signalling sets – were correlated with drug sensitivity, alongside scores for canonical cMYC signalling.

These data identified a significant correlation between lower immune scores and greater sensitivity to HKMTi-1-005 for each immune geneset (figure 5.25). Of note, the 7ISG correlations with sensitivity were indistinguishable from the larger immune genesets (see appendix table A.11 for CIs), suggesting the reduced signature geneset retains most of the sensitivity information that exists in the larger genesets. Immune geneset correlations contrasted with those of canonical cMYC signalling, where higher scores were correlated with greater sensitivity.

Whilst these data implied a relationship between immune signalling and HKMTi-1-005 sensitivity, the relationship was not specific to HKMTi-1-005; 8 of the 11 comparison compounds demonstrated a significant relationship between immune signalling and sensitivity (figure 5.25). Every one of the 11 compounds also demonstrated a significant anticorrelation between canonical cMYC signalling and IC_{50} values, suggesting that higher cMYC signalling generally increases sensitivity to treatment.

The effect of *MYC* deregulation on these relationships was next considered. The relationships between immune/canonical cMYC scores and sensitivity were assessed in *MYC*-amplified lines separately from those without an amplification. These data demonstrated a marked visual difference between the two cohorts: the correlations between sensitivity and immune scores were consistently higher in *MYC*-amplified lines, although a significant difference was not reached (figure 5.26, see appendix table A.12 for CIs). Additionally, in *MYC*-amplified lines the relationship appeared more specific to HKMTi-1-005: only 3 of the 11 comparison compounds consistently demonstrated significant correlations between immune scores and sensitivity, namely UNC0638, olaparib and vorinostat.

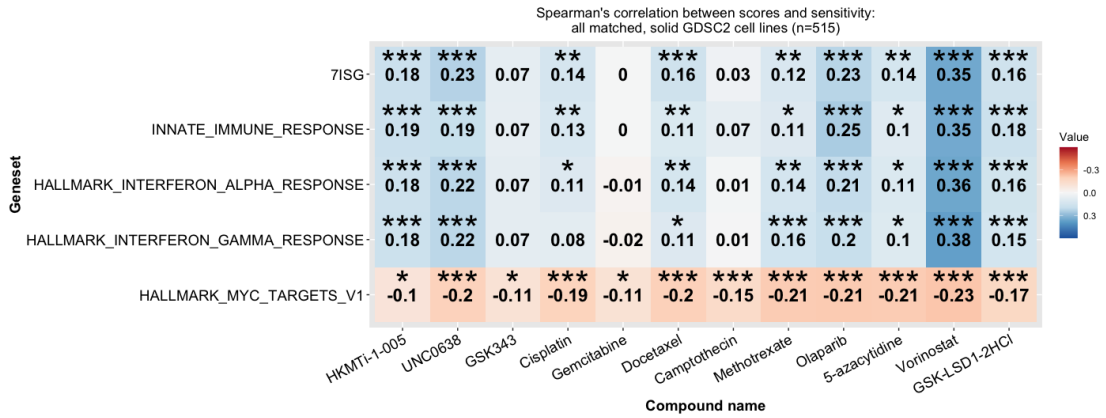


Figure 5.25 **Immune and cMYC scores versus sensitivity in the GDSC2 cell line panel.** Spearman's correlation between sensitivity to a panel of 12 drugs including HKMTi-1-005 (x-axis, as measured by IC_{50} values) versus scores for genesets (y-axis). Genesets include the 7ISG (derived in chapter 4), innate immune response (INNATE_IMMUNE_RESPONSE), IFNA signalling (HALLMARK_INTERFERON_ALPHA_RESPONSE), IFNG signalling (HALLMARK_INTERFERON_GAMMA_RESPONSE) and canonical cMYC signalling (HALLMARK_MYC_TARGETS_V1). Correlation across cell lines in the GDSC2 panel with sensitivity data for each of the 12 compounds was assessed, including cell lines of any solid tumour type with gene expression data (n=515). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

These data demonstrate that a low 7ISG score predicts sensitivity to HKMTi-1-005 as well as to several other compounds, an effect that appears heightened in lines with *MYC* amplification. However, canonical cMYC signalling scores also correlated with sensitivity across the panel of drugs; canonical cMYC scores significantly anticorrelated with the 7ISG in both *MYC*-amplified and non-amplified lines, potentially confounding the 7ISG sensitivity relationships (figure 5.27). Hence, I looked to investigate the individual contribution of the 7ISG score versus the canonical cMYC score to sensitivity correlations.

To investigate this potential confounding effect, a multiple linear regression approach was used. In each regression model, the 7ISG and hallmark cMYC scores acted as independent variables, whilst drug sensitivity was the dependent variable. Alongside HKMTi-1-005, the three compounds with the strongest correlation between sensitivity and the 7ISG were chosen for comparison: UNC0638, olaparib and vorinostat. Cell lines were matched to include those that had sensitivity data for each of these compounds as well as for HKMTi-1-005, resulting in a panel of 563 cell lines (n=65 *MYC*-amplified, n=498 non-amplified).

These investigations illustrated that sensitivity to HKMTi-1-005 was significantly associated with the 7ISG score, independent from canonical cMYC score associations (figure 5.28). This was true in both *MYC*-amplified and non-amplified cohorts, whilst canonical cMYC signalling was not independently associated with sensitivity in either cohort. Interestingly, the same relationships were not

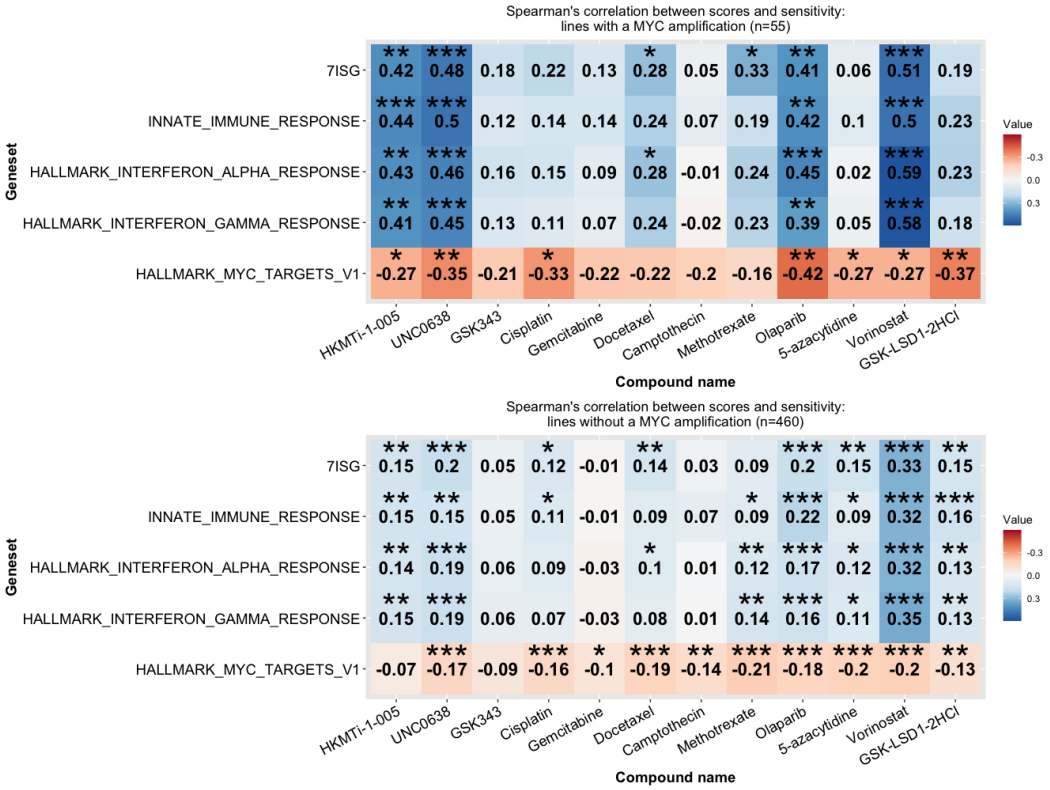


Figure 5.26 **Immune and cMYC scores versus sensitivity in the GDSC2 cell line panel: stratified by *MYC* status.** Spearman's correlation between sensitivity to a panel of 12 drugs including HKMTi-1-005 (x-axes, as measured by IC_{50} values) versus scores for genesets (y-axes) in a) *MYC*-amplified cell lines (n=55) and b) lines without *MYC* amplification (n=460). Genesets include the 7ISG (derived in chapter 4), innate immune response (INNATE_IMMUNE_RESPONSE), IFNA signalling (HALLMARK_INTERFERON_ALPHA_RESPONSE), IFNG signalling (HALLMARK_INTERFERON_GAMMA_RESPONSE) and canonical cMYC signalling (HALLMARK_MYC_TARGETS_V1). Cell lines of any solid tumour type in the GDSC2 panel with with gene expression data and with sensitivity data for each of the 12 compounds were included (see text for details on compound selection) (n=515). * = p < 0.05, ** = p < 0.01, *** = p < 0.001.

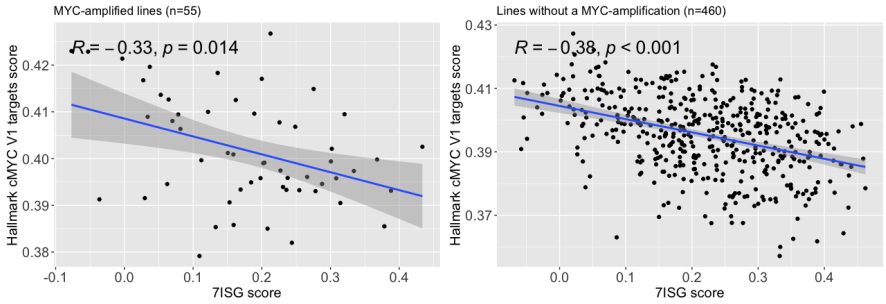
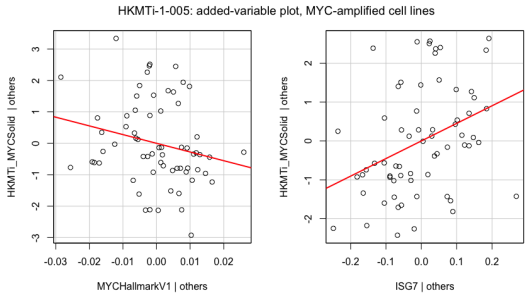


Figure 5.27 **Correlation between canonical cMYC scores and 7ISG scores in the GDSC2 panel, stratified by *MYC* status.** Spearman's correlation between 7ISG scores and canonical cMYC signalling scores (Hallmark cMYC V1 targets score = HALLMARK_MYC_TARGETS_V1), in *MYC*-amplified lines (left, n=55) and lines without *MYC* amplification (right, n=460). R = Spearman's r, p = p-value. Cell lines of any solid tumour type in the GDSC2 panel with with gene expression data and with sensitivity data for each of the 12 compounds were included (see text for details on compound selection) (n=515).

observed for any of the other three compounds examined that had demonstrated a strong relationship between the 7ISG and sensitivity: whilst 7ISG scores were still independently associated with sensitivity, canonical cMYC signalling was an independent predictor of sensitivity for all three of UNC0638, olaparib and vorinostat compounds (figure 5.29, appendix figures A.21 and A.22). Both *MYC*-amplified and non-amplified cohorts demonstrated these relationships, with the single exception of vorinostat sensitivity in the *MYC*-amplified cohort, where canonical cMYC signalling did not independently associate with sensitivity (figure A.22).

Taken together, the data support the hypothesis that the 7ISG score – but not canonical cMYC signalling – acts as an independent predictor of sensitivity to HKMTi-1-005, and suggest that this effect may be strongest in *MYC*-amplified lines. Low basal immune signalling including the 7ISG is associated with higher sensitivity to several other compounds when cell lines of any *MYC* status are considered. However, the data also demonstrate a high degree of specificity to the 7ISG-sensitivity relationships when *MYC*-amplified cell lines are analysed separately. Importantly, whilst the 7ISG is independently associated with sensitivity to several other compounds, the lack of an independent association with canonical cMYC signalling is unique to HKMTi-1-005 amongst compounds assessed.

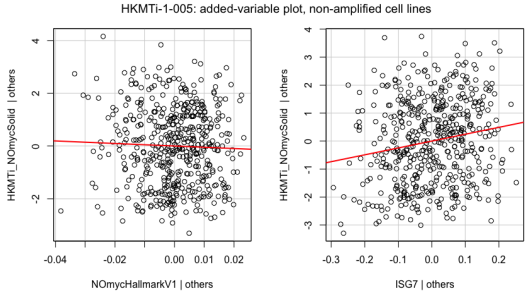


(a)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.932      6.358   2.191   0.0322 *
MYCHallmarkV1 -27.695     15.531  -1.783   0.0795 .
ISG7            4.524      1.490   3.037   0.0035 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.324 on 62 degrees of freedom
Multiple R-squared:  0.2442,    Adjusted R-squared:  0.2199
F-statistic: 10.02 on 2 and 62 DF,  p-value: 0.0001698
```

(b)



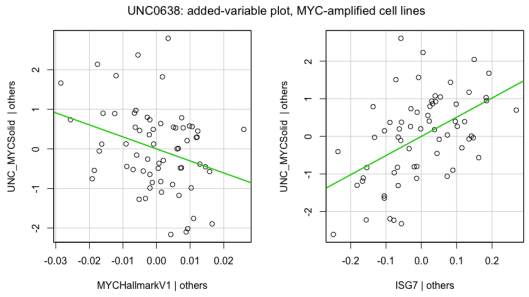
(c)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.5394     2.4904   2.224 0.026580 *
MYCHallmarkV1 -4.7915     6.1414  -0.780 0.435648
ISG7           2.4541     0.6272   3.913 0.000104 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.471 on 495 degrees of freedom
Multiple R-squared:  0.04324,   Adjusted R-squared:  0.03937
F-statistic: 11.18 on 2 and 495 DF,  p-value: 1.775e-05
```

(d)

Figure 5.28 **Multiple linear regression: HKMTi-1-005.** Multiple linear regression model examining HKMTi-1-005 sensitivity. (a) Added-variable plot and (b) summary of the results of the linear model in *MYC*-amplified cell lines (n=65). (c) Added-variable plot and (d) summary of the results of the linear model in non-amplified cell lines (n=498). Cell lines included are solid lines from the GDSC2 panel with sensitivity data for the four compounds assessed (HKMTi-1-005, UNC0638, olaparib and vorinostat, see text for details). MYCHallmarkV1=canonical cMYC signalling score (score for the HALLMARK_MYC_ TARGETS _V1 geneset). ISG7=score for the 7ISG geneset. Note added-variable plots are also called partial-regression plots.

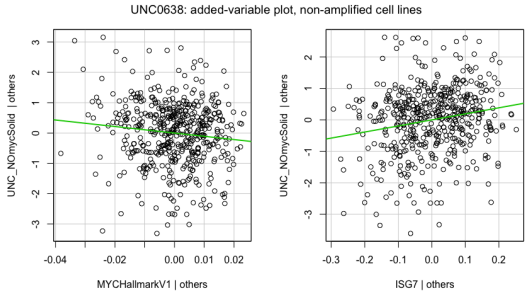


(a)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    15.462      4.863   3.180   0.0023 **
MYCHallmarkV1 -30.399     11.877  -2.559   0.0129 *
ISG7            5.099      1.139   4.475 3.33e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.013 on 62 degrees of freedom
Multiple R-squared:  0.4084,    Adjusted R-squared:  0.3893
F-statistic:  21.4 on 2 and 62 DF,  p-value: 8.582e-08
```

(b)



(c)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.4683     1.7730   4.776 2.36e-06 ***
MYCHallmarkV1 -10.7696     4.3722  -2.463   0.0141 *
ISG7            1.9277     0.4465   4.317 1.91e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.047 on 495 degrees of freedom
Multiple R-squared:  0.07594,   Adjusted R-squared:  0.07221
F-statistic: 20.34 on 2 and 495 DF,  p-value: 3.24e-09
```

(d)

Figure 5.29 **Multiple linear regression: UNC0638.** Multiple linear regression model examining UNC0638 sensitivity. (a) Added-variable plot and (b) summary of the results of the linear model in *MYC*-amplified cell lines (n=65). (c) Added-variable plot and (d) summary of the results of the linear model in non-amplified cell lines (n=498). Cell lines included are solid lines from the GDSC2 panel with sensitivity data for the four compounds assessed (HKMTi-1-005, UNC0638, olaparib and vorinostat, see text for details). MYCHallmarkV1=canonical cMYC signalling score (score for the HALLMARK_MYC_ TARGETS _V1 geneset). ISG7=score for the 7ISG geneset. Note added-variable plots are also called partial-regression plots.

5.6 Assessing prognostic significance of the 7ISG in *MYC*-deregulated tumours

5.6.1 A high 7ISG score is associated with longer survival in *MYC*-amplified HGSC patients

The data presented so far support an inverse relationship between cMYC-G9a/EZH2 repression and immune signalling. They suggest that cell-intrinsic immune repression is greater in *MYC*-amplified cell lines, and support HKMTi-1-005-mediated reversal of cMYC-G9a/EZH2 immune repression in HGSC models. Previous data demonstrated that HKMTi-1-005 immune stimulation can be characterised by the 7ISG response signature: a subset of ISGs that is strongly and consistently induced by treatment of HGSC cell lines (see section 4.2.1). To assess whether there is evidence that expression of the 7ISG may provide benefit *in vivo*, and any *MYC*-dependency of this effect, I investigated whether the 7ISG was associated with prognosis in the patient cohorts described. Especially, I looked to determine whether the prognostic value of the 7ISG was dependent on *MYC* amplification status, given the hypothesis that *MYC*-deregulated cells have a lower threshold for VM and a greater dependency on cell-intrinsic cMYC-G9a/EZH2 repression of immune genes.

The ICGC cohort was used for a preliminary analysis of associations between the 7ISG and overall survival (OS). Patients were not stratified by *MYC* status given the small sample size (n=93). Following testing of the Cox proportional hazards assumptions (appendix figure A.23a), a Cox proportional hazards (CoxPH) model was generated [413]. Age at diagnosis was included as a variable in the CoxPH model given its known impact on HGSC prognosis [337], with stratification at 50 years old as used elsewhere [412]. All patients were FIGO stage III or IV and no data were available for platinum resistance status or residual disease following cytoreduction, hence these features were not included as variables in the model.

The CoxPH model revealed that higher 7ISG scores were associated with significantly longer OS in both uni- and multi-variable analyses (table 5.1). Age was not significantly associated with survival in either analysis. A Kaplan-Meier curve comparing high versus low 7ISG score patients demonstrated a clear separation between the two groups (figure 5.30).

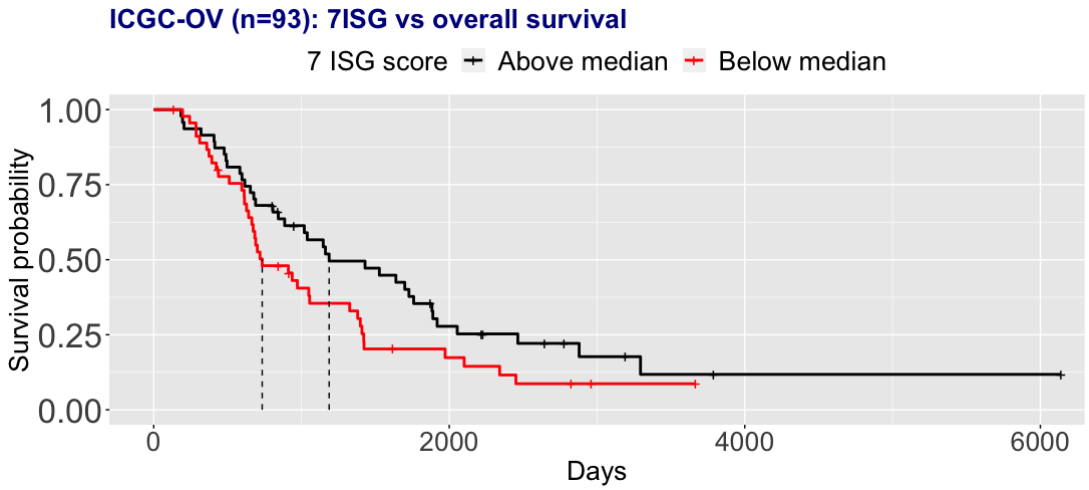


Figure 5.30 **7ISG scores versus overall survival in the ICGC-OV cohort.** Kaplan-Meier curve illustrating the overall survival difference between ICGC-OV patients with a high (above median) versus a low (below median) 7ISG score. Note survival measured in days.

Feature		**Overall survival**			
		Univariable		Multivariable	
		HR	P-value	HR	P-value
7ISG score		0.08	**0.051**	0.05	**0.033**
Age	>50				
	≤50	0.94	0.835	0.75	0.397

Table 5.1 **CoxPH model in the ICGC-OV cohort.** CoxPH uni- and multi-variable analyses results assessing the relationship between the 7ISG score and age versus overall survival in the ICGC-OV cohort (n=93). Significant (p <= 0.05, rounded to one significant figure) values indicated in bold.

Subsequently, the TCGA cohorts were used to validate findings and assess the *MYC*-dependency of the prognostic 7ISG effect. The TCGA datasets include disease-specific survival (DSS) alongside OS; DSS was used as the most relevant survival metric going forward, defined as the time between initial diagnosis and death specifically from disease. For each cohort, patients were annotated by *MYC* amplification status. Too few *MYC*-amplified patients existed for use in survival analyses for the TCGA-BRCA (n=68 samples with DSS, with only 6 events) and TCGA-LUAD cohorts (n=29 samples with DSS, with only 8 events). However, 101 samples with DSS data were retained in the *MYC*-amplified TCGA-OV cohort following filtering, with a total of 52 events. Hence, the association between 7ISG scores and survival was assessed in *MYC*-amplified versus non-amplified TCGA-OV patients using a CoxPH model, following assessment of the model assumptions (appendix figures A.23b and A.23c).

These data indicated that in *MYC*-amplified patients, but not non-amplified patients, the 7ISG was associated with improved DSS (figures 5.31 and 5.32, tables 5.2 and 5.3). In both uni- and multivariable analyses of the *MYC*-amplified cohort, a higher 7ISG significantly associated with prolonged DSS (figure 5.31, table 5.2). This relationship did not reach significance in OS data, but still showed a strong trend in the same direction (p = 0.069) (appendix figure A.24, table 5.2). In non-amplified samples, there was no separation between patients with a high 7ISG and those with a low 7ISG, and no significant relationship of the 7ISG with OS or DSS (figures 5.32 and appendix figure A.25, table 5.3). Taken together, the data suggest prognostic value of the 7ISG in *MYC*-amplified HGSC patients, but not those without *MYC* amplification.

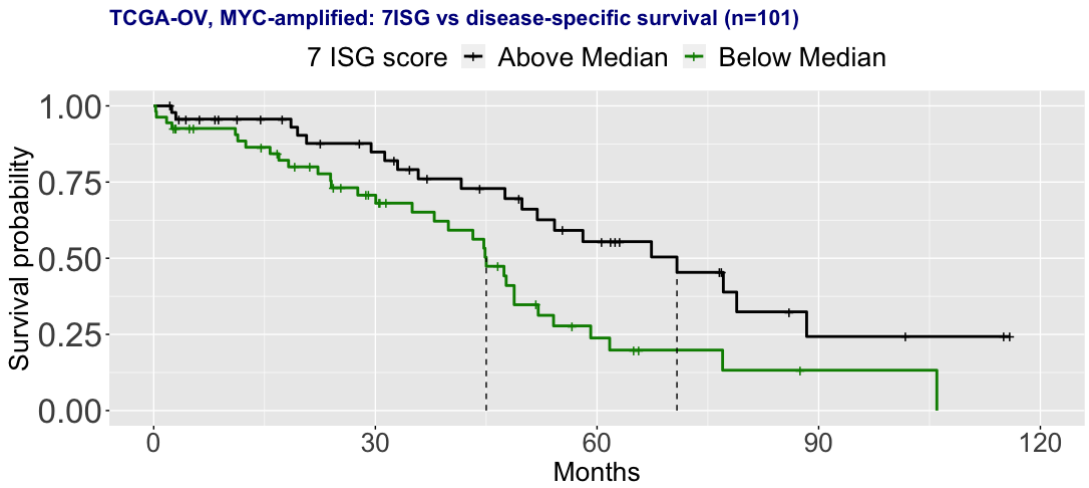


Figure 5.31 **7ISG scores versus disease-specific survival in the *MYC*-amplified TCGA-OV cohort.** Kaplan-Meier curve comparing disease-specific survival between those with a high (above median) versus a low (below median) 7ISG score in the *MYC*-amplified TCGA-OV patient cohort. Note survival measured in months.

Feature		**Disease-specific survival**				**Overall survival**			
		Univariable		Multivariable		Univariable		Multivariable	
		HR	P-value	HR	P-value	HR	P-value	HR	P-value
7ISG score		0.00	**0.012**	0.00	**0.012**	0.02	0.069	0.02	0.069
Age	>50								
	≤50	0.91	0.767	0.91	0.759	0.84	0.552	0.84	0.54

Table 5.2 **CoxPH model in the *MYC*-amplified TCGA-OV cohort.** CoxPH uni- and multivariable analyses results assessing the relationship between the 7ISG score and age versus disease-specific or overall survival in the *MYC*-amplified TCGA-OV cohort. Significant (p <= 0.05) values indicated in bold.

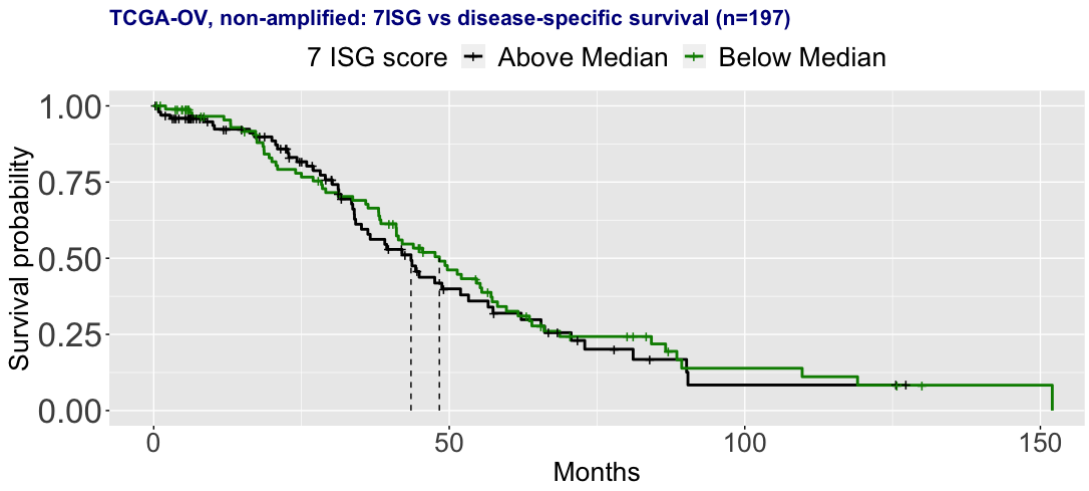


Figure 5.32 **7ISG scores versus disease-specific survival in the non-amplified TCGA-OV cohort.** Kaplan-Meier curve comparing disease-specific survival between those with a high (above median) versus a low (below median) 7ISG score in the non-amplified TCGA-OV cohort. Note survival measured in months.

Feature		**Disease-specific survival**				**Overall survival**			
		Univariable		Multivariable		Univariable		Multivariable	
		HR	P-value	HR	P-value	HR	P-value	HR	P-value
7ISG score		1.18	0.869	1.19	0.862	1.44	0.694	1.39	0.724
Age	>50								
	≤50	0.72	0.217	0.72	0.217	0.69	0.141	0.7	0.143

Table 5.3 **CoxPH model in the non-amplified TCGA-OV cohort.** CoxPH uni- and multivariable analyses results assessing the relationship between the 7ISG score and age versus disease-specific or overall survival in the non-amplified TCGA-OV cohort.

5.7 Examining effects of HKMTi-1-005 treatment in another *MYC*-deregulated tumour model

5.7.1 HKMTi-1-005 treatment induces immune signalling with growth inhibition in lung cancer cell lines

HKMTi-1-005 treatment is accompanied by cMYC-G9a/EZH2 de-repression and immune stimulation in the ID8 *in vivo* model of HGSC, with a survival benefit [155]. The data presented in this chapter support an inverse relationship between cMYC-G9a/EZH2 repression and immune signalling in another *MYC*-deregulated tumour type, LuAd (see 5.2.3 and 5.3.1). Hence, HKMTi-1-005 treatment of LuAd models was investigated to assess anti-tumour effects. Experiments in LuAd models were performed by Dr Sarah Laing and Professor Daniel J Murphy at the CRUK Scotland Institute.

First, the A549 and H358 lung carcinoma cell lines were used to investigate HKMTi-1-005 treatment effect on proliferation. Cell lines were treated with 1, 5, 10 or 20µM HKMTi-1-005 or vehicle H_2O. H358 cells have been shown to be sensitive to MEK inhibition, which targets the RAS/RAF/MEK/ERK pathway [455], hence the MEK inhibitor Trametinib was included as a positive control for growth inhibition of H358 cells at a concentration of 10nM. In contrast, A549 cells are described as resistant to MEK inhibition [455]. Therefore, a higher concentration of 100nM was used in the A549 cell line assay to assess any growth inhibitory effects. Incucyte analysis indicated a significant ($p < 0.001$) growth inhibitory effect on A549 at 10 or 20µM HKMTi-1-005 and on H358 at 5, 10 or 20µM HKMTi-1-005 (figures 5.33a and 5.33b). As anticipated, significant growth inhibition was achieved with 10nM of Trametinib in the H358 cells (figure 5.33b, $p < 0.001$), whilst no growth inhibitory effects of Trametinib treatment were evident in the A549 cells, even at the much higher concentration of 100nM (figure 5.33a).

The effect of HKMTi-1-005 treatment on expression of a subset of ISGs was subsequently assessed in the A549 cell lines. KRAS and cMYC are demonstrated to repress IFN responses in pancreatic cancer [274], an effect that has been demonstrated in other models including LuAd [363]. Further, Trametinib has been shown to induce *IRF7* and *IRF9* expression in the LuAd KMA mouse model [449]. Hence, Trametinib was used as a positive control for ISG induction in the A549 cell line. qRT-PCR demonstrated significant transcriptional upregulation of the key ISGs *IRF7*, *IRF9* and *STAT2* following treatment with either HKMTi-1-005 or Trametinib (figure 5.33c; Trametinib run failed for *STAT2*). These *in vitro* analyses suggest HKMTi-1-005 inhibits proliferation and induces a cell-intrinsic immune response in lung carcinoma cell lines.

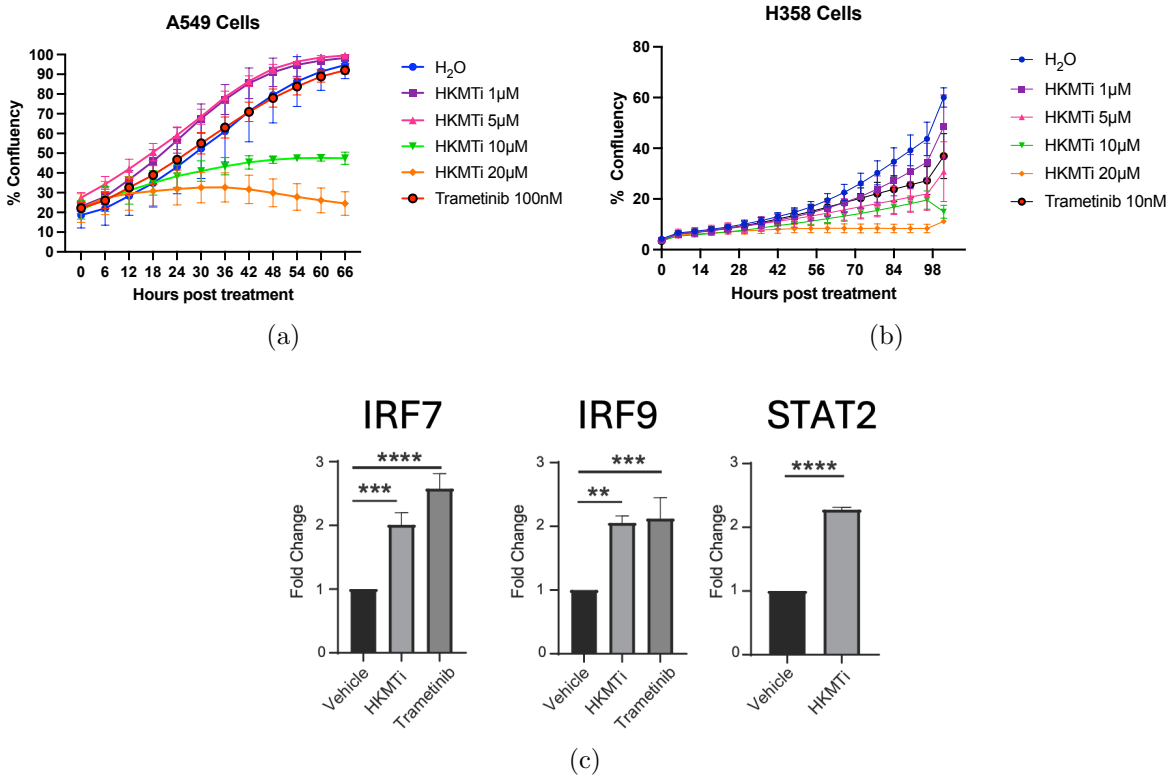


Figure 5.33 **Analysis of HKMTi-1-005 treatment effect on lung adenocarcinoma cell lines.** Incucyte analysis of (a) A549 and (b) H358 cell line proliferation, following treatment with either HKMTi-1-005 at the indicated doses, H_2O vehicle, or the MEK inhibitor Trametinib at 100nM (A549) or 10nM (H358). Error bars show standard deviation (SD) of 3 biological replicates, performed on separate days. $p<0.001$ for A549 treated with 10 or 20μM HKMTi-1-005. $p<0.001$ for H358 treated with 5, 10 or 20μM HKMTi-1-005 (2-way ANOVA). (c) qRT-PCR of *IRF7*, *IRF9* and *STAT2* ISGs in A549 cell lines treated with vehicle, HKMTi-1-005 or Trametinib. Mean and SD shown for n=3 technical replicates, 2 biological repeats. ** = p <0.01, *** = p < 0.001, **** = p <0.0001.

5.7.2 HKMTi-1-005 treatment increases survival of a murine *MYC*-driven LuAd model

Following *in vitro* investigations, the KM^2 (genotype: *LSL-**K**Ras*^{G12D}*;Rosa26*$^{DM.lsl-MYC/lsl-MYC}$) mouse model of LuAd was used to assess HKMTi-1-005 treatment effect on survival. Lung tumours were initiated, and after four weeks of incubation mice were randomly assigned to treatment or vehicle groups, with a treatment schedule of 3 days a week, for 4 weeks. Encouragingly, mice treated with HKMTi-1-005 demonstrated significantly longer survival following tumour initiation (figure 5.34).

A separate cohort of KM^2 mice were treated with the same dosing schedule but with harvest of lungs immediately after the last treatment. These mice were analysed for tumour burden and proliferation effects. Tumour burden was significantly reduced by HKMTi-1-005 as compared to vehicle treatment (figures 5.35a and 5.35b), and Ki67+ staining showed significantly reduced proliferation accompanying this reduction (figures 5.35c and 5.35d). Together, the data provide preliminary evidence that HKMTi-1-005 may be effective for immune stimulation with anti-tumour efficacy and improved survival in LuAd, a *MYC*-deregulated tumour type.

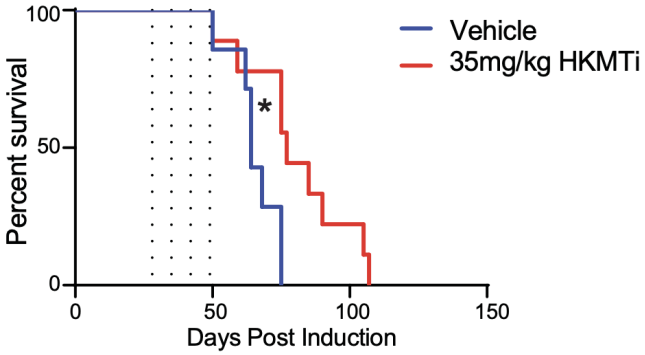


Figure 5.34 **Analysis of HKMTi-1-005 treatment effect on a LuAd model: survival effects.** (a) Kaplan-Meier curve showing overall survival of KM^2 mice treated with vehicle (n=7) or HKMTi-1-005 at 35 mg/kg, twice daily (n=9). Dotted lines indicate the first day of each cycle of three days' treatment. Mantel-Cox log rank test used to compare survival between cohorts. * = p <0.05.

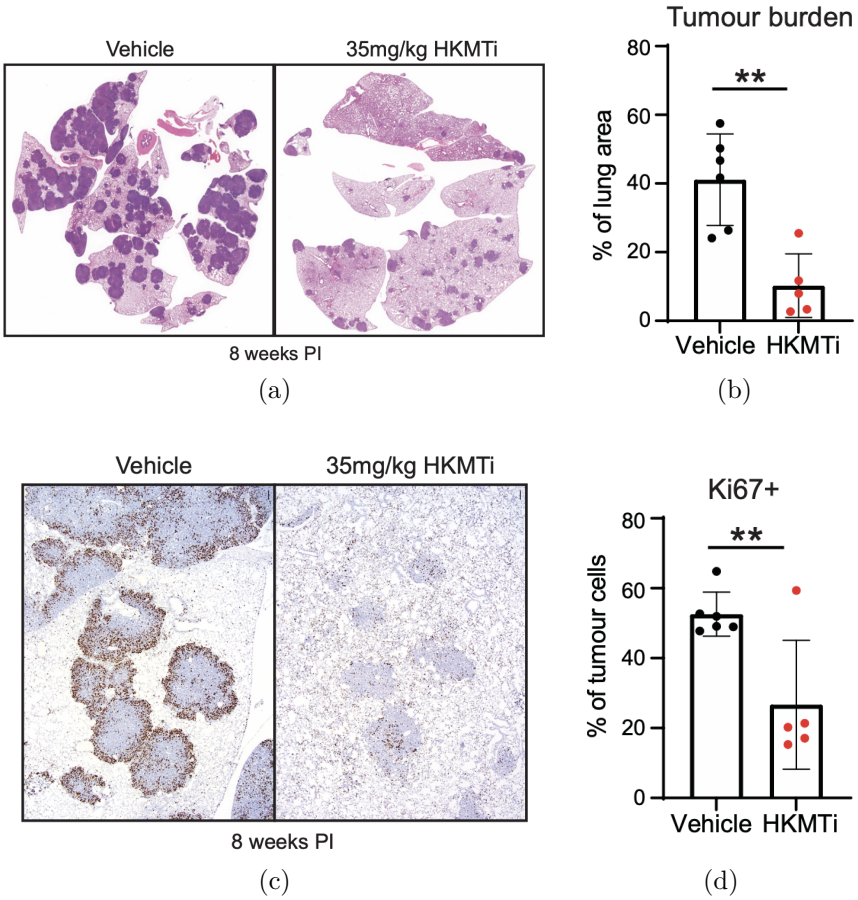


Figure 5.35 **Analysis of HKMTi-1-005 treatment effect on a LuAd model: tumour burden and proliferation.** (a) Representative images of tumour burden histology (H&E stain) of KM^2 mice, comparing mice treated with vehicle or HKMTi-1-005 at 40 mg/kg twice daily, culled following the last treatment cycle. (b) Comparison of overall tumour burden calculated as a percentage of total lung area, between KM^2 mice treated with vehicle (n=6) or HKMTi-1-005 at 40 mg/kg twice daily (n=5), culled following the last treatment cycle. Mean ± Standard Error of the Mean (SEM) shown (unpaired t-test). (c) Representative images of Ki67+ stained tissue sections from KM^2 mice treated with vehicle (n=6) or HKMTi-1-005 at 40 mg/kg twice daily (n=5), culled following the last treatment cycle. (d) Comparison of Ki67+ percentage of tumour cells in vehicle or HKMTi-1-005 treatment conditions, as measured using HALO automated quantification. Each data point represents the average score of all tumours in a section from an individual mouse. Mean ± SEM shown (unpaired t-test). ** = p <0.01.

5.8 Discussion: conclusions, limitations and future work

The data presented in this chapter support a cell-intrinsic relationship between greater cMYC-G9a/EZH2 repression and lower levels of immune signalling, with heightened repression in *MYC*-amplified cell lines. Further, they support HKMTi-1-005 treatment as a potential intervention against this repressive axis, particularly in *MYC*-amplified cancers with low basal immune signalling.

Cell line data hosted by the GDSC2 were first investigated (see 5.2). This included scoring for immune genesets as well as genesets reflecting targets of cMYC-G9a repression, of EZH2 repression and of canonical (hallmark) cMYC activation. Strong positive correlations between cMYC-G9a/EZH2 scores and each of the immune signalling scores were observed, supporting a cell-intrinsic cMYC-G9a/EZH2 immune repression effect. In parallel, a strong anticorrelation existed between canonical cMYC scores and immune signalling, suggesting greater cMYC-mediated transcriptional activation is associated with reduced cell-intrinsic immune signalling. The 7ISG was included amongst the immune genesets and exhibited indistinguishable relationships with cMYC-G9a/EZH2 repression and canonical cMYC scores compared to the larger IFNA, IFNG and innate immune genesets. Analysis of ovarian, breast and LuAd cell lines separately corroborated the wider GDSC2 analysis, suggesting the relationships observed exist across multiple *MYC*-deregulated tumour types and are not confounded by a single tumour type.

To assess these relationships *in vivo*, bulk tumour data from the ICGC and TCGA databases were also examined (see 5.3). Ovarian, breast and LuAd cohorts all revealed similarly strong relationships between lower immune signalling and higher levels of cMYC-G9a/EZH2 repression, particularly for IFNG and innate immune genesets. The relationship between canonical cMYC signalling scores and immune signalling was less consistent across *in vivo* cohorts, but was largely not significant or corroborated the negative *in vitro* correlation. A weak positive correlation was observed between canonical cMYC signalling and several immune genesets in the breast cancer cohort, however in all cases IFNG and innate immune signalling demonstrated a significantly stronger relationship with cMYC-G9a/EZH2 repression.

The effect of *MYC* amplification on the repression-immune axis was also investigated *in vitro*: a stronger relationship between repression scores and immune scores was observed in *MYC*-amplified cell lines compared to non-amplified lines (see 5.2). This led to investigation into the effect of *MYC* deregulation on basal cMYC-G9a/EZH2 repression scores, immune scores or canonical cMYC signalling scores. As anticipated, *MYC* amplification – associated with higher expression of *MYC* RNA and cMYC protein [456] – was associated with greater canonical cMYC signalling. Additionally, *MYC*-amplified cell lines had scores suggesting more cMYC-G9a/EZH2 repression and less immune signalling, including lower 7ISG scores. The results implied a heightened cMYC-G9a/EZH2 immune

repression effect in *MYC*-amplified cell lines.

However, these findings were not verified in the *in vivo* cohorts: in contrast to the GDSC2 cell line results, *MYC* amplification was not consistently associated with a greater strength of correlation between immune scores and cMYC-G9a/EZH2 repression (see 5.3). Additionally, *MYC*-amplified patient samples did not consistently demonstrate greater levels of cMYC-G9a/EZH2 repression, less immune signalling or more canonical cMYC signalling. Whilst the majority of breast and LuAd comparisons demonstrated no significant *MYC* amplification effect on these scores, the ovarian cohort suggested *MYC* amplification was accompanied by greater immune signalling and less cMYC-G9a/EZH2 repression, in contradiction to *in vitro* observations.

Several hypotheses may explain these observations. Of course, there is the possibility that *MYC* amplification does not equate to greater cMYC-G9a/EZH2 repression of immune genes or greater canonical cMYC signalling. Additionally, the stratification of TCGA patients may not be sufficiently accurate to capture *MYC* effects; ploidy is not considered in *MYC* amplification calls using cBioPortal annotations [121–123], which can lead to false negatives and positives. *MYC* can also be deregulated in numerous ways other than amplification, with evidence that oncogenic cMYC activity in HGSC is downstream of deregulation at each of the DNA, protein and RNA levels [457].

Significant literature reveals that cMYC heavily influences the immune composition of the TME [279, 451, 452]. Given the bulk nature of the patient samples, one highly plausible hypothesis is that the contribution of non-tumour cell gene expression masks the cell-intrinsic effect of *MYC* amplification on scores. Particularly, the fact that cohort-level correlation analyses were consistent with *in vitro* observations supports this possibility, as these comparisons would be less sensitive to changes in cellularity than comparison of absolute scores between *MYC*-amplified and non-amplified samples, or comparison of correlation strengths.

To examine evidence for this confounding effect, the CIBERSORTx and cTME platforms were used to infer immune infiltration in the *in vivo* cohorts (see 5.3.3). CIBERSORTx results provide an estimate for absolute immune infiltration [454], whilst cTME estimates are most suited for comparison of relative cell scores across samples [411]. Importantly, there is considerable debate in the field over which immune infiltration platforms are most accurate, with different platforms performing best across different benchmarking experiments [411, 453]. cTME was chosen as it attempts to overcome platform-specificity by compiling genesets from many different approaches and benchmarking results against the TCGA database [411], whilst CIBERSORTx allows intra- and inter-sample comparison, is widely-used and performs well across several benchmarking analyses [458, 459].

The most striking result from immune analyses was the marked increase in relative infiltration of all immune cell types in *MYC*-amplified ovarian samples compared to non-amplified samples by cTME analysis. The overall inferred immune infiltration was higher in this cohort by both

cTME and CIBERSORTx, with significantly higher absolute infiltration of two individual cell types (M1 macrophages and Tfh cells) by CIBERSORTx. Moreover, these were the two cell types that most strongly correlated with the 7ISG immune scores that were higher in *MYC*-amplified samples, suggesting that they could contribute to the higher scores in *MYC*-amplified samples where cell-intrinsic data had demonstrated the opposite effect.

Supporting this, no differences in relative immune cell abundances were apparent in the LuAd cohort and only a single cell type (memory B cells) was significantly different in absolute inferred infiltration, with higher infiltration in non-amplified samples. In the LuAd cohort, scores showed the expected trend towards more cMYC-G9a/EZH2 repression and less immune signalling in *MYC*-amplified samples, with stronger correlations between the scores in this amplified group. In the breast cohort, only canonical cMYC signalling and IFNG response scores were higher in the *MYC*-amplified samples. Canonical cMYC signalling scores were highly significantly correlated with the immune cells that showed greater infiltration in this group: M0 and M1 macrophages, CD4+ memory activated T cells and Tfh cells. Notably, the two cell types that were negatively correlated with canonical cMYC signalling – monocytes and resting mast cells – showed significantly lower infiltration in *MYC*-amplified samples. In both the TCGA-OV and TCGA-BRCA cohorts, the overall immune infiltration highly significantly correlated with all cMYC-G9a/EZH2 repression and immune scores.

In aggregate, the data demonstrate that *MYC*-amplified TCGA-OV and TCGA-BRCA cohorts exhibit considerable differences in immune cell infiltration versus non-amplified groups. Additionally, they demonstrate that cMYC-G9a/EZH2 repression and immune scores are highly related to immune composition. Many possible reasons for the differential immune infiltration in *MYC*-amplified versus non-amplified tumours exist. For example, cMYC activity in tumour cells may directly influence immune infiltration as demonstrated by multiple groups [279, 451, 452, 460]. Alternatively, indirect relationships between tumour cell *MYC* status and immune infiltration may exist; there may be a different composition of tumour subtypes between the *MYC* status groups, and therefore different tumour microenvironments. Hence, the data suggest that comparison between *MYC*-amplified and non-amplified scores in a bulk tumour setting may be confounded by non-tumour cell gene expression. Single-cell RNAseq would enable differentiation of tumour versus non-tumour gene expression, to validate the cell-intrinsic observation that *MYC* amplification is associated with increased cMYC-G9a/EZH2 repression of immune genes.

Having investigated cMYC-G9a/EZH2 relationships with immune signalling, I looked to determine whether the immune stimulation induced by HKMTi-1-005 treatment was accompanied by a reversal in cMYC-G9a/EZH2 repression or any effect on canonical cMYC signalling (see 5.4). fGSEA was used to assess enrichment or depletion of relevant genesets [404]; consistently, HKMTi-1-005 treatment induced enrichment of cMYC-G9a/EZH2-repressed genes, with depletion of canonical

cMYC target genes. Given the broad range of cMYC target genes, fGSEA analyses were extended to assess the effect of treatment on four paired cMYC genesets from the MSigDB [416], including genes shown to be up- or down-regulated by cMYC. Supporting distinct effects of HKMTi-1-005 at genes transactivated by cMYC versus transrepressed genes, a strong trend towards induction of cMYC-repressed genes and downregulation of canonical cMYC-activated genes was observed across cell lines.

The consistent effect of HKMTi-1-005 treatment on cMYC-activated genes in HGSC cell lines was of particular interest given the literature suggesting that EZH2 inhibitors are unable to target cMYC-EZH2 effects [250, 292, 300, 301]. The HKMTi-1-005 effect at these target genes calls to mind results from chapter 3: cell-free assays did not support direct enzymatic inhibition of EZH2, but pointed at alternative targeting effects at EZH2, such as a potential depletion effect on EZH2 or a disruption of the EZH2-G9a interaction. Additionally, recent literature has demonstrated that in the *MYC*-amplified cell line HL-60 (an acute myeloid leukaemia line), HKMTi-1-005 but not GSK343 is able to reduce cell viability and induce differentiation of cells [437], suggesting a potential efficacy of HKMTi-1-005 in *MYC*-deregulated lines that is not achieved with enzymatic EZH2 inhibitors. The current data cannot definitively answer how HKMTi-1-005 affects canonical cMYC targets. However, given the moderate effect at the *MYC* RNA level (see appendix table A.10), future investigations into the HKMTi-1-005 mechanism of action could include assessment of cMYC at the protein level by western blot to determine whether protein stability is altered by treatment. Additionally, ChIPseq experiments to determine whether cMYC chromatin binding is altered following treatment would be valuable in unravelling *MYC*-dependent HKMTi-1-005 effects.

A notable limitation in analyses up to this point is the limited existence of genesets describing cMYC-G9a or cMYC-EZH2 targets; only a single cMYC-G9a geneset has been defined to date [268] and was derived in breast cancer cells, whilst no cMYC-EZH2 genesets were discovered in the literature. This limitation prevents distinction of EZH2-only repression targets versus those of the cMYC-EZH2 complex, whilst the small number of cMYC-G9a targets included in the geneset are likely limited by cell line-specificity. Additionally, it is important to recognise the correlative nature of cMYC-G9a/EZH2 versus immune observations and of HKMTi-1-005 upregulation of cMYC-G9a/EZH2 target genes. Whilst the literature decisively supports cMYC-mediated immune repression (see 1.8.2) and cMYC interaction with G9a and EZH2 (see 1.8.3 and 1.8.4), as well as G9a- and EZH2-mediated immune repression [127, 128, 249], evidence that these axes converge is only beginning to emerge [250, 285].

To validate the causative nature of cMYC-G9a/EZH2 repression of immune targets, several experiments would be of value. For example, co-IP experiments would help to verify the formation of cMYC-G9a and cMYC-EZH2 complexes in cells. Additionally, ChIPseq experiments could pro-

vide insight into where cMYC-G9a or cMYC-EZH2 complexes bind in the genome and into whether cMYC-G9a/EZH2 genome occupancy is accompanied by histone methylation changes. Further, comparison of these analyses between paired *MYC*-amplified and non-amplified cell lines could help to determine whether there are differences in the extent of cMYC-G9a or cMYC-EZH2 binding at target loci dependent on *MYC* status, or in the amount of complex formed. Finally, each of these investigations could be repeated in the context of HKMTi-1-005 treatment; does the compound reduce formation of cMYC-G9a/EZH2 complexes, reduce genome occupancy of these complexes, or reduce their deposition of repressive marks at target loci?

Despite the limitations described, sensitivity data provided further support for a cMYC-G9a/EZH2-immune axis (see 5.5). *MYC* deregulation has been demonstrated to decrease the threshold for VM induced by spliceosome targeting [235], as well as increasing dsRNA accumulation [312], DNA damage and sensitivity to cell death [279]. The data presented in this chapter suggest that *MYC* amplification is associated with increased cell-intrinsic cMYC repression, with reduced immune signalling including lower 7ISG scores. I therefore hypothesised that cell lines with lower basal immune signalling – reflecting greater cMYC-G9a/EZH2 repression – would be more sensitive to HKMTi-1-005 treatment, especially those with *MYC* amplification and a lower threshold for inflammatory cell death. Particularly, *MYC*-amplified lines with low 7ISG scores may comprise a sensitive subgroup with a decreased threshold for VM and with potential for strong upregulation of HKMTi-1-005-induced immune signalling.

Initial analyses into this effect in the GDSC2 panel illustrated the expected heightened sensitivity of *MYC*-amplified versus non-amplified lines to HKMTi-1-005. This was also true of sensitivity to comparator compounds such as cisplatin or single EZH2i/G9ai, suggesting a non-specific increase in sensitivity of *MYC*-amplified cells. Therefore, associations between immune scores and sensitivity were also assessed in this panel, with comparison to a range of compounds such as first-line HGSC treatments, widely used chemotherapies and several epigenetic drugs. Lower immune scores (including for the 7ISG, IFNA, IFNG and innate immune response genesets) were consistently correlated with greater sensitivity to HKMTi-1-005, as well as to the majority of other compounds considered (8 of 11 comparison compounds). Higher canonical cMYC signalling scores correlated with greater sensitivity to every one of the compounds considered, once again supporting the idea that *MYC* deregulation generally increases sensitivity to pharmacological intervention.

Given the limited specificity towards HKMTi-1-005 sensitivity observed in these initial investigations, cell lines were stratified by *MYC* amplification status. In this analysis, correlations suggested a strong trend towards greater predictive value of the 7ISG in the *MYC*-amplified group. Additionally, the relationship between sensitivity and immune scores appeared more specific to HKMTi-1-005 in this cohort, with only 3 of the 11 comparison compounds demonstrating a consistently significant

relationship: UNC0638, olaparib and vorinostat. Yet, canonical cMYC signalling retained a significant relationship with sensitivity for these compounds and for HKMTi-1-005 in the *MYC*-amplified cohort, and canonical cMYC scores strongly anticorrelated with 7ISG scores. Hence, multiple linear regression was used to determine whether 7ISG relationships to sensitivity were independent from canonical cMYC effects. The 7ISG was revealed to be an independent predictor of HKMTi-1-005 sensitivity in both *MYC*-amplified and non-amplified cohorts, whilst canonical cMYC signalling was not. The 7ISG also independently related to sensitivity to UNC0638, olaparib and vorinostat, however so too did canonical cMYC signalling for these drugs.

Taken together, the data support *MYC* amplification and low basal immune signatures – including the reduced 7ISG immune signature – as predictors of sensitivity to the HKMTi-1-005 compound, independent of canonical cMYC activity. Of note, SCLC lines were identified in chapter 3 as particularly sensitive to HKMTi-1-005, an effect that was particularly striking in *MYC*-amplified SCLC lines (appendix figure A.26); literature demonstrates downregulation of IFNG signalling by oncogenic *MYC* activation in SCLC [269], corroborating the hypothesis that *MYC*-deregulated tumours may be a sensitive subgroup to HKMTi-1-005 due to the capacity for immune de-repression in these cells.

HKMTi-1-005 has anti-tumour efficacy *in vitro* and in *in vivo* mouse models [155], with cell line data supporting greater sensitivity in *MYC*-amplified groups. To further assess evidence for HKMTi-1-005 efficacy *in vivo*, I next investigated the prognostic value of the 7ISG in *in vivo* patient cohorts (see 5.6). Particularly, I hypothesised that patient data would support a greater benefit of high 7ISG scores in *MYC*-amplified over non-amplified groups, as *MYC*-amplified tumour cells may be closer to the threshold for VM effects and therefore immunogenic cell death. The ICGC data were first interrogated, without stratification by *MYC* status due to the limited cohort size; they provided preliminary support for association between improved prognosis and higher 7ISG scores, in both uni- and multi-variable analyses.

Insufficient numbers of *MYC*-amplified patients were available for analysis of breast or LuAd patient cohorts, limiting the analysis of the *MYC*-dependency of the 7ISG prognostic benefit. The ovarian TCGA cohort was large enough to be utilised in CoxPH analyses, although was also limited by lack of available data for additional variables such as residual disease following debulking surgery. However, ovarian results were striking; *MYC*-amplified but not non-amplified patients demonstrated a prolonged disease-specific survival when 7ISG scores were high. This held true in both uni- and multi-variable analyses for *MYC*-amplified patients, whilst non-amplified patients showed no relationship between 7ISG scores and survival. Given the earlier findings demonstrating that immune scores may be confounded by non-tumour cell contributions, scores were compared within the different *MYC* status groups rather than between them to reduce the effect of differing immune infiltration.

However, it must be noted that non-tumour contributions cannot be completely negated and high

7ISG scores may reflect beneficial immune infiltration rather than a cell-intrinsic effect. Additionally, the relationship between basal 7ISG score and survival does not confirm that induction of the 7ISG by HKMTi-1-005 will be accompanied by improved prognosis. Further work will need to assess the 7ISG as a predictive biomarker of sensitivity in additional HGSC *in vivo* models, alongside experiments validating the HKMTi-1-005-mediated induction of 7ISG expression and an accompanying increase in survival.

The results discussed so far support sensitivity of *MYC*-deregulated tumour types to HKMTi-1-005 treatment. LuAd was utilised to assess HKMTi-1-005 efficacy in an additional *MYC*-deregulated tumour type aside from HGSC (see 5.7). Treatment of lung carcinoma cell lines supported growth inhibitory effects and cell-intrinsic immune stimulation by HKMTi-1-005, in line with HGSC observations. This included induction of several core ISGs such as *IRF7*, *IRF9* and *STAT2*. Having assessed *in vitro* effects, subsequent analyses assessed *in vivo* HKMTi-1-005 effects in the *MYC*-deregulated KM^2 LuAd model. These data provided evidence that HKMTi-1-005 increases survival of KM^2 LuAd models, with reduced tumour burden and reduced proliferative rate of treated tumours as compared to vehicle.

Several aspects of the LuAd investigations are worth examining further. First, assessing induction of the full 7ISG set in both cell line and mouse models of LuAd will help to determine whether this signature is applicable to other *MYC*-dependent cancers or is HGSC-specific. Measurement of 7ISG scores at baseline in these mouse models will also add to HGSC investigations assessing the predictive value of the 7ISG as a biomarker. Additionally, comparison between the *MYC*-dependent models described and a *MYC*-independent model could validate the increased sensitivity of the former group *in vivo*, alongside the aforementioned comparisons between paired cell lines of different *MYC* statuses to verify the cell-intrinsic nature of this effect.

In conclusion, further *in vivo* experiments will be essential to validate the value of the 7ISG as a predictive biomarker for HKMTi-1-005 sensitivity, as well as to examine the *MYC*-dependency of HKMTi-1-005 effects. Further *in vitro* experiments are required to examine cMYC-G9a/EZH2 complex formation and their immune targets, and to dissect the mechanism of action of HKMTi-1-005-mediated disruption of cMYC axes. However, these preliminary data do support efficacy of HKMTi-1-005 treatment in *MYC*-deregulated cancers and suggest that this effect may be strongest in those with low basal immune signalling. Additionally, they support that induction of cell-intrinsic immune signalling by HKMTi-1-005 is accompanied by relief of cMYC-G9a/EZH2 repression and downregulation of canonical cMYC signalling.

Chapter 6

Overall discussion

6.1 Dual targeting of EZH2/G9a: the clinical and biological context

Current cancer treatment revolves around use of conventional chemotherapy, radiotherapy, hormone therapy, targeted therapy or immunotherapy [461]. One of the major issues limiting the efficacy of these approaches is the emergence of resistance. Epigenetic plasticity is heavily implicated in both primary and acquired resistance, including histone methylation [461]. For example, EZH2 and G9a contribute to resistance to taxanes [195], cisplatin [325] and PARPis [32] in preclinical models. More broadly, epigenetic events have well-described roles in tumour initiation and progression (see 1.5), revealing epigenetic drugs as a promising therapeutic approach. Particularly, epigenetic drugs have potential as combination therapies to sensitise tumours to other treatments or to overcome resistance [461].

Despite these numerous tumourigenic roles for epigenetic modifiers and a wealth of preclinical data supporting the use of epigenetic drugs, conclusive efficacy of epigenetic therapy has been demonstrated almost exclusively in haematological malignancies [136, 462]. Very few epigenetic drugs have been licensed, with drugs targeting HKMTs limited to a single licensed compound targeting EZH2, tazemetostat [143, 144]. The limited efficacy observed may be explained by epigenetic redundancy and the extensive interplay between epigenetic modifiers (such as between EZH2 and G9a, see 1.4.3), rendering the targeting of single epigenetic enzymes insufficient for biological activity. Yet, epigenetic combination therapy trials have largely utilised HDACis and DNMTis and have suffered from limited tolerability to date [461]. Additionally, current approaches are limited by a lack of predictive biomarkers to stratify sensitive patient populations [461] and a lack of specificity for target loci; whilst therapeutic effects are most likely mediated via changes at only a few loci, epigenetic drugs affect loci genome-wide [462]. Hence, research into novel epigenetic drugs including HKMTis and

multi-targeting approaches is vital, to reduce resistance to current anticancer approaches and to broaden the range of responsive tumour types.

Resistance to therapy can also be mediated by the oncogene *MYC* [272], a gene with a multitude of oncogenic roles whose deregulation is found in roughly 70% of tumours [257–261]. *MYC* targeting represents another considerable unmet clinical need; despite the vast numbers of *MYC*-deregulated tumours, direct targeting of the oncogene has proven incredibly challenging due to its nuclear localisation and the lack of a binding pocket [308]. A dominant negative targeting approach – Omomyc – has demonstrated the powerful anticancer efficacy of successful cMYC targeting, where treatment caused complete regression of LuAd tumours in murine models with limited side effects [309]. However, Omomyc is the only cMYC targeting approach to progress beyond phase I clinical trials [463] and a broader arsenal of therapeutic approaches is required to tackle the multitude of ways in which *MYC* can contribute to tumourigenesis.

For example, whilst immunotherapy has revolutionised treatment for several tumour types [269], its efficacy may be limited by oncogenic *MYC* [269, 270, 272]. Preclinical data suggest that cMYC achieves this effect via cell-intrinsic [269, 270, 274, 278] and extrinsic [280] immune suppression. Given the remarkable anticancer potential of immunotherapy, identification of approaches to overcome *MYC*-mediated immunotherapy resistance is highly relevant. Several preclinical investigations support epigenetic drugs as a potential *MYC*-targeting intervention [137, 139, 250, 267, 283, 288, 289, 311]. This is exemplified by the interaction of cMYC with EZH2 and G9a; targeting of either enzyme in various tumour models results in a reduction of *MYC* signalling (see 1.8.3 and 1.8.4). Importantly, preclinical data support the hypothesis that *MYC*-amplified patients may have a particular vulnerability to the combination of epigenetic *MYC* targeting and immunotherapy, via boosting of immune responses [269, 311]. Particularly, patients exhibiting low immune but high *MYC* signalling may be sensitive to *MYC* intervention using epigenetic drugs [311].

Taken together, the literature demonstrates a need for novel methods of *MYC* targeting and suggests that epigenetic drugs may prove valuable in this context, with potential for heightened efficacy in *MYC*-deregulated immune-suppressed cancers. Given the limited efficacy of epigenetic approaches when used as single agents on unselected patient populations [461], the literature also recommends identification of predictive biomarkers for patient stratification. Whilst *MYC* amplification is frequently used as a proxy for cMYC deregulation, other forms of cMYC deregulation such as overexpression or downstream signalling alterations exist. Identification of biomarkers based on the primary downstream effects of *MYC* deregulation in a specific tumour – for example, whether immune or metabolic pathways predominate – would aid in appropriate patient stratification for different cMYC targeting approaches [279].

Clinical advances rely on improved understanding of the underlying biology. How cMYC achieves

its immune repression and the roles of EZH2 and G9a in the process have not been entirely established. Current data demonstrate a direct repression by cMYC at various immune genes [269–278], including those of the IFNG signalling pathway [269, 275]. EZH2 has been shown to drive suppression of IFNG signalling in a *MYC*-dependent manner [250] and G9a directly cooperates with cMYC to repress various ISGs and apoptosis genes in chronic inflammatory contexts [285]. Further, G9a and EZH2 interact with cMYC both directly and indirectly (sections 1.8.3 and 1.8.4), with additional distinct roles in binding and repressing immune targets individually [249, 250].

Despite strong literature unravelling each of these axes individually, few groups have investigated in detail their convergence in the form of direct immune gene repression by cMYC-G9a/cMYC-EZH2 complexes [250, 285]. Given the well-documented EZH2-G9a interplay (section 1.4.3), there is also the possibility that a larger trimeric complex exists (cMYC-G9a-EZH2) with immunosuppressive activity. Further, EZH2 and G9a canonically repress gene expression via histone methylation of their respective targets, yet each have additional non-canonical roles [54, 64]. Hence, the ways in which cMYC-G9a/EZH2 complexes could achieve immune silencing are also unresolved.

In contrast, the role of cMYC as a global amplifier of transcription has been clearly established [258]. cMYC is also involved in splicing regulation [235] and in the aberrant activation of RNA polymerases I and III [444], effects that can promote formation of dsRNAs [312]. Further, *MYC* amplification can promote DNA damage and generally increase the sensitivity of cells to apoptosis [279]. Accumulation of dsRNA in tumour cells would usually stimulate viral mimicry immune signalling with accompanying apoptosis [235]. Yet, a paradox exists: direct comparison of *MYC* activated versus inactivated cells reveals lower immune signalling in *MYC*-deregulated cells [269, 270], reflecting the cell-intrinsic immune repression mediated by cMYC [269–278] (figure 6.1). These findings propose the hypothesis that *MYC* deregulation brings cells closer to the threshold for viral mimicry and inflammatory cell death due to elevated dsRNA levels, yet simultaneously prevents inflammatory signalling via direct immune gene inhibition. Hence, deregulated *MYC* may create a particular vulnerability to EZH2/G9a targeting.

However, questions remain regarding *MYC* deregulation and priming for viral mimicry and cell death. Are all *MYC*-deregulated cells primed for VM, or is an immune-repressed subset the most relevant target group? Will the VM priming effect of *MYC* deregulation observed *in vitro* – reflecting cell-intrinsic *MYC* effects – translate to *in vivo* cohorts, given the strong influence of *MYC* on non-tumour cells in the TME? Is the priming effect of *MYC* deregulation specific to certain cancer types, or ubiquitous? The literature supports further investigation into *MYC* deregulation in the context of VM to resolve these questions and to identify other potential priming events, alongside investigation into the specifics of VM pathways and their induction.

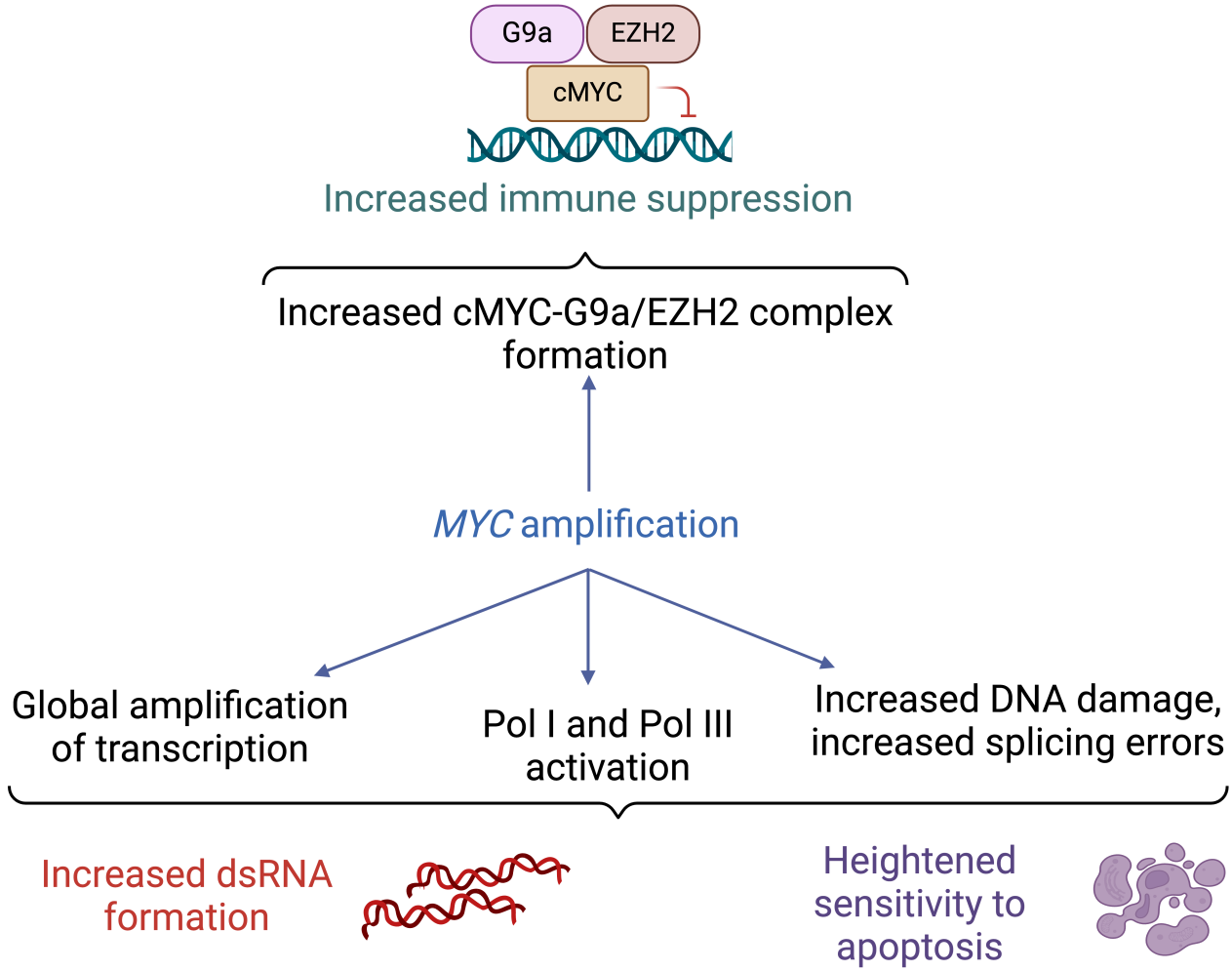


Figure 6.1 **Schematic illustrating the paradox surrounding inflammatory signalling in *MYC*-amplified cells.** Bottom, *MYC* amplification is known to increase global transcription [258], Pol I and Pol III activation [444], DNA damage [279] and splicing errors [235], which may contribute to the increased basal levels of dsRNA [312] and increased sensitivity to apoptosis [235, 279] observed in these cells. Top, *MYC* deregulation reduces immune signalling [269, 270]; this is hypothesised to be the result of increased formation/genome binding of cMYC-G9a/EZH2 complexes and therefore increased repression of immune targets.

6.2 Overview of findings

The data presented in this thesis support inhibitory activity of HKMTi-1-005 at EZH2 and G9a targets. Global accessibility is increased in HGSC cell lines and *in vivo* models as measured by ATACseq, alongside increased accessibility specifically at loci associated with H3K9me3 and H3K27me3 marks (see 3.3). RNAseq assays in the same models revealed that chromatin accessibility changes were accompanied by transcriptional changes at dual EZH2 and G9a targets (see 3.2). Despite evidence for EZH2i-like behaviour of the compound in HGSC models and evidence for H3K27me3 reduction in cell lines, the direct activity of HKMTi-1-005 against EZH2 was weak in cell-free assays (see 3.4.4). These latter data suggested an alternative mechanism of action of the compound – which was originally based on an enzymatic G9ai – against EZH2 targets, aside from direct enzymatic EZH2 inhibition. Although the mechanism of action against EZH2 could not be conclusively addressed in the current work, several hypotheses arose based on these results, such as disruption of EZH2-G9a crosstalk, loss of chromatin targeting or depletion of PRC2 components.

Accompanying G9a and EZH2 targeting effects, HKMTi-1-005 treatment stimulated immune signalling in HGSC models (see 4.2.1). In cell lines, RNAseq data supported upregulation of several intracellular signalling pathways with consistent upregulation of innate immune and IFNG response pathways. The upregulated immune genes included many ISGs, with particularly strong and consistent upregulation of 7 ISGs that were used to create a signature geneset – the 7ISG – characterising the cell-intrinsic immune signalling response induced by the compound. Re-analysis of previously published data examining HKMTi-1-005 treatment of the ID8 *in vivo* HGSC model demonstrated an even stronger upregulation of immune pathways in this model, including the innate immune, IFNG and 7ISG genesets stimulated *in vitro*. It was interesting to note the divergent response of the PEO4 cell line, which showed downregulation of immune pathways. Analysis of the PEO4 line compared to PEO1, derived from the same patient but before (PEO1) or after (PEO4) the onset of platinum resistance, demonstrated that PEO4 had a significantly higher basal expression of ERVs and immune genes including the 7ISG (see 4.4). The data supported the hypothesis described [214] that basal ERV/IFN status may predict the extent of immune induction in response to epigenetic perturbation.

A growing body of literature suggests that TE transcription and downstream VM may be responsible for immune stimulation induced by epigenetic therapy (see 1.7.5). Hence, the effect of HKMTi-1-005 treatment on TEs was assessed in HGSC models at the sub-family and locus-specific levels (see 4.3). These analyses demonstrated that, both *in vitro* and *in vivo*, HKMTi-1-005 upregulates many TEs. ERVs made up the majority of TEs induced in HGSC cell lines and demonstrated the greatest concordance between lines: HERVH-int sub-family elements were upregulated across all cell lines included, and LTR7Y and LTR1C1 were upregulated in most. Notably, the evolutionarily

older ERV1 family was overrepresented amongst upregulated ERVs, supporting the de-repression of older/intermediate age elements that are hypothesised to be more reliant on histone methylation for transcriptional silencing, as compared to younger elements (such as the ERVK family) that rely more on DNA methylation [153]. The HGSC cell line data suggested TE induction may be distinct from that observed with DNMTi [311, 439].

Despite these *in vitro* data showing consistent support for induction of older ERVs, data derived from the ID8 *in vivo* murine model of HGSC were more mixed. ERVs were still the predominant TE type induced by treatment, but within the ERV superfamily the younger ERVK family was most prevalent. Locus-level analyses demonstrated slightly greater ERV1 induction than expected based on genomic abundance, although not to the same extent as *in vitro* data. The data suggested that treatment effect on TEs may be confounded by cellularity changes in the bulk RNAseq performed on *in vivo* samples: extensive ERV changes were observed at the locus-specific level, yet there was a lack of concordance between ERVs induced at the sub-family level. Additionally, the highly distinct mouse TE landscape may have contributed to *in vitro* versus *in vivo* observations [162, 163] and the role of different epigenetic marks in silencing particular ERV families within the mouse genome remains disputed [198, 199].

Expanding analyses to include a publicly available pan-cancer cell line panel demonstrated that *in vitro* sensitivity to the compound extended to many different tumour types (see 3.4). The sensitivity profile of HKMTi-1-005 was highly similar both to single EZH2i and to single G9ai, with increased sensitivity of EZH2/G9a-altered cell lines and of lines derived from haematological malignancies. Yet, median overall sensitivity to the dual targeting compound was heightened as compared to these single inhibitors. Particular sensitivity was evident for cell lines derived from SCLC, sarcoma and melanoma tumour types, with a skew towards low IC_{50} values highlighting the existence of a subset of highly sensitive cells.

Several of the most sensitive cell lines were those with *MYC* amplification. EZH2 and G9a have both been shown to physically interact with cMYC protein as well as demonstrating significant bidirectional regulation at the RNA level (see 1.8.3 and 1.8.4). cMYC mediates transcriptional repression alongside its canonical transcriptional activation, and several of its targets of repression are immune genes including innate immune targets (see 1.8.2). EZH2 and G9a are also each known to repress immune genes [249, 250], with initial evidence supporting the convergence of the cMYC and EZH2/G9a immune repression axes, especially in the context of inflammation and cancer [250, 285]. Hence, the hypothesis that cMYC-G9a/EZH2 complexes repress immune signalling was investigated in the pan-cancer cell line panel and in HKMTi-1-005-treated models.

The pan-cancer cell line data demonstrated a strong anticorrelation between the extent of cMYC-G9a/EZH2 repression and the extent of cell-autonomous immune signalling (see 5.2). These effects

were heightened in cell lines with *MYC* amplification, which had greater basal levels of cMYC-G9a/EZH2 repression and lower basal levels of immune signalling. Assessing these pathways in the context of HKMTi-1-005 treatment demonstrated that the compound was capable of upregulating targets of cMYC-G9a/EZH2 repression alongside immune stimulation in HGSC models (see 5.4). Importantly, the de-repression of cMYC targets was accompanied by transcriptional downregulation of canonical cMYC activation targets, that are considered to be predominantly pro-oncogenic [258].

These data suggested involvement of the cMYC-G9a/EZH2 repression axis in the HKMTi-1-005 mechanism of action, leading to interrogation of the pan-cancer cell line panel for related predictive markers of sensitivity (see 5.5). *MYC*-amplified cell lines were significantly more sensitive to HKMTi-1-005. This effect was true for an array of other drugs, as was the heightened sensitivity in cell lines with lower basal immune signalling. However, considering *MYC*-amplified lines separately, low basal immune signalling was predictive of greater sensitivity to HKMTi-1-005 but not to most other compounds examined. Alongside larger immune genesets, the 7ISG demonstrated potential as a predictive biomarker for HKMTi-1-005 sensitivity in *MYC*-amplified samples. Importantly, this relationship was independent of canonical cMYC signalling.

Taken together, the results supported the relief of *MYC*-mediated immune repression as a potential cause of HKMTi-1-005 efficacy. They suggested that loss of cell viability induced by the compound could be downstream of inflammatory VM, with amplification via direct de-repression of inflammatory pathways targeted by cMYC-G9a/EZH2 complexes. Additionally, they suggested that *MYC*-amplified cell lines with low levels of immune signalling may be particularly sensitive to this pharmacological approach: *MYC*-amplified lines had greater basal levels of cMYC-G9a/EZH2 immune repression, yet literature also suggests *MYC* deregulation may push cells closer to the threshold for inflammatory cell death [235, 279, 312]. Figure 6.2 includes a schematic summarising these hypotheses.

The strong anticorrelation between cMYC-G9a/EZH2 repression and immune signalling *in vitro* held true in bulk RNA data from cohorts of ovarian, breast and LuAd patients (see 5.3.1). In these bulk data, *MYC* amplification was no longer associated with greater levels of repression and lower levels of immune signalling (see 5.3.2). However, analysis of immune infiltration in the patient cohorts highlighted substantial immune composition differences based on *MYC* status and demonstrated the potential for non-tumour cell contribution to measurements of *MYC*-mediated repression or immune signalling (see 5.3.3). The data supported the hyothesis that *MYC* amplification effects on cMYC-G9a/EZH2 repression and immune signalling occur in a cancer cell-intrinsic manner, whilst highlighting that measurement of this effect that may be obscured by cellularity changes in bulk data.

Further analysis of patient data suggested that the HKMTi-1-005-induced 7ISG may have prog-

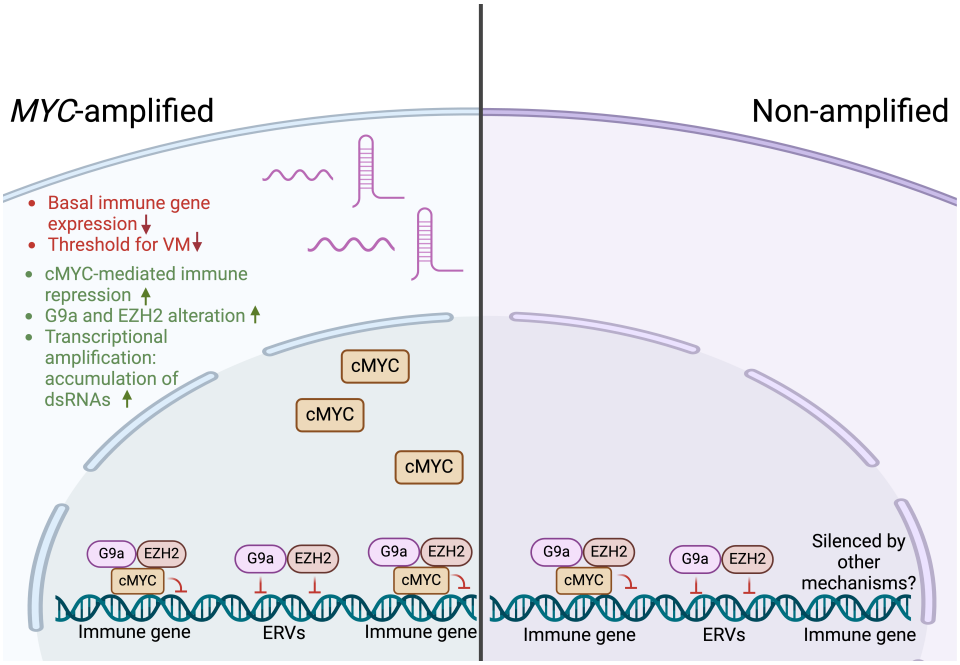
nostic value in ovarian patients (see 5.6). In ovarian cohorts, a higher 7ISG predicted improved survival. Strikingly, analysis of ovarian patients stratified by *MYC* status demonstrated this effect was evident in *MYC*-amplified but not non-amplified ovarian cohorts. HGSC is a highly *MYC*-amplified tumour type and may therefore be particularly sensitive to HKMTi-1-005 treatment and 7ISG induction. Indeed, the immune stimulation described in the ID8 model of HGSC is accompanied by prolonged survival [155].

To corroborate HGSC findings, LuAd was investigated as another highly *MYC*-deregulated tumour type [359] (see 5.7). HKMTi-1-005 had growth inhibitory effects in the *MYC*-overexpressing KM^2 LuAd model, with accompanying immune stimulation and a survival benefit. Importantly, induction of immune signalling pathways including ISGs was observed in LuAd cell lines, supporting the cell-intrinsic immune effect of HKMTi-1-005 treatment.

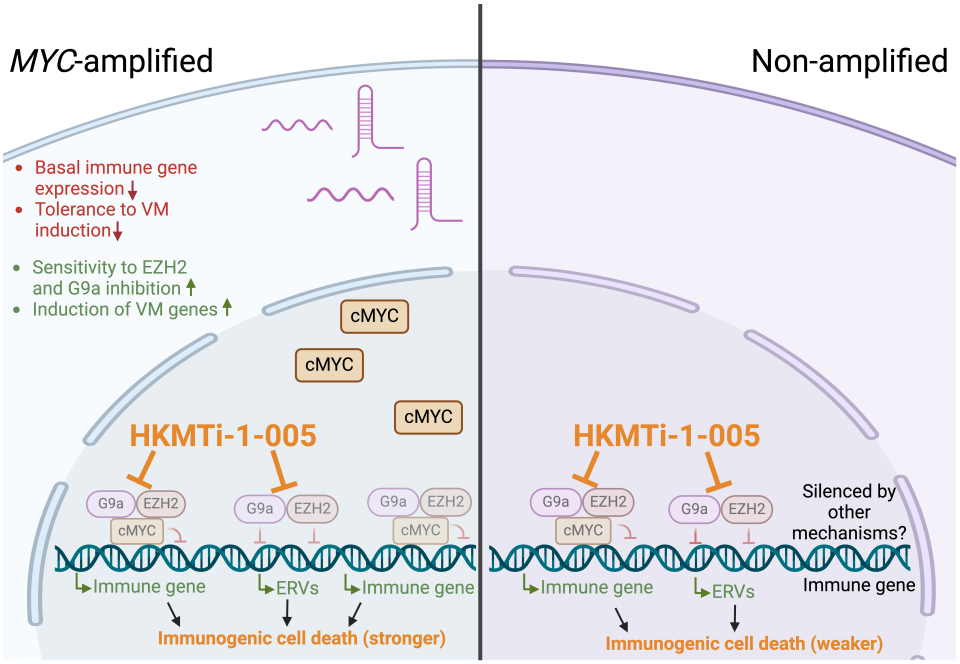
In summary, HKMTi-1-005 represents a candidate compound to tackle some of the clinical issues and biological questions described. The dual EZH2/G9a-targeting nature of the compound may overcome epigenetic redundancy to promote full realisation of HKMT targeting effects. This feature of HKMTi-1-005 would also allow easier delivery and scheduling than the use of two single inhibitors [364], whilst facilitating combination approaches with compounds of different modalities. Additionally, the effect of HKMTi-1-005 on cMYC activity may contribute to efficacy against a broader range of tumour types aside from haematological tumours.

One crucial feature of the next generation of epigenetic drugs is the identification of sensitive patient populations [461]. *MYC*-deregulated lines show heightened sensitivity to HKMTi-1-005; given the large pool of *MYC*-deregulated tumour types, a sizeable target patient population exists. Within *MYC*-amplified tumours, preliminary data suggest the 7ISG may act as a more specific predictive biomarker for HKMTi-1-005 sensitivity. Further, literature suggests that patients with low immune signalling but high *MYC* signalling will be most vulnerable to epigenetic *MYC* targeting combined with immunotherapy [311], supporting analysis of HKMTi-1-005 in this context.

Overall, HKMTi-1-005 represents a novel epigenetic compound that has demonstrated strong promise in preclinical analyses to date, with a dual EZH2/G9a targeting effect, efficacy against a wide range of tumour types *in vitro*, and with potential predictive biomarkers to guide further experiments.



(a)



(b)

Figure 6.2 **Schematic overview of hypothesised cMYC-G9a/EZH2 and EZH2/G9a repression of immune genes and ERVs, the effect of *MYC* amplification on these axes and the disruption of repression by HKMTi-1-005.** (a) In untreated cancer cells, cMYC-G9a/EZH2 complexes repress immune genes. G9a and EZH2 also separately repress immune genes and TEs including ERVs. *MYC* amplification (left) leads to more repression by cMYC-G9a/EZH2 than in non-amplified cells (right) and lower immune gene expression, as well as increased accumulation of dsRNAs and increased global transcription, with a heightened sensitivity to cell death (see 6.3.3). *MYC* alteration also significantly co-occurs with EZH2 and G9a alterations [268] (appendix table A.2). (b) HKMTi-1-005 treatment induces immunogenic cell death via a combination of direct de-repression of immune genes and of ERVs. This effect is stronger in *MYC*-amplified cells (left), where there is greater basal cMYC-G9a/EZH2 repression of immune genes, and a lower threshold for inflammatory cell death.

6.3 Limitations and future directions

6.3.1 Overview

Due to the largely bioinformatic nature of analyses, the current data are limited with regards to causative evidence supporting the hypotheses proposed. To aid in the progression of HKMTi-1-005 towards the clinic, several outstanding research avenues should be addressed.

First, the mechanism of action (MoA) of HKMTi-1-005 requires clarification (figure 6.3). How does the compound exert its EZH2i effect, and how does it target cMYC? Does HKMTi-1-005 inhibit cMYC-G9a/EZH2 complexes to induce immune stimulation or are effects entirely VM-dependent, and how does *MYC* deregulation influence these axes?

Second, further *in vivo* data are required. The use of *MYC* status and the 7ISG as predictive biomarkers need validation in *in vivo* models. In addition, combination approaches and dosing schedules should be investigated to optimise anti-tumour efficacy whilst minimising treatment side effects. Finally, deeper investigation into some of the pharmacological features of the drug will be necessary for HKMTi-1-005 progression, including comparison of its efficacy to other EZH2/G9a-targeting approaches, assessment of its tolerability profile, off-target effects, pharmacokinetics and pharmacodynamics.

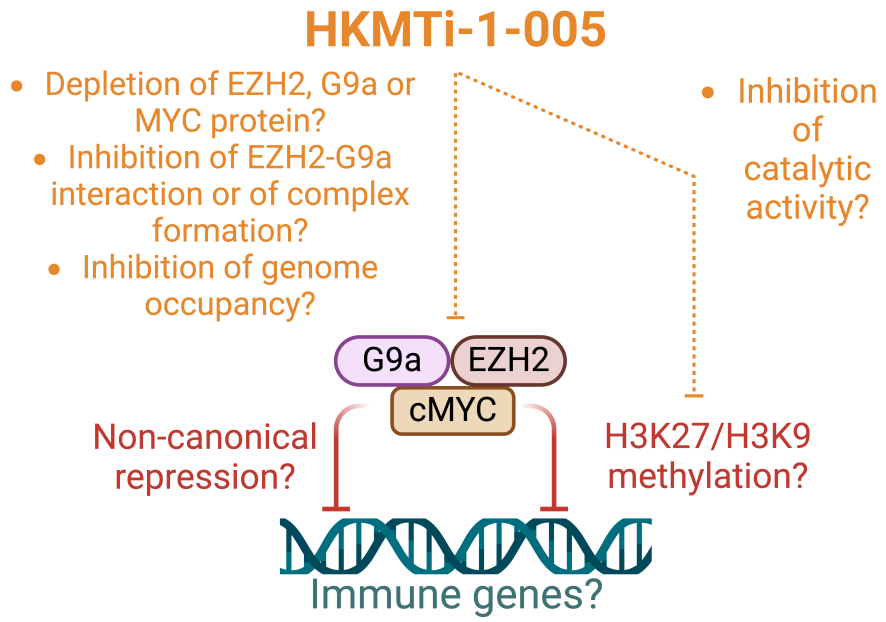


Figure 6.3 **Outstanding questions: cMYC-G9a/EZH2 repression and HKMTi-1-005 treatment mechanisms.** Hypotheses regarding the mechanism by which the cMYC-G9a/EZH2 complex(es) repress immune genes, and the manner in which HKMTi-1-005 disrupts this repression.

6.3.2 cMYC-G9a/EZH2 complexes: formation, genomic targets and treatment effect

To address the MoA questions posed, several pivotal experiments are required; particularly, experiments investigating the proposed cMYC-G9a/EZH2 complexes. It will be important to identify where cMYC-G9a and cMYC-EZH2 complexes bind in the genome and how HKMTi-1-005 treatment affects them. Experiments could involve ChIPseq to measure cMYC, G9a and EZH2 binding to chromatin alongside measuring H3K9/H3K27 methylation, with and without HKMTi-1-005 treatment. This would allow measurement of the extent of cMYC/G9a/EZH2 co-occupancy in the genome as well as identification of target genomic loci and methylation at these loci. Combined with co-IP experiments, these analyses could also reveal whether cMYC-G9a and cMYC-EZH2 are all part of one larger trimeric complex. Comparison of control versus HKMTi-1-005 treatment conditions will contribute to understanding the MoA of HKMTi-1-005; is the formation of cMYC-G9a/EZH2 complexes inhibited, is genome occupancy reduced, or are repressive histone methylation marks inhibited?

These experiments may also provide insight into cMYCi and EZH2i effects of HKMTi-1-005. If cMYC or EZH2 binding to the chromatin is reduced by treatment, ChIPseq data should reflect this. Additional western blot analysis could be used to reveal whether there is protein depletion of PRC2 components including EZH2, alongside analysis of any depletion effect on cMYC protein. Potentially, neither genome occupancy nor protein levels of PRC2 components will be perturbed by HKMTi-1-005. In this instance, disruption of EZH2-G9a crosstalk by HKMTi-1-005 could be measured using co-IP to investigate this third hypothesis for EZH2i effects. Additional comparisons between HKMTi-1-005 and EZH2-targeting PROteolysis TArgeting Chimeras (PROTACs) could be valuable in this context. Given that PROTACs degrade EZH2 [136], similarity between HKMTi-1-005 and PROTAC treatment effects would provide support for HKMTi-1-005-mediated destabilisation of the PRC2 complex.

6.3.3 Understanding the priming effect of *MYC* amplification and basal 7ISG scores

It has already been demonstrated that *MYC* activation reduces immune signalling, using using paired lines of differing *MYC* activation or expression statuses [269, 270]. Further analysis of *MYC*-amplified compared to non-amplified cell lines from the same tumour type will allow investigation into the cMYC-dependency of HKMTi-1-005 effects and sensitivity, and the value of the 7ISG in this context. First, basal levels of dsRNA could be compared between untreated lines of different *MYC* statuses using the J2 anti-dsRNA monoclonal antibody. Combined with the co-IP and ChIPseq analyses described, as well as RNAseq, these experiments will examine how *MYC* status predisposes

cells to VM and increases HKMTi-1-005 sensitivity. Specifically, they will examine whether *MYC* amplification increases basal dsRNA accumulation, whether it alters the amount of cMYC-G9a/EZH2 complex formation or genome occupancy, and how these features relate to basal inflammatory gene expression including 7ISG scores.

Second, RNAseq analyses of HKMTi-1-005 treatment effect in the same lines will help to assess the differences in transcriptional responses by *MYC* and 7ISG statuses, to validate whether increased relief of inflammatory gene expression contributes to the greater sensitivity of *MYC*-amplified immune-repressed lines. Particularly, these analyses will address the hypothesis that *MYC*-amplified lines with low 7ISG scores will show greater induction of inflammatory genes, accompanied by greater sensitivity to the HKMTi-1-005 (as measured by cell viability assays). In contrast, non-amplified lines with high basal 7ISG scores are expected to show lower sensitivity and poor induction of an immune signalling response. Together, these data will verify whether basal *MYC* and 7ISG statuses have value as predictive biomarkers and give insight into the underlying biology of the effect.

6.3.4 *In vivo* experiments

A suite of *in vivo* experiments will also be essential for the progression of HKMTi-1-005. Strong evidence for efficacy, safety and on-target effects of the drug are a requirement for entry into phase I clinical trials. Given the hypothesised role for immune responses and *MYC* axes in anti-tumour efficacy, immunocompetent *MYC*-deregulated models should be used. HGSC and LuAd models meet these criteria and have demonstrated survival benefits with HKMTi-1-005 treatment, therefore represent promising tumour types for further analyses.

For example, the PPNM model of HGSC with genotype $Tr\boldsymbol{p}53^{-/-R172H}\boldsymbol{P}ten^{-/-}\boldsymbol{N}f1^{-/-}\boldsymbol{M}yc^{OE}$ may help to validate anti-tumour efficacy in another representative *MYC*-driven HGSC model. LuAd investigations could include the KP model (genotype: $lsl\text{-}\boldsymbol{K}Ras^{G12D};lsl\text{-}Tr\boldsymbol{p}53^{R172H/+}$) without *MYC* OE, but likely to have *MYC* deregulation due to the co-dependent nature of *KRAS* and *MYC* in LuAd [309]. Baseline 7ISG scores could be measured in each of the models and correlated with HKMTi-1-005 sensitivity, to assess the 7ISG as a predictive biomarker *in vivo*. Additional experiments should involve testing of pharmacodynamic responses to HKMTi-1-005 using ELISA and immunohistochemistry to validate H3K9/H3K27me reduction, with assessment of the pharmacokinetics, toxicity and tolerability of the compound in tandem to provide the pharmacological basis for entry into clinical trials. If *in vivo* experiments in HGSC and LuAd models continue to support HKMTi-1-005 efficacy, investigations could be extended to assess other tumour types that demonstrate *in vitro* sensitivity such as melanoma or sarcoma, tumour types that also show frequent *MYC* deregulation [121–123].

Importantly, *in vivo* experiments should include determination of whether HKMTi-1-005 is the

optimum compound for dual EZH2i/G9ai effects: the anti-tumour efficacy and pharmacology of HKMTi-1-005 should first be compared to single EZH2i or G9ai compounds, as well as to other compounds in the original HKMTi screening series that demonstrated high EZH2i activity. At the time of writing, only a single other dual EZH2/G9a-targeting compound has been described, Compound 15h (also known as as SU08) [154, 445]; this compound should be included in HKMTi-1-005 comparison experiments. Regardless of the outcome of such investigations, proof-of-principle analyses of dual EZH2/G9a targeting effects and sensitivity are a vital component of progressing EZH2/G9a inhibitors towards phase I clinical trials.

Finally, although not essential for entry into clinical trials, the *in vivo* experiments should assess combination approaches. Evidence exists for many potential combination approaches with epigenetic targeting, such as classic chemotherapy, alternative epigenetic agents, PARPi or immunotherapy [461]. Our group has preliminarily assessed HKMTi-1-005 in combination with cisplatin or anti-PD1, but with limited ability to draw conclusions from the mixed results due to small numbers of mice used and a lack of optimisation for drug scheduling [464]. Further experimentation should first include analysis of synergistic effects *in vitro*, followed by *in vivo* testing guided by the results. Testing of different drug schedules should be included *in vivo*, as these can have significant impact on combination efficacy [461].

6.4 Final remarks

Epigenetic therapy represents a highly promising avenue for effective anticancer treatment. Current limitations – such as inadequate efficacy in solid tumours and a lack of predictive biomarkers for patient selection – may prove surmountable with novel epigenetic compounds and a greater understanding of the underlying biology. The dual EZH2/G9a-targeting HKMTi-1-005 has efficacy against a broad range of tumour types *in vitro* and provides a survival benefit with immune stimulation in HGSC and LuAd models *in vivo*. Preliminary work suggests that *MYC*-deregulated tumours may be particularly sensitive to the compound and propose the use of a 7ISG immune signature as a stratification biomarker within this cohort. These data reinforce recent advances in the literature regarding epigenetic contributions to *MYC*-mediated immune repression. Further research is warranted into the mechanism of action of HKMTi-1-005 and its *in vivo* efficacy to assess whether the compound should be taken forward into first-in-human phase I clinical trials.

Chapter 7

Bibliography

Bibliography

1. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., *et al. Molecular Biology of the Cell. 4th edition.* Chromosomal DNA and Its Packaging in the Chromatin Fiber. Available at https://www.ncbi.nlm.nih.gov/books/NBK26834/. (Garland Science, New York, 2002).
2. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128,** 693–705 (2007).
3. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421,** 448–453 (2003).
4. Trojer, P. & Reinberg, D. Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Mol. Cell* **28,** 1–13 (2007).
5. Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326,** 289–293 (2009).
6. Fortin, J.-P. & Hansen, K. D. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biol.* **16,** 180 (2015).
7. Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
8. Grubert, F., Srivas, R., Spacek, D., Kasowski, M., Ruiz-Velasco, M., *et al.* Landscape of cohesin-mediated chromatin loops in the human genome. *Nature* **583,** 737–743 (2020).
9. Ji, X., Dadon, D. B., Powell, B. E., Fan, Z. P., Borges-Rivera, D., *et al.* 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell* **18,** 262–75 (2016).
10. Rao, S. P., Huntley, M., Durand, N., Stamenova, E., Bochkov, I., *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159,** 1665–1680 (2014).
11. Boltsis, I., Grosveld, F., Giraud, G. & Kolovos, P. Chromatin Conformation in Development and Disease. *Front. Cell Dev. Biol.* **9,** 723859 (2021).
12. Smith, E. M., Lajoie, B. R., Jain, G. & Dekker, J. Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus. *Am. J. Hum. Genet.* **98,** 185–201 (2016).

13. Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., *et al.* Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24,** 390–400 (2014).

14. Long, H. S., Greenaway, S., Powell, G., Mallon, A.-M., Lindgren, C. M., *et al.* Making sense of the linear genome, gene function and TADs. *Epigenet. Chromatin* **15,** 4 (2022).

15. Dixon, J., Gorkin, D. & Ren, B. Chromatin Domains: The Unit of Chromosome Organization. *Mol. Cell* **62,** 668–680 (2016).

16. Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453,** 948–951 (2008).

17. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. U. S. A.* **115,** E6697–E6706 (2018).

18. Harris, H. L., Gu, H., Olshansky, M., Wang, A., Farabella, I., *et al.* Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. *Nat. Commun.* **14,** 3303 (2023).

19. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* **28,** 817–825 (2010).

20. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17,** 487–500 (2016).

21. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403,** 41–45 (2000).

22. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293,** 1074–80 (2001).

23. Lee, J. S., Smith, E. & Shilatifard, A. The language of histone crosstalk. *Cell* **142,** 682–5 (2010).

24. Millán-Zambrano, G., Burton, A., Bannister, A. J. & Schneider, R. Histone post-translational modifications — cause and consequence of genome function. *Nat. Rev. Genet.* **23,** 563–580 (2022).

25. Barth, T. K. & Imhof, A. Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem. Sci.* **35,** 618–626 (2010).

26. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21,** 381–95 (2011).

27. Kabi, M. & Filion, G. J. Heterochromatin: did H3K9 methylation evolve to tame transposons? *Genome Biol.* **22,** 325 (2021).

28. Padeken, J., Methot, S. P. & Gasser, S. M. Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. *Nat. Rev. Mol. Cell Biol.* **23,** 623–640 (2022).

29. Shankar, S. R., Bahirvani, A. G., Rao, V. K., Bharathy, N., Ow, J. R., *et al.* G9a, a multipotent regulator of gene expression. *Ciba. F. Symp.* **8,** 16–22 (2013).

30. Fritsch, L., Robin, P., Mathieu, J. R., Souidi, M., Hinaux, H., *et al.* A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. *Mol. Cell* **37,** 46–56 (2010).

31. Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* **49,** e324 (2017).

32. Watson, Z. L., Yamamoto, T. M., McMellen, A., Kim, H., Hughes, C. J., *et al.* Histone methyltransferases EHMT1 and EHMT2 (GLP/G9A) maintain PARP inhibitor resistance in high-grade serous ovarian carcinoma. *Clin. Epigenetics* **11,** 165 (2019).

33. Shinkai, Y. & Tachibana, M. H3K9 methyltransferase G9a and the related molecule GLP. *Genes Dev.* **25,** 781–8 (2011).

34. Chaturvedi, C.-P. & Brand, M. *EHMT2 (euchromatic histone-lysine N-methyltransferase 2).* Web Page, available at https://atlasgeneticsoncology.org/gene/51148/ehmt2-(euchromatic-histone-lysine-n-methyltransferase-2). 2013.

35. Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., *et al.* G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* **16,** 1779–91 (2002).

36. Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., *et al.* Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev.* **19,** 815–26 (2005).

37. Tachibana, M., Matsumura, Y., Fukuda, M., Kimura, H. & Shinkai, Y. G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. *EMBO J.* **27,** 2681–90 (2008).

38. Yan, Z., Ji, L., Huo, X., Wang, Q., Zhang, Y., *et al.* G9a/GLP-sensitivity of H3K9me2 Demarcates Two Types of Genomic Compartments. *Genom. Proteom. Bioinform.* **18,** 359–370 (2020).

39. Loyola, A., Tagami, H., Bonaldi, T., Roche, D., Quivy, J. P., *et al.* The HP1alpha-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. *EMBO Rep.* **10,** 769–75 (2009).

40. Patnaik, D., Chin, H. G., Estève, P. O., Benner, J., Jacobsen, S. E., *et al.* Substrate specificity and kinetic mechanism of mammalian G9a histone H3 methyltransferase. *J. Biol. Chem.* **279,** 53248–58 (2004).

41. Mozzetta, C., Boyarchuk, E., Pontis, J. & Ait-Si-Ali, S. Sound of silence: the properties and functions of repressive Lys methyltransferases. *Nat. Rev. Mol. Cell Biol.* **16,** 499–513 (2015).

42. Mozzetta, C., Pontis, J. & Ait-Si-Ali, S. Functional Crosstalk Between Lysine Methyltransferases on Histone Substrates: The Case of G9A/GLP and Polycomb Repressive Complex 2. *Antioxid. Redox Signal.* **22,** 1365–81 (2015).

43. Ginjala, V., Rodriguez-Colon, L., Ganguly, B., Gangidi, P., Gallina, P., *et al.* Protein-lysine methyltransferases G9a and GLP1 promote responses to DNA damage. *Sci. Rep.* **7,** 16613 (2017).

44. Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410,** 120–4 (2001).

45. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410,** 116–20 (2001).

46. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19,** 229–244 (2018).

47. Canzio, D., Chang, E. Y., Shankar, S., Kuchenbecker, K. M., Simon, M. D., *et al.* Chromodomain-Mediated Oligomerization of HP1 Suggests a Nucleosome-Bridging Mechanism for Heterochromatin Assembly. *Mol. Cell* **41,** 67–81 (2011).

48. El Gazzar, M., Yoza, B. K., Chen, X., Hu, J., Hawkins, G. A., *et al.* G9a and HP1 couple histone and DNA methylation to TNFalpha transcription silencing during endotoxin tolerance. *J. Biol. Chem.* **283,** 32198–208 (2008).

49. Smallwood, A., Estève, P. O., Pradhan, S. & Carey, M. Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev.* **21,** 1169–78 (2007).

50. Briand, N. & Collas, P. Lamina-associated domains: peripheral matters and internal affairs. *Genome Biol.* **21,** 85 (2020).

51. See, K., Kiseleva, A. A., Smith, C. L., Liu, F., Li, J., *et al.* Histone methyltransferase activity programs nuclear peripheral genome positioning. *Dev. Biol.* **466,** 90–98 (2020).

52. Haebe, J. R., Bergin, C. J., Sandouka, T. & Benoit, Y. D. Emerging role of G9a in cancer stemness and promises as a therapeutic target. *Oncogenesis* **10,** 76 (2021).

53. Jiang, Q., Ang, J. Y. J., Lee, A. Y., Cao, Q., Li, K. Y., *et al.* G9a Plays Distinct Roles in Maintaining DNA Methylation, Retrotransposon Silencing, and Chromatin Looping. *Cell Rep.* **33,** 108315 (2020).

54. Casciello, F., Windloch, K., Gannon, F. & Lee, J. S. Functional Role of G9a Histone Methyltransferase in Cancer. *Front. Immunol.* **6,** 487 (2015).

55. Chin, H. G., Estève, P. O., Pradhan, M., Benner, J., Patnaik, D., *et al.* Automethylation of G9a and its implication in wider substrate specificity and HP1 binding. *Nucleic Acids Res.* **35,** 7313–23 (2007).

56. Poulard, C., Bittencourt, D., Wu, D. Y., Hu, Y., Gerke, D. S., *et al.* A post-translational modification switch controls coactivator function of histone methyltransferases G9a and GLP. *EMBO Rep.* **18,** 1442–1459 (2017).

57. Lee, J. S., Kim, Y., Kim, I. S., Kim, B., Choi, H. J., *et al.* Negative regulation of hypoxic responses via induced Reptin methylation. *Mol. Cell* **39,** 71–85 (2010).

58. Lee, J. S., Kim, Y., Bhin, J., Shin, H.-J. R., Nam, H. J., *et al.* Hypoxia-induced methylation of a pontin chromatin remodeling factor. *Proc. Natl. Acad. Sci.* **108,** 13510–13515 (2011).

59. Purcell, D. J., Jeong, K. W., Bittencourt, D., Gerke, D. S. & Stallcup, M. R. A Distinct Mechanism for Coactivator versus Corepressor Function by Histone Methyltransferase G9a in Transcriptional Regulation. *J. Biol. Chem.* **286,** 41963–41971 (2011).

60. Jeon, H. Y., Hussain, A. & Qi, J. Role of H3K9 demethylases in DNA double-strand break repair. *J. Cancer Biol.* **1,** 10–15 (2020).

61. Laugesen, A., Højfeldt, J. W. & Helin, K. Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol. Cell* **74,** 8–18 (2019).

62. Chammas, P., Mocavini, I. & Di Croce, L. Engaging chromatin: PRC2 structure meets function. *Br. J. Cancer* **122,** 315–328 (2020).

63. Laugesen, A., Højfeldt, J. W. & Helin, K. Role of the Polycomb Repressive Complex 2 (PRC2) in Transcriptional Regulation and Cancer. *Cold Spring Harb. Perspect. Med.* **6,** a026575 (2016).

64. Duan, R., Du, W. & Guo, W. EZH2: a novel target for cancer treatment. *J. Hematol. Oncol.* **13,** 104 (2020).

65. Lee, S. H., Li, Y., Kim, H., Eum, S., Park, K., *et al.* The role of EZH1 and EZH2 in development and cancer. *BMB Rep.* **55,** 595–601 (12 2022).

66. Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., *et al.* Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* **32,** 503–18 (2008).

67. Healy, E., Mucha, M., Glancy, E., Fitzpatrick, D. J., Conway, E., *et al.* PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. *Mol. Cell* **76,** 437–452.e6 (2019).

68. Shi, Y., Wang, X.-x., Zhuang, Y.-w., Jiang, Y., Melcher, K., *et al.* Structure of the PRC2 complex and application to drug discovery. *Acta Pharmacol. Sin.* **38,** 963–976 (2017).

69. Blackledge, N. P. & Klose, R. J. The molecular principles of gene regulation by Polycomb repressive complexes. *Nat. Rev. Mol. Cell Biol.* **22,** 815–833 (2021).

70. Skalska, L., Begley, V., Beltran, M., Lukauskas, S., Khandelwal, G., *et al.* Nascent RNA antagonizes the interaction of a set of regulatory proteins with chromatin. *Mol. Cell* **81,** 2944–2959.e10 (2021).

71. Ferrari, K. J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., *et al.* Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol. Cell* **53,** 49–62 (2014).

72. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469,** 343–349 (2011).

73. Wiles, E. T. & Selker, E. U. H3K27 methylation: a promiscuous repressive chromatin mark. *Curr. Opin. Genet. Dev.* **43,** 31–37 (2017).

74. Tan, J.-z., Yan, Y., Wang, X.-x., Jiang, Y. & Xu, H. E. EZH2: biology, disease, and structure-based drug discovery. *Acta Pharmacol. Sin.* **35,** 161–174 (2014).

75. Pauler, F. M., Sloane, M. A., Huang, R., Regha, K., Koerner, M. V., *et al.* H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Res.* **19,** 221–33 (2009).

76. Liu, X. PRC2, Chromatin Regulation, and Human Disease: Insights From Molecular Structure and Function. *Front. Oncol.* **12,** 894585 (2022).

77. Kraft, K., Yost, K. E., Murphy, S. E., Magg, A., Long, Y., *et al.* Polycomb-mediated genome architecture enables long-range spreading of H3K27 methylation. *Proc. Natl. Acad. Sci.* **119,** e2201883119 (2022).

78. Cruz-Molina, S., Respuela, P., Tebartz, C., Kolovos, P., Nikolic, M., *et al.* PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell* **20,** 689–705.e9 (2017).

79. Ngan, C. Y., Wong, C. H., Tjong, H., Wang, W., Goldfeder, R. L., *et al.* Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. *Nat. Genet.* **52,** 264–272 (2020).

80. He, A., Shen, X., Ma, Q., Cao, J., von Gise, A., *et al.* PRC2 directly methylates GATA4 and represses its transcriptional activity. *Genes Dev.* **26,** 37–42 (2012).

81. Kim, J., Lee, Y., Lu, X., Song, B., Fong, K.-W., *et al.* Polycomb- and Methylation-Independent Roles of EZH2 as a Transcription Activator. *Cell Rep.* **25,** 2808–2820.e4 (2018).

82. Vasanthakumar, A., Xu, D., Lun, A. T. L., Kueh, A. J., van Gisbergen, K. P. J. M., *et al.* A non-canonical function of Ezh2 preserves immune homeostasis. *EMBO Rep.* **18,** 619–631 (2017).

83. Swigut, T. & Wysocka, J. H3K27 Demethylases, at Long Last. *Cell* **131,** 29–32 (2007).

84. Pan, M. R., Hsu, M. C., Chen, L. T. & Hung, W. C. Orchestration of H3K27 methylation: mechanisms and therapeutic implication. *Cell Mol. Life Sci.* **75,** 209–223 (2018).

85. Xu, C., Bian, C., Yang, W., Galka, M., Ouyang, H., *et al.* Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proc. Natl. Acad. Sci. U. S. A.* **107,** 19266–71 (2010).

86. Binda, O. On your histone mark, SET, methylate! *Epigenetics* **8,** 457–463 (2013).

87. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20,** 590–607 (2019).

88. Dhar, G. A., Saha, S., Mitra, P. & Nag Chaudhuri, R. DNA methylation and regulation of gene expression: Guardian of our health. *Nucleus (Calcutta)* **64,** 259–270 (2021).

89. Jin, B., Li, Y. & Robertson, K. D. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer* **2,** 607–17 (2011).

90. De Mendoza, A., Nguyen, T. V., Ford, E., Poppe, D., Buckberry, S., *et al.* Large-scale manipulation of promoter DNA methylation reveals context-specific transcriptional responses and stability. *Genome Biol.* **23,** 163 (2022).

91. Miller, J. L. & Grant, P. A. The role of DNA methylation and histone modifications in transcriptional regulation in humans. *Subcell Biochem.* **61,** 289–317 (2013).

92. Du, J., Johnson, L. M., Jacobsen, S. E. & Patel, D. J. DNA methylation pathways and their crosstalk with histone methylation. *Nat. Rev. Mol. Cell Biol.* **16,** 519–32 (2015).

93. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **10,** 295–304 (2009).

94. Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125,** 315–326 (2006).

95. Matsumura, Y., Nakaki, R., Inagaki, T., Yoshida, A., Kano, Y., *et al.* H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. *Mol. Cell* **60,** 584–596 (2015).

96. Crispatzu, G., Rehimi, R., Pachano, T., Bleckwehl, T., Cruz-Molina, S., *et al.* The chromatin, topological and regulatory properties of pluripotency-associated poised enhancers are conserved in vivo. *Nat. Commun.* **12,** 4344 (2021).

97. Yu, Y., Li, X., Jiao, R., Lu, Y., Jiang, X., *et al.* H3K27me3-H3K4me1 transition at bivalent promoters instructs lineage specification in development. *Cell & Bioscience* **13,** 66 (2023).

98. Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H. F., *et al.* Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8,** 532–8 (2006).

99. Pietersen, A. M. & van Lohuizen, M. Stem cell regulation by polycomb repressors: postponing commitment. *Curr. Opin. Cell Biol.* **20,** 201–7 (2008).

100. Kumar, D., Cinghu, S., Oldfield, A. J., Yang, P. & Jothi, R. Decoding the function of bivalent chromatin in development and cancer. *Genome Res.* **31,** 2170–84 (2021).

101. Sachs, M., Onodera, C., Blaschke, K., Ebata, K. T., Song, J. S., *et al.* Bivalent chromatin marks developmental regulatory genes in the mouse embryonic germline in vivo. *Cell Rep.* **3,** 1777–84 (2013).

102. Barral, A., Pozo, G., Ducrot, L., Papadopoulos, G. L., Sauzet, S., *et al.* SETDB1/NSD-dependent H3K9me3/H3K36me3 dual heterochromatin maintains gene expression profiles by bookmarking poised enhancers. *Mol. Cell* **82,** 816–832.e12 (2022).

103. Mauser, R., Kungulovski, G., Keup, C., Reinhardt, R. & Jeltsch, A. Application of dual reading domains as novel reagents in chromatin biology reveals a new H3K9me3 and H3K36me2/3 bivalent chromatin state. *Epigenet. Chromatin* **10,** 45 (2017).

104. Lu-Culligan, W. J., Connor, L. J., Xie, Y., Ekundayo, B. E., Rose, B. T., *et al.* Acetyl-methyllysine marks chromatin at active transcription start sites. *Nature* **622,** 173–179 (2023).

105. Zentner, G. E., Tesar, P. J. & Scacheri, P. C. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* **21,** 1273–83 (2011).

106. Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., *et al.* The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. *Mol. Cell* **53,** 277–89 (2014).

107. Casciello, F., Kelly, G. M., Ramarao-Milne, P., Kamal, N., Stewart, T. A., *et al.* Combined Inhibition of G9a and EZH2 Suppresses Tumor Growth via Synergistic Induction of IL24-Mediated Apoptosis. *Cancer Res.* **82,** 1208–1221 (2022).

108. Coward, W. R., Brand, O. J., Pasini, A., Jenkins, G., Knox, A. J., *et al.* Interplay between EZH2 and G9a Regulates CXCL10 Gene Repression in Idiopathic Pulmonary Fibrosis. *Am. J. Respir. Cell Mol. Biol.* **58,** 449–460 (2018).

109. Curry, E., Green, I., Chapman-Rothe, N., Shamsaei, E., Kandil, S., *et al.* Dual EZH2 and EHMT2 histone methyltransferase inhibition increases biological efficacy in breast cancer cells. *Clin. Epigenetics* **7,** 84 (2015).

110. Ishiguro, K., Kitajima, H., Niinuma, T., Maruyama, R., Nishiyama, N., *et al.* Dual EZH2 and G9a inhibition suppresses multiple myeloma cell proliferation by regulating the interferon signal and IRF4-MYC axis. *Cell Death Discov.* **7,** 7 (2021).

111. Liu, F., Chen, X., Allali-Hassani, A., Quinn, A. M., Wigle, T. J., *et al.* Histone methyltransferase G9a contributes to H3K27 methylation in vivo. *Cell Res.* **21,** 365–367 (2011).

112. Boros, J., Arnoult, N., Stroobant, V., Collet, J. F. & Decottignies, A. Polycomb repressive complex 2 and H3K27me3 cooperate with H3K9 methylation to maintain heterochromatin protein 1α at chromatin. *Mol. Cell Biol.* **34,** 3662–74 (2014).

113. Fong, K.-w., Zhao, J. C., Lu, X., Kim, J., Piunti, A., *et al.* PALI1 promotes tumor growth through competitive recruitment of PRC2 to G9A-target chromatin for dual epigenetic silencing. *Mol. Cell* **82,** 4611–4626.e7 (2022).

114. Pham, V., Pitti, R., Tindell, C. A., Cheung, T. K., Masselot, A., *et al.* Proteomic Analyses Identify a Novel Role for EZH2 in the Initiation of Cancer Cell Drug Tolerance. *J. Proteome Res.* **19,** 1533–1547 (2020).

115. Tachibana, M., Sugimoto, K., Fukushima, T. & Shinkai, Y. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J. Biol. Chem.* **276,** 25309–17 (2001).

116. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* **16,** 2893–905 (2002).

117. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* **12,** 31–46 (2022).

118. Darwiche, N. Epigenetic mechanisms and the hallmarks of cancer: an intimate affair. *Am. J. Cancer Res.* **10,** 1954–1978 (2020).

119. Jiang, T., Wang, Y., Zhou, F., Gao, G., Ren, S., *et al.* Prognostic value of high EZH2 expression in patients with different types of cancer: a systematic review with meta-analysis. *Oncotarget* **7,** 4584–97 (2016).

120. Segovia, C., San José-Enériz, E., Munera-Maravilla, E., Martínez-Fernández, M., Garate, L., *et al.* Inhibition of a G9a/DNMT network triggers immune-mediated bladder cancer regression. *Nat. Med.* **25,** 1073–1081 (2019).

121. Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2,** 401–4 (2012).

122. De Bruijn, I., Kundra, R., Mastrogiacomo, B., Tran, T. N., Sikina, L., *et al.* Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. *Cancer Res.* **83,** 3861–3867 (2023).

123. Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6,** l1 (2013).

124. Poulard, C., Noureddine, L. M., Pruvost, L. & Le Romancer, M. Structure, Activity, and Function of the Protein Lysine Methyltransferase G9a. *Life (Basel)* **11,** 1082 (2021).

125. Jan, S., Dar, M. I., Wani, R., Sandey, J., Mushtaq, I., *et al.* Targeting EHMT2/ G9a for cancer therapy: Progress and perspective. *Eur. J. Pharmacol.* **893,** 173827 (2021).

126. Dobson, T. H. W., Hatcher, R. J., Swaminathan, J., Das, C. M., Shaik, S., *et al.* Regulation of USP37 Expression by REST-Associated G9a-Dependent Histone Methylation. *Mol. Cancer Res.* **15,** 1073–1084 (2017).

127. Burr, M. L., Sparbier, C. E., Chan, K. L., Chan, Y. C., Kersbergen, A., *et al.* An Evolutionarily Conserved Function of Polycomb Silences the MHC Class I Antigen Presentation Pathway and Enables Immune Evasion in Cancer. *Cancer Cell* **36,** 385–401.e8 (2019).

128. Peng, D., Kryczek, I., Nagarsheth, N., Zhao, L., Wei, S., *et al.* Epigenetic silencing of TH1-type chemokines shapes tumour immunity and immunotherapy. *Nature* **527,** 249–53 (2015).

129. Guler, G. D., Tindell, C. A., Pitti, R., Wilson, C., Nichols, K., *et al.* Repression of Stress-Induced LINE-1 Expression Protects Cancer Cell Subpopulations from Lethal Drug Exposure. *Cancer Cell* **32,** 221–237.e13 (2017).

130. Curry, E., Zeller, C., Masrour, N., Patten, D. K., Gallon, J., *et al.* Genes Predisposed to DNA Hypermethylation during Acquired Resistance to Chemotherapy Are Identified in Ovarian Tumors by Bivalent Chromatin Domains at Initial Diagnosis. *Cancer Res.* **78,** 1383–1391 (2018).

131. Terranova, C. J., Tang, M., Maitituoheti, M., Raman, A. T., Ghosh, A. K., *et al.* Reprogramming of bivalent chromatin states in NRAS mutant melanoma suggests PRC2 inhibition as a therapeutic strategy. *Cell Rep.* **36,** 109410 (2021).

132. Kim, E., Kim, M., Woo, D. H., Shin, Y., Shin, J., *et al.* Phosphorylation of EZH2 activates STAT3 signaling via STAT3 methylation and promotes tumorigenicity of glioblastoma stem-like cells. *Cancer Cell* **23,** 839–52 (2013).

133. Xu, K., Wu, Z. J., Groner, A. C., He, H. H., Cai, C., *et al.* EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. *Science* **338,** 1465–9 (2012).

134. Huang, J., Dorsey, J., Chuikov, S., Zhang, X., Jenuwein, T., *et al.* G9a and Glp methylate lysine 373 in the tumor suppressor p53. *J. Biol. Chem.* **285,** 9636–9641 (2010).

135. Gray, Z. H., Chakraborty, D., Duttweiler, R. R., Alekbaeva, G. D., Murphy, S. E., *et al.* Epigenetic balance ensures mechanistic control of MLL amplification and rearrangement. *Cell* **186,** 4528–4545.e18 (2023).

136. Yang, X., Xu, L. & Yang, L. Recent advances in EZH2-based dual inhibitors in the treatment of cancers. *Eur. J. Med. Chem.* **256,** 115461 (2023).

137. Ke, X. X., Zhang, R., Zhong, X., Zhang, L. & Cui, H. Deficiency of G9a Inhibits Cell Proliferation and Activates Autophagy via Transcriptionally Regulating c-Myc Expression in Glioblastoma. *Front. Cell Dev. Biol.* **8,** 593964 (2020).

138. Li, H., Cai, Q., Godwin, A. K. & Zhang, R. Enhancer of zeste homolog 2 promotes the proliferation and invasion of epithelial ovarian cancer cells. *Mol. Cancer Res.* **8,** 1610–8 (2010).

139. De Smedt, E., Devin, J., Muylaert, C., Robert, N., Requirand, G., *et al.* G9a/GLP targeting in MM promotes autophagy-associated apoptosis and boosts proteasome inhibitor-mediated cell death. *Blood Adv.* **5,** 2325–2338 (2021).

140. Yao, Y., Hu, H., Yang, Y., Zhou, G., Shang, Z., *et al.* Downregulation of Enhancer of Zeste Homolog 2 (EZH2) is essential for the Induction of Autophagy and Apoptosis in Colorectal Cancer Cells. *Genes (Basel)* **7,** 83 (2016).

141. Chibaya, L., Murphy, K. C., DeMarco, K. D., Gopalan, S., Liu, H., *et al.* EZH2 inhibition remodels the inflammatory senescence-associated secretory phenotype to potentiate pancreatic cancer immune surveillance. *Nat. Cancer* **4,** 872–892 (2023).

142. Mourits, V. P., van Puffelen, J. H., Novakovic, B., Bruno, M., Ferreira, A. V., *et al.* Lysine methyltransferase G9a is an important modulator of trained immunity. *Clin. Transl. Immunology* **10,** e1253 (2021).

143. U.S. Food & Drug Administration (FDA). *FDA approves first treatment option specifically for patients with epithelioid sarcoma, a rare soft tissue cancer.* Web Page, available at https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-option-specifically-patients-epithelioid-sarcoma-rare-soft-tissue. 2020.

144. U.S. Food & Drug Administration (FDA). *FDA granted accelerated approval to tazemetostat for follicular lymphoma.* Web Page, available at https://www.fda.gov/drugs/fda-granted-accelerated-approval-tazemetostat-follicular-lymphoma. 2020.

145. Li, B. & Chng, W. J. EZH2 abnormalities in lymphoid malignancies: underlying mechanisms and therapeutic implications. *J. Hematol. Oncol.* **12,** 118 (2019).

146. Yap, T. A., Winter, J. N., Giulino-Roth, L., Longley, J., Lopez, J., *et al.* Phase I Study of the Novel Enhancer of Zeste Homolog 2 (EZH2) Inhibitor GSK2816126 in Patients with Advanced Hematologic and Solid Tumors. *Clin. Cancer Res.* **25,** 7331–7339 (2019).

147. Liao, Q., Yang, J., Ge, S., Chai, P., Fan, J., *et al.* Novel insights into histone lysine methyltransferases in cancer therapy: From epigenetic regulation to selective drugs. *J. Pharm. Anal.* **13,** 127–141 (2023).

148. Cao, H., Li, L., Yang, D., Zeng, L., Yewei, X., *et al.* Recent progress in histone methyltransferase (G9a) inhibitors as anticancer agents. *Eur. J. Med. Chem.* **179,** 537–546 (2019).

149. Huang, R., Wu, Y. & Zou, Z. Combining EZH2 inhibitors with other therapies for solid tumors: more choices for better effects. *Epigenomics-uk.* **14,** 1449–1464 (2022).

150. Huang, X., Yan, J., Zhang, M., Wang, Y., Chen, Y., *et al.* Targeting Epigenetic Crosstalk as a Therapeutic Strategy for EZH2-Aberrant Solid Tumors. *Cell* **175,** 186–199.e19 (2018).

151. Zhang, L., Li, H. T., Shereda, R., Lu, Q., Weisenberger, D. J., *et al.* DNMT and EZH2 inhibitors synergize to activate therapeutic targets in hepatocellular carcinoma. *Cancer Lett.* **548,** 215899 (2022).

152. Liu, M., Thomas, S. L., DeWitt, A. K., Zhou, W., Madaj, Z. B., *et al.* Dual Inhibition of DNA and Histone Methyltransferases Increases Viral Mimicry in Ovarian Cancer Cells. *Cancer Res.* **78,** 5754–5766 (2018).

153. Ohtani, H., Liu, M., Zhou, W., Liang, G. & Jones, P. A. Switching roles for DNA and histone methylation depend on evolutionary ages of human endogenous retroviruses. *Genome Res.* **28,** 1147–1157 (2018).

154. Shi, Y., Zhang, Q., Zhang, M., Chen, Y., Sun, J., *et al.* Discovery of Dual Lysine Methyltransferase G9a and EZH2 Inhibitors with In Vivo Efficacy against Malignant Rhabdoid Tumor. *J. Med. Chem.* **66,** 5685–5702 (2023).

155. Spiliopoulou, P., Spear, S., Mirza, H., Garner, I., McGarry, L., *et al.* Dual G9A/EZH2 Inhibition Stimulates Antitumor Immune Response in Ovarian High-Grade Serous Carcinoma. *Mol. Cancer Ther.* **21,** 522–534 (2022).

156. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).

157. Wells, J. N. & Feschotte, C. A Field Guide to Eukaryotic Transposable Elements. *Annu. Rev. Genet.* **54,** 539–561 (2020).

158. Sotero-Caio, C. G., Platt Roy N., I., Suh, A. & Ray, D. A. Evolution and Diversity of Transposable Elements in Vertebrate Genomes. *Genome Biol. Evol.* **9,** 161–177 (2017).

159. Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., *et al.* Ten things you should know about transposable elements. *Genome Biol.* **19,** 199 (2018).

160. Fleury, H., Carmona, E., Morin, V. G., Meunier, L., Masson, J. Y., *et al.* Cumulative defects in DNA repair pathways drive the PARP inhibitor response in high-grade serous epithelial ovarian cancer cell lines. *Oncotarget* **8,** 40152–40168 (2017).

161. Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8,** 973–982 (2007).

162. Geis, F. K. & Goff, S. P. Silencing and Transcriptional Regulation of Endogenous Retroviruses: An Overview. *Viruses* **12,** 884 (2020).

163. Chinwalla, A. T., Cook, L. L., Delehaunty, K. D., Fewell, G. A., Fulton, L. A., *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).

164. Qu, Y., Izsvák, Z. & Wang, J. Retrotransposon: a versatile player in human preimplantation development and health. *Life Medicine* **2,** lnac041 (2023).

165. Vargiu, L., Rodriguez-Tomé, P., Sperber, G. O., Cadeddu, M., Grandi, N., *et al.* Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology* **13,** 7 (2016).

166. Zheng, J., Wei, Y. & Han, G.-Z. The diversity and evolution of retroviruses: Perspectives from viral "fossils". *Virol. Sin.* **37,** 11–18 (2022).

167. Van de Lagemaat, L. N., Medstrand, P. & Mager, D. L. Multiple effects govern endogenous retrovirus survival patterns in human gene introns. *Genome Biol.* **7,** R86 (2006).

168. Ayarpadikannan, S. & Kim, H. S. The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases. *Genomics Inform.* **12,** 98–104 (2014).

169. Fueyo, R., Judd, J., Feschotte, C. & Wysocka, J. Roles of transposable elements in the regulation of mammalian transcription. *Nat. Rev. Mol. Cell Biol.* **23,** 481–497 (2022).

170. Mills, R. E., Bennett, E. A., Iskow, R. C. & Devine, S. E. Which transposable elements are active in the human genome? *Trends Genet.* **23,** 183–91 (2007).

171. McClintock, B. Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* **21,** 197–216 (1956).

172. Biémont, C. A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics* **186,** 1085–93 (2010).

173. Senft, A. D. & Macfarlan, T. S. Transposable elements shape the evolution of mammalian development. *Nat. Rev. Genet.* **22,** 691–711 (2021).

174. Choudhary, M. N. K., Friedman, R. Z., Wang, J. T., Jang, H. S., Zhuo, X., *et al.* Co-opted transposons help perpetuate conserved higher-order chromosomal structures. *Genome Biol.* **21,** 16 (2020).

175. Diehl, A. G., Ouyang, N. & Boyle, A. P. Transposable elements contribute to cell and species-specific chromatin looping and gene regulation in mammalian genomes. *Nat. Commun.* **11,** 1796 (2020).

176. Lawson, H. A., Liang, Y. & Wang, T. Transposable elements in mammalian chromatin organization. *Nat. Rev. Genet.* **24,** 712–723 (2023).

177. Sexton, C. E., Tillett, R. L. & Han, M. V. The essential but enigmatic regulatory role of HERVH in pluripotency. *Trends Genet.* **38,** 12–21 (2022).

178. Zhang, Y., Li, T., Preissl, S., Amaral, M. L., Grinstein, J. D., *et al.* Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat. Genet.* **51,** 1380–1388 (2019).

179. Joly-Lopez, Z. & Bureau, T. E. Exaptation of transposable element coding sequences. *Curr. Opin. Genet. Dev.* **49,** 34–42 (2018).

180. Naville, M., Warren, I. A., Haftek-Terreau, Z., Chalopin, D., Brunet, F., *et al.* Not so bad after all: retroviruses and long terminal repeat retrotransposons as a source of new genes in vertebrates. *Clin. Microbiol. Infec.* **22,** 312–323 (2016).

181. Fort, V., Khelifi, G. & Hussein, S. M. I. Long non-coding RNAs and transposable elements: A functional relationship. *Biochim. Biophys. Acta. Mol. Cell Res.* **1868,** 118837 (2021).

182. Alan, Y. D., Jason, D. C., Xiaoyu, Z., Cédric, F. & Ting, W. Regulatory Transposable Elements in the Encyclopedia of DNA Elements. *bioRxiv,* doi: 10.1101/2023.09.05.556380 (2023).

183. Buttler, C. A. & Chuong, E. B. Emerging roles for endogenous retroviruses in immune epigenetic regulation. *Immunol. Rev.* **305,** 165–178 (2022).

184. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351,** 1083–7 (2016).

185. Simonti, C. N., Pavličev, M. & Capra, J. A. Transposable Element Exaptation into Regulatory Regions Is Rare, Influenced by Evolutionary Age, and Subject to Pleiotropic Constraints. *Mol. Biol. Evol.* **34,** 2856–2869 (2017).

186. Di Stefano, L. All Quiet on the TE Front? The Role of Chromatin in Transposable Element Silencing. *Cells* **11,** 2501 (2022).

187. Yoth, M., Jensen, S. & Brasset, E. The Intricate Evolutionary Balance between Transposable Elements and Their Host: Who Will Kick at Goal and Convert the Next Try? *Biology (Basel)* **11,** 710 (2022).

188. Huda, A., Mariño-Ramírez, L. & Jordan, I. K. Epigenetic histone modifications of human transposable elements: genome defense versus exaptation. *Mob. DNA* **1,** 2 (2010).

189. Almeida, M. V., Vernaz, G., Putman, A. L. K. & Miska, E. A. Taming transposable elements in vertebrates: from epigenetic silencing to domestication. *Trends Genet.* **38,** 529–553 (2022).

190. Deniz, Ö., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* **20,** 417–431 (2019).

191. Zhou, W., Liang, G., Molloy, P. L. & Jones, P. A. DNA methylation enables transposable element-driven genome expansion. *Proc. Natl. Acad. Sci.* **117,** 19359–19366 (2020).

192. Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev.* **24,** 265–76 (2010).

193. Liu, S., Brind'Amour, J., Karimi, M. M., Shirane, K., Bogutz, A., *et al.* Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev.* **28,** 2041–55 (2014).

194. Déléris, A., Berger, F. & Duharcourt, S. Role of Polycomb in the control of transposable elements. *Trends Genet.* **37,** 882–889 (2021).

195. Deblois, G., Tonekaboni, S. A. M., Grillo, G., Martinez, C., Kao, Y. I., *et al.* Epigenetic Switch-Induced Viral Mimicry Evasion in Chemotherapy-Resistant Breast Cancer. *Cancer Discov.* **10,** 1312–1329 (2020).

196. Hon, G. C., Hawkins, R. D., Caballero, O. L., Lo, C., Lister, R., *et al.* Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* **22,** 246–58 (2012).

197. Walter, M., Teissandier, A., Pérez-Palacios, R. & Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife* **5,** e11418 (2016).

198. Montavon, T., Shukeir, N., Erikson, G., Engist, B., Onishi-Seebacher, M., *et al.* Complete loss of H3K9 methylation dissolves mouse heterochromatin organization. *Nat. Commun.* **12,** 4359 (2021).

199. Maksakova, I. A., Thompson, P. J., Goyal, P., Jones, S. J., Singh, P. B., *et al.* Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenet. Chromatin* **6,** 15 (2013).

200. Kong, Y., Rose, C. M., Cass, A. A., Williams, A. G., Darwish, M., *et al.* Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. *Nat. Commun.* **10,** 5228 (2019).

201. Kassiotis, G. Endogenous retroviruses and the development of cancer. *J. Immunol.* **192,** 1343–9 (2014).

202. Jansz, N. & Faulkner, G. J. Endogenous retroviruses in the origins and treatment of cancer. *Genome Biol.* **22,** 147 (2021).

203. Anwar, S. L., Wulaningsih, W. & Lehmann, U. Transposable Elements in Human Cancer: Causes and Consequences of Deregulation. *Int. J. Mol. Sci.* **18,** 974 (2017).

204. Burns, K. H. Transposable elements in cancer. *Nat. Rev. Cancer* **17,** 415–424 (2017).

205. Karttunen, K., Patel, D., Xia, J., Fei, L., Palin, K., *et al.* Transposable elements as tissue-specific enhancers in cancers of endodermal lineage. *Nat. Commun.* **14,** 5313 (2023).

206. Bhat, A., Ghatage, T., Bhan, S., Lahane, G. P., Dhar, A., *et al.* Role of Transposable Elements in Genome Stability: Implications for Health and Disease. *Int. J. Mol. Sci.* **23,** 7802 (2022).

207. Zhu, X., Fang, H., Gladysz, K., Barbour, J. A. & Wong, J. W. H. Overexpression of transposable elements is associated with immune evasion and poor outcome in colorectal cancer. *Eur. J. Cancer* **157,** 94–107 (2021).

208. Grundy, E. E., Diab, N. & Chiappinelli, K. B. Transposable element regulation and expression in cancer. *FEBS J.* **289,** 1160–1179 (2022).

209. Teugels, E., De Brakeleer, S., Goelen, G., Lissens, W., Sermijn, E., *et al.* De novo Alu element insertions targeted to a sequence common to the BRCA1 and BRCA2 genes. *Hum. Mutat.* **26,** 284 (2005).

210. Jang, H. S., Shah, N. M., Du, A. Y., Dailey, Z. Z., Pehrsson, E. C., *et al.* Transposable elements drive widespread expression of oncogenes in human cancers. *Nat. Genet.* **51,** 611–617 (2019).

211. Deniz, Ö., Ahmed, M., Todd, C. D., Rio-Machin, A., Dawson, M. A., *et al.* Endogenous retroviruses are a source of enhancers with oncogenic potential in acute myeloid leukaemia. *Nat. Commun.* **11,** 3506 (2020).

212. Grillo, G., Keshavarzian, T., Linder, S., Arlidge, C., Mout, L., *et al.* Transposable Elements Are Co-opted as Oncogenic Regulatory Elements by Lineage-Specific Transcription Factors in Prostate Cancer. *Cancer Discov.* **13,** 2470–2487 (2023).

213. Chiappinelli, K. B., Strissel, P. L., Desrichard, A., Li, H., Henke, C., *et al.* Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* **162,** 974–86 (2015).

214. Natoli, M., Gallon, J., Lu, H., Amgheib, A., Pinato, D. J., *et al.* Transcriptional analysis of multiple ovarian cancer cohorts reveals prognostic and immunomodulatory consequences of ERV expression. *J. Immunother. Cancer* **9,** e001519 (2021).

215. Topham, J. T., Titmuss, E., Pleasance, E. D., Williamson, L. M., Karasinska, J. M., *et al.* Endogenous Retrovirus Transcript Levels Are Associated with Immunogenic Signatures in Multiple Metastatic Cancer Types. *Mol. Cancer Ther.* **19,** 1889 (2020).

216. Attig, J., Pape, J., Doglio, L., Kazachenka, A., Ottina, E., *et al.* Human endogenous retrovirus onco-exaptation counters cancer cell senescence through calbindin. *J. Clin. Invest.* **133** (2023).

217. Desai, N., Sajed, D., Arora, K. S., Solovyov, A., Rajurkar, M., *et al.* Diverse repetitive element RNA expression defines epigenetic and immunologic features of colon cancer. *JCI Insight* **2,** e91078 (2017).

218. Smith, C. C., Beckermann, K. E., Bortone, D. S., De Cubas, A. A., Bixby, L. M., *et al.* Endogenous retroviral signatures predict immunotherapy response in clear cell renal cell carcinoma. *J. Clin. Invest.* **128,** 4804–4820 (2018).

219. Sun, S., Qi, W., Rehrauer, H., Ronner, M., Hariharan, A., *et al.* Viral Mimicry Response Is Associated With Clinical Outcome in Pleural Mesothelioma. *JTO Clin. Res. Rep.* **3,** 100430 (2022).

220. Chen, R., Ishak, C. A. & De Carvalho, D. D. Endogenous Retroelements and the Viral Mimicry Response in Cancer Therapy and Cellular Homeostasis. *Cancer Discov.* **11,** 2707–2725 (2021).

221. Jones, P. A., Ohtani, H., Chakravarthy, A. & De Carvalho, D. D. Epigenetic therapy in immune-oncology. *Nat. Rev. Cancer* **19,** 151–161 (2019).

222. Gazquez-Gutierrez, A., Witteveldt, J., S, R. H. & Macias, S. Sensing of transposable elements by the antiviral innate immune system. *RNA* **27,** 735–52 (2021).

223. Goyal, A., Bauer, J., Hey, J., Papageorgiou, D. N., Stepanova, E., *et al.* DNMT and HDAC inhibition induces immunogenic neoantigens from human endogenous retroviral element-derived transcripts. *Nat. Commun.* **14,** 6731 (2023).

224. Mullins, C. S. & Linnebacher, M. Endogenous retrovirus sequences as a novel class of tumor-specific antigens: an example of HERV-H env encoding strong CTL epitopes. *Cancer Immunol. Immunother.* **61,** 1093–100 (2012).

225. Cuellar, T. L., Herzner, A. M., Zhang, X., Goyal, Y., Watanabe, C., *et al.* Silencing of retrotransposons by SETDB1 inhibits the interferon response in acute myeloid leukemia. *J. Cell Biol.* **216,** 3535–3549 (2017).

226. Fresquet, V., Garcia-Barchino, M. J., Larrayoz, M., Celay, J., Vicente, C., *et al.* Endogenous Retroelement Activation by Epigenetic Therapy Reverses the Warburg Effect and Elicits Mitochondrial-Mediated Cancer Cell Death. *Cancer Discov.* **11,** 1268 (2021).

227. Jin, L., He, J., Feng, H., Li, S., Liu, H., *et al.* Transposable elements activation triggers necroptosis in mouse embryonic stem cells. *Cell Death Discov.* **14,** 184 (2023).

228. Liu, M., Ohtani, H., Zhou, W., Ørskov, A. D., Charlet, J., *et al.* Vitamin C increases viral mimicry induced by 5-aza-2'-deoxycytidine. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 10238–44 (2016).

229. Roulois, D., Loo Yau, H., Singhania, R., Wang, Y., Danesh, A., *et al.* DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell* **162,** 961–973 (2015).

230. Alexandraki, A. & Strati, K. Decitabine Treatment Induces a Viral Mimicry Response in Cervical Cancer Cells and Further Sensitizes Cells to Chemotherapy. *Int. J. Mol. Sci.* **23,** 14042 (2022).

231. Brocks, D., Schmidt, C. R., Daskalakis, M., Jang, H. S., Shah, N. M., *et al.* DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat. Genet.* **49,** 1052–1060 (2017).

232. Lee, J. B., Khan, D. H., Hurren, R., Xu, M., Na, Y., *et al.* Venetoclax enhances T cell-mediated antileukemic activity by increasing ROS production. *Blood* **138,** 234–245 (2021).

233. Liu, J., Min, S., Kim, D., Park, J., Park, E., *et al.* Epigenetic priming improves salvage chemotherapy in diffuse large B-cell lymphoma via endogenous retrovirus-induced cGAS-STING activation. *Clin. Epigenetics* **15,** 75 (2023).

234. Mehdipour, P., Marhon, S. A., Ettayebi, I., Chakravarthy, A., Hosseini, A., *et al.* Epigenetic therapy induces transcription of inverted SINEs and ADAR1 dependency. *Nature* **588,** 169–173 (2020).

235. Bowling, E. A., Wang, J. H., Gong, F., Wu, W., Neill, N. J., *et al.* Spliceosome-targeted therapies trigger an antiviral immune response in triple-negative breast cancer. *Cell* **184,** 384–403.e21 (2021).

236. Rajurkar, M., Parikh, A. R., Solovyov, A., You, E., Kulkarni, A. S., *et al.* Reverse Transcriptase Inhibition Disrupts Repeat Element Life Cycle in Colorectal Cancer. *Cancer Discov.* **12,** 1462–1481 (2022).

237. Zhou, X., Singh, M., Sanz Santos, G., Guerlavais, V., Carvajal, L. A., *et al.* Pharmacologic Activation of p53 Triggers Viral Mimicry Response Thereby Abolishing Tumor Immune Evasion and Promoting Antitumor Immunity. *Cancer Discov.* **11,** 3090–3105 (2021).

238. Ghosh, R., Roy, S. & Franco, S. PARP1 depletion induces RIG-I-dependent signaling in human cancer cells. *PloS One* **13,** e0194611 (2018).

239. Liu, Y., Hu, L., Wu, Z., Yuan, K., Hong, G., *et al.* Loss of PHF8 induces a viral mimicry response by activating endogenous retrotransposons. *Nat. Commun.* **14,** 4225 (2023).

240. Sheng, W., LaFleur, M. W., Nguyen, T. H., Chen, S., Chakravarthy, A., *et al.* LSD1 Ablation Stimulates Anti-tumor Immunity and Enables Checkpoint Blockade. *Cell* **174,** 549–563.e19 (2018).

241. Cañadas, I., Thummalapalli, R., Kim, J. W., Kitajima, S., Jenkins, R. W., *et al.* Tumor innate immunity primed by specific interferon-stimulated endogenous retroviruses. *Nat. Med.* **24,** 1143–1150 (2018).

242. Kang, Z., Fu, P., Ma, H., Li, T., Lu, K., *et al.* Distinct functions of EHMT1 and EHMT2 in cancer chemotherapy and immunotherapy. *bioRxiv,* doi: 10.1101/2023.10.03.560719 (2023).

243. Morel, K. L., Sheahan, A. V., Burkhart, D. L., Baca, S. C., Boufaied, N., *et al.* EZH2 inhibition activates a dsRNA–STING–interferon stress axis that potentiates response to PD-1 checkpoint blockade in prostate cancer. *Nat. Cancer* **2,** 444–456 (2021).

244. Pan, D., Bao, X., Hu, M., Jiao, M., Li, F., *et al.* SETDB1 Restrains Endogenous Retrovirus Expression and Antitumor Immunity during Radiotherapy. *Cancer Res.* **82,** 2748–2760 (2022).

245. Shen, J. Z. & Spruck, C. Targeting FBXO44/SUV39H1 elicits tumor cell-specific DNA replication stress and viral mimicry. *Cell Stress* **5,** 37–39 (2021).

246. Zhang, L., Jin, J., Qin, W., Jiang, J., Bao, W., *et al.* Inhibition of EZH2 Causes Retrotransposon Derepression and Immune Activation in Porcine Lung Alveolar Macrophages. *Int. J. Mol. Sci.* **24,** 2394 (2023).

247. Bolis, M., Paroni, G., Fratelli, M., Vallerga, A., Guarrera, L., *et al.* All-Trans Retinoic Acid Stimulates Viral Mimicry, Interferon Responses and Antigen Presentation in Breast-Cancer Cells. *Cancers (Basel)* **12,** 1169 (2020).

248. Ghosh, M., Saha, S., Bettke, J., Nagar, R., Parrales, A., *et al.* Mutant p53 suppresses innate immune signaling to promote tumorigenesis. *Cancer Cell* **39,** 494–508.e5 (2021).

249. Fang, T. C., Schaefer, U., Mecklenbrauker, I., Stienen, A., Dewell, S., *et al.* Histone H3 lysine 9 di-methylation as an epigenetic signature of the interferon response. *J. Exp. Med.* **209,** 661–9 (2012).

250. Wee, Z., Li, Z., Lee, P., Lee, S., Lim, Y., *et al.* EZH2-Mediated Inactivation of IFN-γ-JAK-STAT1 Signaling Is an Effective Therapeutic Target in MYC-Driven Prostate Cancer. *Cell Rep.* **8,** 204–216 (2014).

251. Vitiello, G. A. F., Ferreira, W. A. S., Cordeiro de Lima, V. C. & Medina, T. D. S. Antiviral Responses in Cancer: Boosting Antitumor Immunity Through Activation of Interferon Pathway in the Tumor Microenvironment. *Front. Immunol.* **12,** 782852 (2021).

252. Ishak, C. A., Classon, M. & De Carvalho, D. D. Deregulation of Retroelements as an Emerging Therapeutic Opportunity in Cancer. *Trends Cancer* **4,** 583–597 (2018).

253. Li, H. T., Jang, H. J., Rohena-Rivera, K., Liu, M., Gujar, H., *et al.* RNA mis-splicing drives viral mimicry response after DNMTi therapy in SETD2-mutant kidney cancer. *Cell Rep.* **42,** 112016 (2023).

254. Patel, A. J., Warda, S., Maag, J. L. V., Misra, R., Miranda-Román, M. A., *et al.* PRC2-Inactivating Mutations in Cancer Enhance Cytotoxic Response to DNMT1-Targeted Therapy via Enhanced Viral Mimicry. *Cancer Discov.* **12,** 2120–2139 (2022).

255. Scheller, M., Ludwig, A. K., Göllner, S., Rohde, C., Krämer, S., *et al.* Hotspot DNMT3A mutations in clonal hematopoiesis and acute myeloid leukemia sensitize cells to azacytidine via viral mimicry response. *Nat. Cancer* **2,** 527–544 (2021).

256. Zhou, X., Sekino, Y., Li, H. T., Fu, G., Yang, Z., *et al.* SETD2 Deficiency Confers Sensitivity to Dual Inhibition of DNA Methylation and PARP in Kidney Cancer. *Cancer Res.* **83,** 3813–3826 (2023).

257. Madden, S. K., de Araujo, A. D., Gerhardt, M., Fairlie, D. P. & Mason, J. M. Taking the Myc out of cancer: toward therapeutic strategies to directly inhibit c-Myc. *Mol. Cancer* **20,** 3 (2021).

258. Poole, C. J. & van Riggelen, J. MYC-Master Regulator of the Cancer Epigenome and Transcriptome. *Genes (Basel)* **8,** 142 (2017).

259. Duffy, M. J., O'Grady, S., Tang, M. & Crown, J. MYC as a target for cancer treatment. *Cancer Treat. Rev.* **94,** 102154 (2021).

260. Hanahan, D. & Robert. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674 (2011).

261. Llombart, V. & Mansour, M. R. Therapeutic targeting of "undruggable" MYC. *eBioMedicine* **75,** 103756 (2022).

262. Blackwell, T. K., Huang, J., Ma, A., Kretzner, L., Alt, F. W., *et al.* Binding of myc proteins to canonical and noncanonical DNA sequences. *Mol. Cell Biol.* **13,** 5216–24 (1993).

263. Pelengaris, S., Khan, M. & Evan, G. c-MYC: more than just a matter of life and death. *Nat. Rev. Cancer* **2,** 764–776 (2002).

264. Murphy, D. J., Junttila, M. R., Pouyet, L., Karnezis, A., Shchors, K., *et al.* Distinct thresholds govern Myc's biological output in vivo. *Cancer Cell* **14,** 447–57 (2008).

265. Ahmadi, S. E., Rahimi, S., Zarandi, B., Chegeni, R. & Safa, M. MYC: a multipurpose oncogene with prognostic and therapeutic implications in blood malignancies. *J. Hematol. Oncol.* **14,** 121 (2021).

266. Herkert, B. & Eilers, M. Transcriptional repression: the dark side of myc. *Genes Cancer* **1,** 580–6 (2010).

267. Clavería-Cabello, A., Herranz, J. M., Latasa, M. U., Arechederra, M., Uriarte, I., *et al.* Identification and experimental validation of druggable epigenetic targets in hepatoblastoma. *J. Hepatol.* **79,** 989–1005 (2023).

268. Tu, W. B., Shiah, Y. J., Lourenco, C., Mullen, P. J., Dingar, D., *et al.* MYC Interacts with the G9a Histone Methyltransferase to Drive Transcriptional Repression and Tumorigenesis. *Cancer Cell* **34,** 579–595.e8 (2018).

269. Alburquerque-Bejar, J. J., Navajas-Chocarro, P., Saigi, M., Ferrero-Andres, A., Morillas, J. M., *et al.* MYC activation impairs cell-intrinsic IFNγ signaling and confers resistance to anti-PD1/PD-L1 therapy in lung cancer. *Cell Rep. Med.* **4,** 101006 (2023).

270. Dhanasekaran, R., Hansen, A. S., Park, J., Lemaitre, L., Lai, I., *et al.* MYC Overexpression Drives Immune Evasion in Hepatocellular Carcinoma That Is Reversible through Restoration of Proinflammatory Macrophages. *Cancer Res.* **83,** 626–640 (2023).

271. Kim, T. W., Hong, S., Lin, Y., Murat, E., Joo, H., *et al.* Transcriptional Repression of IFN Regulatory Factor 7 by MYC Is Critical for Type I IFN Production in Human Plasmacytoid Dendritic Cells. *J. Immunol.* **197,** 3348–3359 (2016).

272. Lee, K. M., Lin, C. C., Servetto, A., Bae, J., Kandagatla, V., *et al.* Epigenetic Repression of STING by MYC Promotes Immune Evasion and Resistance to Immune Checkpoint Inhibitors in Triple-Negative Breast Cancer. *Cancer Immunol. Res.* **10,** 829–843 (2022).

273. Markovits, E., Harush, O., Baruch, E. N., Shulman, E. D., Debby, A., *et al.* MYC induces immunotherapy and interferon-γ resistance through downregulation of JAK2. *Cancer Immunol. Res.* **11,** 909–924 (2023).

274. Muthalagu, N., Monteverde, T., Raffo-Iraolagoitia, X., Wiesheu, R., Whyte, D., *et al.* Repression of the Type I Interferon Pathway Underlies MYC- and KRAS-Dependent Evasion of NK and B Cells in Pancreatic Ductal Adenocarcinoma. *Cancer Discov.* **10,** 872–887 (2020).

275. Schlee, M., Hölzel, M., Bernard, S., Mailhammer, R., Schuhmacher, M., *et al.* C-myc activation impairs the NF-kappaB and the interferon response: implications for the pathogenesis of Burkitt's lymphoma. *Int. J. Cancer* **120,** 1387–95 (2007).

276. Si-Yu, W., Yi, X., Jin-Li, W., Xiao-En, X., Xi, J., *et al.* MYC suppresses STING-dependent innate immunity by transcriptionally upregulating DNMT1 in triple-negative breast cancer. *J. Immunother. Cancer* **9,** e002528 (2021).

277. Swaminathan, S., Hansen, A. S., Heftdal, L. D., Dhanasekaran, R., Deutzmann, A., *et al.* MYC functions as a switch for natural killer cell-mediated immune surveillance of lymphoid malignancies. *Nat. Commun.* **11,** 2860 (2020).

278. Zimmerli, D., Brambillasca, C. S., Talens, F., Bhin, J., Linstra, R., *et al.* MYC promotes immune-suppression in triple-negative breast cancer via inhibition of interferon signaling. *Nat. Commun.* **13,** 6579 (2022).

279. Dhanasekaran, R., Deutzmann, A., Mahauad-Fernandez, W. D., Hansen, A. S., Gouw, A. M., *et al.* The MYC oncogene - the grand orchestrator of cancer growth and immune evasion. *Nat. Rev. Clin. Oncol.* **19,** 23–36 (2022).

280. Casey, S. C., Tong, L., Li, Y., Do, R., Walz, S., *et al.* MYC regulates the antitumor immune response through CD47 and PD-L1. *Science* **352,** 227–31 (2016).

281. Wiese, K. E., Walz, S., von Eyss, B., Wolf, E., Athineos, D., *et al.* The role of MIZ-1 in MYC-dependent tumorigenesis. *Cold Spring Harb. Perspect. Med.* **3,** a014290 (2013).

282. Walz, S., Lorenzin, F., Morton, J., Wiese, K. E., von Eyss, B., *et al.* Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles. *Nature* **511,** 483–7 (2014).

283. Thng, D. K. H., Hooi, L., Toh, C. C. M., Lim, J. J., Rajagopalan, D., *et al.* Histone-lysine N-methyltransferase EHMT2 (G9a) inhibition mitigates tumorigenicity in Myc-driven liver cancer. *Mol. Oncol.* **17,** 2275–2294 (2023).

284. Ali, A., Shafarin, J., Unnikannan, H., Al-Jabi, N., Jabal, R. A., *et al.* Co-targeting BET bromodomain BRD4 and RAC1 suppresses growth, stemness and tumorigenesis by disrupting the c-MYC-G9a-FTH1axis and downregulating HDAC1 in molecular subtypes of breast cancer. *Int. J. Biol. Sci.* **17,** 4474–4492 (2021).

285. Liu, C., Yu, Y., Liu, F., Wei, X., Wrobel, J. A., *et al.* A chromatin activity-based chemoproteomic approach reveals a transcriptional repressome for gene-specific silencing. *Nat. Commun.* **5,** 5733 (2014).

286. Zhang, J., Chen, W., Ma, W., Song, K., Lee, S., *et al.* Epigenetic Silencing of 15-Hydroxyprostaglandin Dehydrogenase by Histone Methyltransferase EHMT2/G9a in Cholangiocarcinoma. *Mol. Cancer Res.* **20,** 350–360 (2022).

287. Alzrigat, M., Párraga, A. A., Agarwal, P., Zureigat, H., Österborg, A., *et al.* EZH2 inhibition in multiple myeloma downregulates myeloma associated oncogenes and upregulates microRNAs with potential tumor suppressor functions. *Oncotarget* **8,** 10213–10224 (2017).

288. Sun, Y., Fan, J., Wang, B., Meng, Z., Ren, D., *et al.* The aberrant expression of ADAR1 promotes resistance to BET inhibitors in pancreatic cancer by stabilizing c-Myc. *Am. J. Cancer Res.* **10,** 148–163 (2020).

289. Suvà, M. L., Riggi, N., Janiszewska, M., Radovanovic, I., Provero, P., *et al.* EZH2 is essential for glioblastoma cancer stem cell maintenance. *Cancer Res.* **69,** 9211–8 (2009).

290. Ito, T., Teo, Y. V., Evans, S. A., Neretti, N. & Sedivy, J. M. Regulation of Cellular Senescence by Polycomb Chromatin Modifiers through Distinct DNA Damage- and Histone Methylation-Dependent Pathways. *Cell Rep.* **22,** 3480–3492 (2018).

291. Salvatori, B., Iosue, I., Djodji Damas, N., Mangiavacchi, A., Chiaretti, S., *et al.* Critical Role of c-Myc in Acute Myeloid Leukemia Involving Direct Regulation of miR-26a and Histone Methyltransferase EZH2. *Genes Cancer* **2,** 585–92 (2011).

292. Wang, L., Chen, C., Song, Z., Wang, H., Ye, M., *et al.* EZH2 depletion potentiates MYC degradation inhibiting neuroblastoma and small cell carcinoma tumor formation. *Nat. Commun.* **13,** 12 (2022).

293. Ding, N., You, A., Zhao, S., Yang, H., Lai, C., *et al.* EZH2 inhibitor Tazemetostat synergizes with JQ-1 in esophageal cancer by inhibiting c-Myc signaling pathway. *Med. Oncol.* **40,** 281 (2023).

294. Negri, A., Marozzi, M., Trisciuoglio, D., Rotili, D., Mai, A., *et al.* Simultaneous administration of EZH2 and BET inhibitors inhibits proliferation and clonogenic ability of metastatic prostate cancer cells. *J. Enzyme Inhib. Med. Chem.* **38,** 2163242 (2023).

295. Neri, F., Zippo, A., Krepelova, A., Cherubini, A., Rocchigiani, M., *et al.* Myc regulates the transcription of the PRC2 gene to control the expression of developmental genes in embryonic stem cells. *Mol. Cell Biol.* **32,** 840–51 (2012).

296. Wu, X., Liu, D., Tao, D., Xiang, W., Xiao, X., *et al.* BRD4 Regulates EZH2 Transcription through Upregulation of C-MYC and Represents a Novel Therapeutic Target in Bladder Cancer. *Mol. Cancer Ther.* **15,** 1029–42 (2016).

297. Sun, J., Cai, X., Yung, M. M., Zhou, W., Li, J., *et al.* miR-137 mediates the functional link between c-Myc and EZH2 that regulates cisplatin resistance in ovarian cancer. *Oncogene* **38,** 564–580 (2019).

298. Qu, Y., Yang, Q., Liu, J., Shi, B., Ji, M., *et al.* c-Myc is Required for BRAF(V600E)-Induced Epigenetic Silencing by H3K27me3 in Tumorigenesis. *Theranostics* **7,** 2092–2107 (2017).

299. Wang, L., Zhang, X., Jia, L. T., Hu, S. J., Zhao, J., *et al.* c-Myc-mediated epigenetic silencing of MicroRNA-101 contributes to dysregulation of multiple pathways in hepatocellular carcinoma. *Hepatology* **59,** 1850–63 (2014).

300. Wang, J., Yu, X., Gong, W., Liu, X., Park, K. S., *et al.* EZH2 noncanonically binds cMyc and p300 through a cryptic transactivation domain to mediate gene activation and promote oncogenesis. *Nat. Cell Biol.* **24,** 384–399 (2022).

301. Yu, X., Wang, J., Gong, W., Ma, A., Shen, Y., *et al.* Dissecting and targeting noncanonical functions of EZH2 in multiple myeloma via an EZH2 degrader. *Oncogene* **42,** 994–1009 (2023).

302. Zhao, X., Lwin, T., Zhang, X., Huang, A., Wang, J., *et al.* Disruption of the MYC-miRNA-EZH2 loop to suppress aggressive B-cell lymphoma survival and clonogenicity. *Leukemia* **27,** 2341–50 (2013).

303. Xiao, J. F., Sun, Q. Y., Ding, L. W., Chien, W., Liu, X. Y., *et al.* The c-MYC-BMI1 axis is essential for SETDB1-mediated breast tumourigenesis. *J. Pathol.* **246,** 89–102 (2018).

304. Wu, M. J., Chen, C. J., Lin, T. Y., Liu, Y. Y., Tseng, L. L., *et al.* Targeting KDM4B that coactivates c-Myc-regulated metabolism to suppress tumor growth in castration-resistant prostate cancer. *Theranostics* **11,** 7779–7796 (2021).

305. Zhao, P., Sun, X., Li, H., Liu, Y., Cui, Y., *et al.* c-Myc Targets HDAC3 to Suppress NKG2DL Expression and Innate Immune Response in N-Type SCLC through Histone Deacetylation. *Cancers (Basel)* **14,** 457 (2022).

306. See, Y. X., Chen, K. & Fullwood, M. J. MYC overexpression leads to increased chromatin interactions at super-enhancers and MYC binding sites. *Genome Res.* **32,** 629–642 (2022).

307. Faiola, F., Liu, X., Lo, S., Pan, S., Zhang, K., *et al.* Dual regulation of c-Myc by p300 via acetylation-dependent control of Myc protein turnover and coactivation of Myc-induced transcription. *Mol. Cell Biol.* **25,** 10220–34 (2005).

308. Wang, C., Zhang, J., Yin, J., Gan, Y., Xu, S., *et al.* Alternative approaches to target Myc for cancer treatment. *Signal Transduct. Target. Ther.* **6,** 117 (2021).

309. Soucek, L., Whitfield, J., Martins, C. P., Finch, A. J., Murphy, D. J., *et al.* Modelling Myc inhibition as a cancer therapy. *Nature* **455,** 679–83 (2008).

310. Han, H., Jain, A. D., Truica, M. I., Izquierdo-Ferrer, J., Anker, J. F., *et al.* Small-Molecule MYC Inhibitors Suppress Tumor Growth and Enhance Immunotherapy. *Cancer Cell* **36,** 483–497.e15 (2019).

311. Topper, M. J., Vaz, M., Chiappinelli, K. B., DeStefano Shields, C. E., Niknafs, N., *et al.* Epigenetic Therapy Ties MYC Depletion to Reversing Immune Evasion and Treating Lung Cancer. *Cell* **171,** 1284–1300.e21 (2017).

312. Chen, Y. G. & Hur, S. Cellular origins of dsRNA, their recognition and consequences. *Nat. Rev. Mol. Cell Biol.* **23,** 286–301 (2022).

313. Kandalaft, L. E., Dangaj Laniti, D. & Coukos, G. Immunobiology of high-grade serous ovarian cancer: lessons for clinical translation. *Nat. Rev. Cancer* **22,** 640–656 (2022).

314. Berns, E. M. J. J. & Bowtell, D. D. The Changing View of High-Grade Serous Ovarian Cancer. *Cancer Res.* **72,** 2701–2704 (2012).

315. Lisio, M. A., Fu, L., Goyeneche, A., Gao, Z. H. & Telleria, C. High-Grade Serous Ovarian Cancer: Basic Sciences, Clinical and Therapeutic Standpoints. *Int. J. Mol. Sci.* **20,** 952 (2019).

316. Talbot, T., Lu, H. & Aboagye, E. O. Amplified therapeutic targets in high-grade serous ovarian carcinoma – a review of the literature with quantitative appraisal. *Cancer Gene Ther.* **30,** 955–963 (2023).

317. Bhattacharya, R., Ghosh, A. & Mukhopadhyay, S. High-grade serous ovarian carcinoma, the "Achiles' hill" for clinicians and molecular biologists: a molecular insight. *Mol. Biol. Rep.* **50,** 9511–9519 (2023).

318. Ahmed, A. A., Etemadmoghadam, D., Temple, J., Lynch, A. G., Riad, M., *et al.* Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. *J. Pathol.* **221,** 49–56 (2010).

319. Patch, A.-M., Christie, E. L., Etemadmoghadam, D., Garsed, D. W., George, J., *et al.* Whole–genome characterization of chemoresistant ovarian cancer. *Nature* **521,** 489–494 (2015).

320. Garsed, D. W., Pandey, A., Fereday, S., Kennedy, C. J., Takahashi, K., *et al.* The genomic and immune landscape of long-term survivors of high-grade serous ovarian cancer. *Nat. Genet.* **54,** 1853–1864 (2022).

321. Jones, B. A., Varambally, S. & Arend, R. C. Histone Methyltransferase EZH2: A Therapeutic Target for Ovarian Cancer. *Mol. Cancer Ther.* **17,** 591–602 (2018).

322. Rao, Z.-Y., Cai, M.-Y., Yang, G.-F., He, L.-R., Mai, S.-J., *et al.* EZH2 supports ovarian carcinoma cell invasion and/or metastasis via regulation of TGF-β1 and is a predictor of outcome in ovarian carcinoma patients. *Carcinogenesis* **31,** 1576–1583 (2010).

323. Hua, K. T., Wang, M. Y., Chen, M. W., Wei, L. H., Chen, C. K., *et al.* The H3K9 methyltransferase G9a is a marker of aggressive ovarian cancer that promotes peritoneal metastasis. *Mol. Cancer* **13,** 189 (2014).

324. Kang, J., Shin, S. H., Yoon, H., Huh, J., Shin, H. W., *et al.* FIH Is an Oxygen Sensor in Ovarian Cancer for G9a/GLP-Driven Epigenetic Regulation of Metastasis-Related Genes. *Cancer Res.* **78,** 1184–1199 (2018).

325. Hu, S., Yu, L., Li, Z., Shen, Y., Wang, J., *et al.* Overexpression of EZH2 contributes to acquired cisplatin resistance in ovarian cancer cells in vitro and in vivo. *Cancer Biol. Ther.* **10,** 788–95 (2010).

326. Naskou, J., Beiter, Y., van Rensburg, R., Honisch, E., Rudelius, M., *et al.* EZH2 Loss Drives Resistance to Carboplatin and Paclitaxel in Serous Ovarian Cancers Expressing ATM. *Mol. Cancer Res.* **18,** 278–286 (2020).

327. Sun, Y., Jin, L., Liu, J. H., Sui, Y. X., Han, L. L., *et al.* Interfering EZH2 Expression Reverses the Cisplatin Resistance in Human Ovarian Cancer by Inhibiting Autophagy. *Cancer Biother. Radiopharm.* **31,** 246–52 (2016).

328. Chapman-Rothe, N., Curry, E., Zeller, C., Liber, D., Stronach, E., *et al.* Chromatin H3K27me3/H3K4me3 histone marks define gene sets in high-grade serous ovarian cancer that distinguish malignant, tumour-sustaining and chemo-resistant ovarian tumour cells. *Oncogene* **32,** 4586–4592 (2013).

329. Wen, Y., Hou, Y., Yi, X., Sun, S., Guo, J., *et al.* EZH2 activates CHK1 signaling to promote ovarian cancer chemoresistance by maintaining the properties of cancer stem cells. *Theranostics* **11,** 1795–1813 (2021).

330. Zong, X., Wang, W., Ozes, A., Fang, F., Sandusky, G. E., *et al.* EZH2-Mediated Downregulation of the Tumor Suppressor DAB2IP Maintains Ovarian Cancer Stem Cells. *Cancer Res.* **80,** 4371–4385 (2020).

331. Chen, J., Hong, J. H., Huang, Y., Liu, S., Yin, J., *et al.* EZH2 mediated metabolic rewiring promotes tumor growth independently of histone methyltransferase activity in ovarian cancer. *Mol. Cancer* **22,** 85 (2023).

332. Moufarrij, S., Dandapani, M., Arthofer, E., Gomez, S., Srivastava, A., *et al.* Epigenetic therapy for ovarian cancer: promise and progress. *Clin. Epigenetics* **11,** 7 (2019).

333. Bradbury, M., Borràs, E., Vilar, M., Castellví, J., Sánchez-Iglesias, J. L., *et al.* A combination of molecular and clinical parameters provides a new strategy for high-grade serous ovarian cancer patient management. *J. Transl. Med.* **20,** 611 (2022).

334. Colombo, P.-E., Fabbro, M., Theillet, C., Bibeau, F., Rouanet, P., *et al.* Sensitivity and resistance to treatment in the primary management of epithelial ovarian cancer. *Crit. Rev. Oncol. Hematol.* **89,** 207–216 (2014).

335. González-Martín, A., Harter, P., Leary, A., Lorusso, D., Miller, R. E., *et al.* Newly diagnosed and relapsed epithelial ovarian cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann. Oncol.* **34,** 833–848 (2023).

336. Colombo, N., Sessa, C., du Bois, A., Ledermann, J., McCluggage, W. G., *et al.* ESMO-ESGO consensus conference recommendations on ovarian cancer: pathology and molecular biology, early and advanced stages, borderline tumours and recurrent disease. *Ann. Oncol.* **30,** 672–705 (2019).

337. Davidson, B. & Elstrand, M. B. Clinicopathological prognostic parameters in patients with tubo-ovarian carcinoma effusions. *Cytopathology* **33,** 479–492 (2022).

338. Pennington, K. P., Walsh, T., Harrell, M. I., Lee, M. K., Pennil, C. C., *et al.* Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. *Clin. Cancer Res.* **20,** 764–75 (2014).

339. Su, R., Liu, Y., Wu, X., Xiang, J. & Xi, X. Dynamically Accumulating Homologous Recombination Deficiency Score Served as an Important Prognosis Factor in High-Grade Serous Ovarian Cancer. *Front. Mol. Biosci.* **8,** 762741 (2021).

340. Chen, G. M., Kannan, L., Geistlinger, L., Kofia, V., Safikhani, Z., *et al.* Consensus on Molecular Subtypes of High-Grade Serous Ovarian Carcinoma. *Clin. Cancer Res.* **24,** 5037–5047 (2018).

341. Hwang, W. T., Adams, S. F., Tahirovic, E., Hagemann, I. S. & Coukos, G. Prognostic significance of tumor-infiltrating T cells in ovarian cancer: a meta-analysis. *Gynecol. Oncol.* **124,** 192–8 (2012).

342. (OTTA), O. T. T. A. C. Dose-Response Association of CD8+ Tumor-Infiltrating Lymphocytes and Survival Time in High-Grade Serous Ovarian Cancer. *JAMA Oncology* **3,** e173290–e173290 (2017).

343. Hoogstad-van Evert, J. S., Bekkers, R., Ottevanger, N., Jansen, J. H., Massuger, L., *et al.* Harnessing natural killer cells for the treatment of ovarian cancer. *Gynecol. Oncol.* **157,** 810–816 (2020).

344. Nielsen, Julie S and Sahota, Rob A and Milne, Katy and Kost, Sara E and Nesslinger, Nancy J and Watson, Peter H and Nelson, Brad H. CD20+ tumor-infiltrating lymphocytes have an atypical CD27- memory phenotype and together with CD8+ T cells promote favorable prognosis in ovarian cancer. *Clin. Cancer Res.* **18,** 3281–3292 (2012).

345. Ghaffari, A., Peterson, N., Khalaj, K., Vitkin, N., Robinson, A., *et al.* STING agonist therapy in combination with PD-1 immune checkpoint blockade enhances response to carboplatin chemotherapy in high-grade serous ovarian cancer. *Br. J. Cancer* **119,** 440–449 (2018).

346. Dangaj, D., Bruand, M., Grimm, A. J., Ronet, C., Barras, D., *et al.* Cooperation between Constitutive and Inducible Chemokines Enables T Cell Engraftment and Immune Attack in Solid Tumors. *Cancer Cell* **35,** 885–900.e10 (2019).

347. Corrales, L., Glickman, L., McWhirter, S., Kanne, D., Sivick, K., *et al.* Direct Activation of STING in the Tumor Microenvironment Leads to Potent and Systemic Tumor Regression and Immunity. *Cell Rep.* **11,** 1018–1030 (2015).

348. Ding, L., Kim, H.-J., Wang, Q., Kearns, M., Jiang, T., *et al.* PARP Inhibition Elicits STING-Dependent Antitumor Immunity in Brca1-Deficient Ovarian Cancer. *Cell Rep.* **25,** 2972–2980.e5 (2018).

349. Woo, S.-R., Fuertes, M., Corrales, L., Spranger, S., Furdyna, M., *et al.* STING-Dependent Cytosolic DNA Sensing Mediates Innate Immune Recognition of Immunogenic Tumors. *Immunity* **41,** 830–842 (2014).

350. Au, K. K., Le Page, C., Ren, R., Meunier, L., Clément, I., *et al.* STAT1-associated intratumoural T(H)1 immunity predicts chemotherapy resistance in high-grade serous ovarian cancer. *J. Pathol. Clin. Res.* **2,** 259–270 (2016).

351. Bruand, M., Barras, D., Mina, M., Ghisoni, E., Morotti, M., *et al.* Cell-autonomous inflammation of *BRCA1*-deficient ovarian cancers drives both tumor-intrinsic immunoreactivity and immune resistance *via* STING. *Cell Rep.* **36,** 109412 (2021).

352. Stone, M. L., Chiappinelli, K. B., Li, H., Murphy, L. M., Travers, M. E., *et al.* Epigenetic therapy activates type I interferon signaling in murine ovarian cancer to reduce immunosuppression and tumor burden. *Proc. Natl. Acad. Sci. U. S. A.* **114,** E10981–E10990 (2017).

353. Liu, T., Li, Y., Wang, X., Yang, X., Fu, Y., *et al.* The role of interferons in ovarian cancer progression: Hinderer or promoter? *Front. Immunol.* **13,** 1087620 (2022).

354. Shakfa, N., Li, D., Conseil, G., Lightbody, E. D., Wilson-Sanchez, J., *et al.* Cancer cell genotype associated tumor immune microenvironment exhibits differential response to therapeutic STING pathway activation in high-grade serous ovarian cancer. *J. Immunother. Cancer* **11,** e006170 (2023).

355. Herbst, R. S., Morgensztern, D. & Boshoff, C. The biology and management of non-small cell lung cancer. *Nature* **553,** 446–454 (2018).

356. Hoang, P. H. & Landi, M. T. DNA Methylation in Lung Cancer: Mechanisms and Associations with Histological Subtypes, Molecular Alterations, and Major Epidemiological Factors. *Cancers (Basel)* **14,** 961 (2022).

357. Chalela, R., Curull, V., Enríquez, C., Pijuan, L., Bellosillo, B., *et al.* Lung adenocarcinoma: from molecular basis to genome-guided therapy and immunotherapy. *J. Thorac. Dis.* **9,** 2142–2158 (2017).

358. Collisson, E. A., Campbell, J. D., Brooks, A. N., Berger, A. H., Lee, W., *et al.* Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511,** 543–550 (2014).

359. Wallbillich, N. J. & Lu, H. Role of c-Myc in lung cancer: Progress, challenges, and prospects. *Chin. Med. J. Pulmonary and Crit. Care Med.* **1,** 129–138 (2023).

360. Seo, A. N., Yang, J. M., Kim, H., Jheon, S., Kim, K., *et al.* Clinicopathologic and prognostic significance of c-MYC copy number gain in lung adenocarcinomas. *Br. J. Cancer* **110,** 2688–99 (2014).

361. Kim, E. Y., Kim, A., Kim, S. K. & Chang, Y. S. MYC expression correlates with PD-L1 expression in non-small cell lung cancer. *Lung Cancer* **110,** 63–67 (2017).

362. Kortlever, R. M., Sodir, N. M., Wilson, C. H., Burkhart, D. L., Pellegrinet, L., *et al.* Myc Cooperates with Ras by Programming Inflammation and Immune Suppression. *Cell* **171,** 1301–1315.e14 (2017).

363. Mugarza, E., van Maldegem, F., Boumelha, J., Moore, C., Rana, S., *et al.* Therapeutic KRAS(G12C) inhibition drives effective interferon-mediated antitumor immunity in immunogenic lung cancers. *Sci. Adv.* **8,** eabm8780 (2022).

364. Lopez, J. S. & Banerji, U. Combine and conquer: challenges for targeted therapy combinations in early phase trials. *Nat. Rev. Clin. Oncol.* **14,** 57–66 (2017).

365. Kubicek, S., O'Sullivan, R. J., August, E. M., Hickey, E. R., Zhang, Q., *et al.* Reversal of H3K9me2 by a Small-Molecule Inhibitor for the G9a Histone Methyltransferase. *Mol. Cell* **25,** 473–481 (2007).

366. R Core Team. *R: A Language and Environment for Statistical Computing.* Available at https://www.R-project.org/. R Foundation for Statistical Computing (Vienna, Austria, 2023).

367. RStudio Team. *RStudio: Integrated Development Environment for R.* Available at http://www.rstudio.com/. RStudio, PBC. (Boston, MA, 2020).

368. Langdon, S. P., Lawrie, S. S., Hay, F. G., Hawkes, M. M., McDonald, A., *et al.* Characterization and properties of nine human ovarian adenocarcinoma cell lines. *Cancer Res.* **48,** 6166–72 (1988).

369. Schilder, R. J., Hall, L., Monks, A., Handel, L. M., Fornace A. J., J., *et al.* Metallothionein gene expression and resistance to cisplatin in human ovarian cancer. *Int. J. Cancer* **45,** 416–22 (1990).

370. Domcke, S., Sinha, R., Levine, D. A., Sander, C. & Schultz, N. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat. Commun.* **4,** 2126 (2013).

371. Andrews, S. *FastQC: a quality control tool for high throughput sequence data.* Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc (2010).

372. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32,** 3047–3048 (2016).

373. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–20 (2014).

374. Dobin, A. & Gingeras, T. R. Mapping RNA-seq Reads with STAR. *Curr. Protoc. Bioinformatics* **51,** 11.14.1–11.14.19 (2015).

375. Birney, E., Andrews, T. D., Bevan, P., Caccamo, M., Chen, Y., *et al.* An overview of Ensembl. *Genome Res.* **14,** 925–8 (2004).

376. Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M., *et al.* Ensembl 2022. *Nucleic Acids Res.* **50,** D988–D995 (2022).

377. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

378. Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10,** giab008 (2021).

379. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–30 (2014).

380. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

381. Pagès, H., Carlson, M., Falcon, S. & Li, N. *AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor.* Available at https://bioconductor.org/packages/AnnotationDbi (2023).

382. Jin, Y. & Hammell, M. Analysis of RNA-Seq Data Using TEtranscripts. *Methods Mol. Biol.* **1751,** 153–167 (2018).

383. Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31,** 3593–3599 (2015).

384. Bendall, M. L., de Mulder, M., Iñiguez, L. P., Lecanda-Sánchez, A., Pérez-Losada, M., *et al.* Telescope: Characterization of the retrotranscriptome by accurate estimation of transposable element expression. *PLoS Comput. Biol.* **15,** e1006453 (2019).

385. Brunton, H., Garner, I. M., Bailey, U.-M., Upstill-Goddard, R. & Bailey, P. J. Using Chromatin Accessibility to Delineate Therapeutic Subtypes in Pancreatic Cancer Patient-Derived Cell Lines. *STAR Protocols* **1,** 100079 (2020).

386. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

387. Benjamin, C. H., Jin-Wook, L., Otto, J., Meenakshi, S. K., Keenan, G., *et al.* The ENCODE Uniform Analysis Pipelines. *bioRxiv,* doi: 10.1101/2023.04.04.535623 (2023).

388. Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44,** W160–W165 (2016).

389. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).

390. Stark, R. & Brown, G. *DiffBind: differential binding analysis of ChIP-Seq peak data.* Available at http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf (2011).

391. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31,** 2382–2383 (2015).

392. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38,** 576–89 (2010).

393. Roby, K. F., Taylor, C. C., Sweetwood, J. P., Cheng, Y., Pace, J. L., *et al.* Development of a syngeneic mouse model for events related to ovarian cancer. *Carcinogenesis* **21,** 585–91 (2000).

394. Walton, J., Blagih, J., Ennis, D., Leung, E., Dowson, S., *et al.* CRISPR/Cas9-Mediated Trp53 and Brca2 Knockout to Generate Improved Murine Models of Ovarian High-Grade Serous Carcinoma. *Cancer Res.* **76,** 6118–6129 (2016).

395. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–60 (2009).

396. Walton, J. B., Farquharson, M., Mason, S., Port, J., Kruspig, B., *et al.* CRISPR/Cas9-derived models of ovarian high grade serous carcinoma targeting Brca1, Pten and Nf1, and correlation with platinum sensitivity. *Sci. Rep.* **7,** 16827 (2017).

397. Blighe, K., Rana, S. & Lewis, M. *EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling.* Available at https://bioconductor.org/packages/EnhancedVolcano (2023).

398. Subhash, S. & Kanduri, C. GeneSCF: a real-time based functional enrichment tool with support for multiple organisms. *Bmc Bioinformatics* **17,** 365 (2016).

399. Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *Bmc Bioinformatics* **14,** 128 (2013).

400. Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44,** W90–W97 (2016).

401. Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., *et al.* Gene Set Knowledge Discovery with Enrichr. *Curr. Protoc.* **1,** e90 (2021).

402. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47,** W199–W205 (2019).

403. Ma, Y., Sun, S., Shang, X., Keller, E. T., Chen, M., *et al.* Integrative differential expression and gene set enrichment analysis using summary statistics for scRNA-seq studies. *Nat. Commun.* **11,** 1585 (2020).

404. Gennady, K., Vladimir, S., Nikolay, B., Boris, S., Maxim, N. A., *et al.* Fast gene set enrichment analysis. *bioRxiv,* doi: https://doi.org/10.1101/060012 (2021).

405. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25,** 1841–1842 (2009).

406. Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* **16,** 22 (2015).

407. Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47,** D941–D947 (2018).

408. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).

409. Mounir, M., Lucchetta, M., Silva, T. C., Olsen, C., Bontempi, G., *et al.* New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. *PLoS Comput. Biol.* **15,** e1006701 (2019).

410. Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., *et al.* TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44,** e71 (2016).

411. Jiménez-Sánchez, A., Cast, O. & Miller, M. L. Comprehensive Benchmarking and Integration of Tumor Microenvironment Cell Estimation Methods. *Cancer Res.* **79,** 6238–6246 (2019).

412. Ku, F. C., Wu, R. C., Yang, L. Y., Tang, Y. H., Chang, W. Y., *et al.* Clear cell carcinomas of the ovary have poorer outcomes compared with serous carcinomas: Results from a single-center Taiwanese study. *J. Formos. Med. Assoc.* **117,** 117–125 (2018).

413. Therneau, T. M. *A Package for Survival Analysis in R.* Available at https://CRAN.R-project.org/package=survival (2024).

414. Fox, J. & Weisberg, S. *An R Companion to Applied Regression.* Third (Sage, Thousand Oaks CA, 2019).

415. Kassambara, A. *ggpubr: 'ggplot2' Based Publication Ready Plots.* Available at https://rpkgs.datanovia.com/ggpubr/ (2023).

416. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27,** 1739–1740 (2011).

417. Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., *et al.* BioMart – biological queries made easy. *Bmc Genomics* **10,** 22 (2009).

418. Foroutan, M., Bhuva, D. D., Lyu, R., Horan, K., Cursons, J., *et al.* Single sample scoring of molecular phenotypes. *Bmc Bioinformatics* **19,** 404 (2018).

419. Wang, M., Zhao, Y. & Zhang, B. Efficient Test and Visualization of Multi-Set Intersections. *Sci. Rep.* **5,** 16923 (2015).

420. Jackson, E. L., Olive, K. P., Tuveson, D. A., Bronson, R., Crowley, D., *et al.* The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res.* **65,** 10280–8 (2005).

421. Cortez, A. J., Tudrej, P., Kujawa, K. A. & Lisowska, K. M. Advances in ovarian cancer therapy. *Cancer Chemother. Pharmacol.* **81,** 17–38 (2018).

422. Zeller, C., Dai, W., Steele, N. L., Siddiq, A., Walley, A. J., *et al.* Candidate DNA methylation drivers of acquired cisplatin resistance in ovarian cancer identified by methylome and expression profiling. *Oncogene* **31,** 4567–4576 (2012).

423. Read, J. A., Tart, J., Rawlins, P. B., Gregson, C., Jones, K., *et al.* Rapid Identification of Novel Allosteric PRC2 Inhibitors. *Acs Chem. Biol.* **14,** 2134–2140 (2019).

424. Brown, R., Fuchter, M. J., Chapman-Rothe, N., Srimongkolpithak, N., Caron, J., *et al. EP-2828250-A1* Patent. 2012.

425. Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41,** D955–61 (2013).

426. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9,** 9354 (2019).

427. Wolf, C. R., Hayward, I. P., Lawrie, S. S., Buckton, K., McIntyre, M. A., *et al.* Cellular heterogeneity and drug resistance in two ovarian adenocarcinoma cell lines derived from a single patient. *Int. J. Cancer* **39,** 695–702 (1987).

428. Davidovich, C. & Cech, T. R. The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. *RNA* **21,** 2007–22 (2015).

429. Wu, Z., Nicoll, M. & Ingham, R. J. AP-1 family transcription factors: a diverse family of proteins that regulate varied cellular activities in classical hodgkin lymphoma and ALK+ALCL. *Exp. Hematol. Oncol.* **10,** 4 (2021).

430. Barral, A. & Déjardin, J. The chromatin signatures of enhancers and their dynamic regulation. *Nucleus* **14,** 2160551 (2023).

431. Schoenfelder, S. & Fraser, P. Long-range enhancer–promoter contacts in gene expression control. *Nat. Rev. Genet.* **20,** 437–455 (2019).

432. Bejjani, F., Tolza, C., Boulanger, M., Downes, D., Romero, R., *et al.* Fra-1 regulates its target genes via binding to remote enhancers without exerting major control on chromatin architecture in triple negative breast cancers. *Nucleic Acids Res.* **49,** 2488–2508 (2021).

433. Tremblay-LeMay, R., Rastgoo, N., Pourabdollah, M. & Chang, H. EZH2 as a therapeutic target for multiple myeloma and other haematological malignancies. *Biomark. Res.* **6,** 34 (2018).

434. Lu, R. & Wang, G. G. Pharmacologic Targeting of Chromatin Modulators As Therapeutics of Acute Myeloid Leukemia. *Front. Oncol.* **7,** 241 (2017).

435. Kato, S., Weng, Q. Y., Insco, M. L., Chen, K. Y., Muralidhar, S., *et al.* Gain-of-Function Genetic Alterations of G9a Drive Oncogenesis. *Cancer Discov.* **10,** 980–997 (2020).

436. Whetstine, J. R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., *et al.* Reversal of Histone Lysine Trimethylation by the JMJD2 Family of Histone Demethylases. *Cell* **125,** 467–481 (2006).

437. Sbirkov, Y., Schenk, T., Kwok, C., Stengel, S., Brown, R., *et al.* Dual inhibition of EZH2 and G9A/GLP histone methyltransferases by HKMTI-1-005 promotes differentiation of acute myeloid leukemia cells. *Front. Cell Dev. Biol.* **11,** 1076458 (2023).

438. Li, Y., Goldberg, E. M., Chen, X., Xu, X., McGuire, J. T., *et al.* Histone methylation antagonism drives tumor immune evasion in squamous cell carcinomas. *Mol. Cell* **82,** 3901–3918.e7 (2022).

439. McDonald, J. I., Diab, N., Arthofer, E., Hadley, M., Kanholm, T., *et al.* Epigenetic Therapies in Ovarian Cancer Alter Repetitive Element Expression in a TP53-Dependent Manner. *Cancer Res.* **81,** 5176–5189 (2021).

440. Wellenstein, M. D. & de Visser, K. E. Cancer-Cell-Intrinsic Mechanisms Shaping the Tumor Immune Landscape. *Immunity* **48,** 399–416 (2018).

441. Von Locquenghien, M., Rozalén, C. & Celià-Terrassa, T. Interferons in cancer immunoediting: sculpting metastasis and immunotherapy response. *J. Clin. Invest.* **131,** e143296 (2021).

442. Gebrie, A. Transposable elements as essential elements in the control of gene expression. *Mobile DNA* **14,** 9 (2023).

443. Jin, L. *Interruptional Activity and Simulation of Transposable Elements.* Thesis (University of Saskatchewan, 2017).

444. Campbell, K. J. & White, R. J. MYC regulation of cell growth through control of transcription by RNA polymerases I and III. *Cold Spring Harb. Perspect. Med.* **4,** a018408 (2014).

445. Zhang, Q., Shi, Y., Liu, S., Yang, W., Chen, H., *et al.* EZH2/G9a interact to mediate drug resistance in non-small-cell lung cancer by regulating the SMAD4/ERK/c-Myc signaling axis. *Cell Rep.* **43** (2024).

446. Refsland, E. W. & Harris, R. S. The APOBEC3 family of retroelement restriction factors. *Curr. Top. Microbiol. Immunol.* **371,** 1–27 (2013).

447. Henderson, S. & Fenton, T. APOBEC3 genes: retroviral restriction factors to cancer drivers. *Trends Mol. Med.* **21,** 274–84 (2015).

448. Petljak, M., Alexandrov, L. B., Brammeld, J. S., Price, S., Wedge, D. C., *et al.* Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. *Cell* **176,** 1282–1294.e20 (2019).

449. Laing, S., Kruspig, B., Shaw, R., Officer-Jones, L., Edwards, S., *et al.* ERBB signalling contributes to immune evasion in KRAS-driven lung adenocarcinoma. *bioRxiv,* doi: 10.1101/2023.07.24.550274 (2023).

450. Neidler, S., Kruspig, B., Hewit, K., Monteverde, T., Gyuraszova, K., *et al.* Identification of a Clinically Relevant Signature for Early Progression in KRAS-Driven Lung Adenocarcinoma. *Cancers (Basel)* **11,** 600 (2019).

451. Casey, S. C., Baylot, V. & Felsher, D. W. The MYC oncogene is a global regulator of the immune response. *Blood* **131,** 2007–2015 (2018).

452. Li, J., Dong, T., Wu, Z., Zhu, D. & Gu, H. The effects of MYC on tumor immunity and immunotherapy. *Cell Death Discov.* **9,** 103 (2023).

453. Sturm, G., Finotello, F., Petitprez, F., Zhang, J. D., Baumbach, J., *et al.* Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics* **35,** i436–i445 (2019).

454. Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37,** 773–782 (2019).

455. Yoon, Young-Kwang and Kim, Hwang-Phill and Han, Sae-Won and Oh, Do Youn and Im, Seock-Ah and Bang, Yung-Jue and Kim, Tae-You. KRAS mutant lung cancer cells are differentially responsive to MEK inhibitor due to AKT or STAT3 activation: Implication for combinatorial approach. *Mol. Carcinog.* **49,** 353–362 (2010).

456. Blancato, J., Singh, B., Liu, A., Liao, D. J. & Dickson, R. B. Correlation of amplification and overexpression of the c-myc oncogene in high-grade breast cancer: FISH, in situ hybridisation and immunohistochemical analyses. *Br. J. Cancer* **90,** 1612–9 (2004).

457. Reyes-González, J. M. & Vivas-Mejía, P. E. c-MYC and Epithelial Ovarian Cancer. *Front. Oncol.* **11,** 601512 (2021).

458. Jin, H. & Liu, Z. A benchmark for RNA-seq deconvolution analysis under dynamic testing environments. *Genome Biol.* **22,** 102 (2021).

459. Saito, N., Sato, Y., Abe, H., Wada, I., Kobayashi, Y., *et al.* Selection of RNA-based evaluation methods for tumor microenvironment by comparing with histochemical and flow cytometric analyses in gastric cancer. *Sci. Rep.* **12,** 8576 (2022).

460. Sodir, N. M., Kortlever, R. M., Barthet, V. J. A., Campos, T., Pellegrinet, L., *et al.* MYC Instructs and Maintains Pancreatic Adenocarcinoma Phenotype. *Cancer Discov.* **10,** 588–607 (2020).

461. Morel, D., Jeffery, D., Aspeslagh, S., Almouzni, G. & Postel-Vinay, S. Combining epigenetic drugs with other therapies for solid tumours — past lessons and future promise. *Nat. Rev. Clin. Oncol.* **17,** 91–107 (2020).

462. Feehley, T., O'Donnell, C. W., Mendlein, J., Karande, M. & McCauley, T. Drugging the epigenome in the age of precision medicine. *Clin. Epigenetics* **15,** 6 (2023).

463. Garralda, E., Beaulieu, M.-E., Moreno, V., Casacuberta-Serra, S., Martínez-Martín, S., *et al.* MYC targeting by OMO-103 in solid tumors: a phase 1 trial. *Nat. Med.,* 762–771 (2024).

464. Spiliopoulou, P. *G9a and EZH2 histone methyltransferase inhibition modulates the immune microenvironment of ovarian cancer.* Thesis (University of Glasgow, 2020).

Appendix A

Appendix

A	B	Neither	A Not B	B Not A	Both	Log2 Odds Ratio	p-Value	q-Value	Tendency
EZH2	EHMT2	5677	2111	788	1559	2.412	<0.001	<0.001	Co-occurrence

Table A.1 **Co-occurrence of *EZH2* and *G9a* alterations in TCGA patient cohorts.** Statistical testing of the co-occurence of *EZH2* and *EHMT2* (*G9a*) alterations in The Cancer Genome Atlas (TCGA) cohorts (2-sided Fisher Exact Test). Alterations include mutations, structural variants, amplifications, deep deletions and over- or under-expression as defined by a z-score ± 2. Table produced using cBioPortal [121–123].

A	B	Neither	A Not B	B Not A	Both	Log2 Odds Ratio	p-Value	q-Value	Tendency
EZH2	MYC	5325	2366	1140	1304	1.364	<0.001	<0.001	Co-occurrence
EHMT2	MYC	6234	1457	1554	890	1.293	<0.001	<0.001	Co-occurrence

Table A.2 **Co-occurrence of *EZH2* or *G9a* alterations with *MYC* alterations in TCGA patient cohorts.** Statistical testing of the co-occurence of *EZH2* or *EHMT2* (*G9a*) alterations with *MYC* alterations in The Cancer Genome Atlas (TCGA) cohorts (2-sided Fisher Exact Test). Alterations include mutations, structural variants, amplifications, deep deletions and over- or under-expression as defined by a z-score ± 2. Table produced using cBioPortal [121–123].

Intersections	Degree	Observed.Overlap	Expected.Overlap	FE	P.value
mm39sigDN & mm10sigDN	2	411	9.205629	44.6466	0
mm39sigUP & mm10sigUP	2	828	23.82303	34.75629	0
mm39sigDN & mm10sigUP	2	0	14.37987	0	1
mm39sigUP & mm10sigDN	2	0	15.2509	0	1

Table A.3 **Statistical comparison of RNAseq results using mm39 versus mm10 reference genome annotation.** Statistical overlap between both up- ("sigUP") and down-regulated ("sigDN") DEGs according to the analysis performed in this thesis (using the mm39 mouse reference genome annotation) as compared to the analysis published in [155] (using the mm10 annotation). All DEGs are defined using p-adj < 0.05.

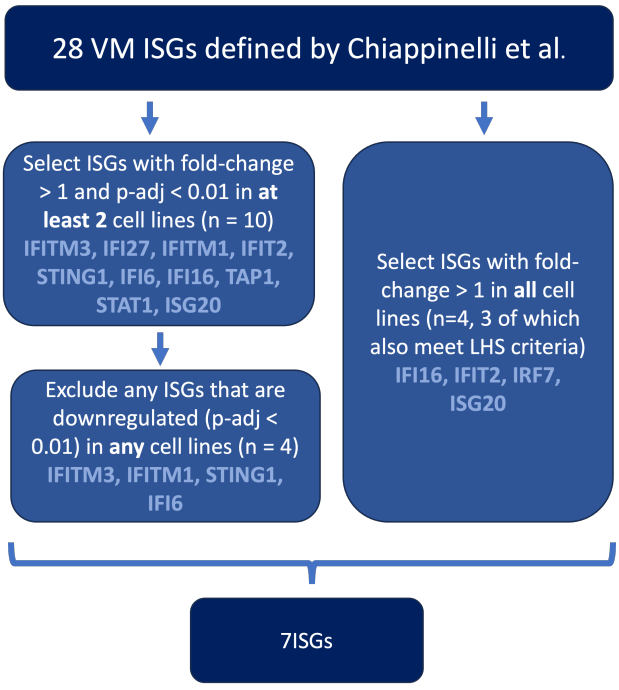


Figure A.1 **7ISG selection criteria.** Selection criteria for the 7 ISGs comprising the 7ISG geneset, selected from the broader panel of 28 VM ISGs defined in [439] (see table 4.1).

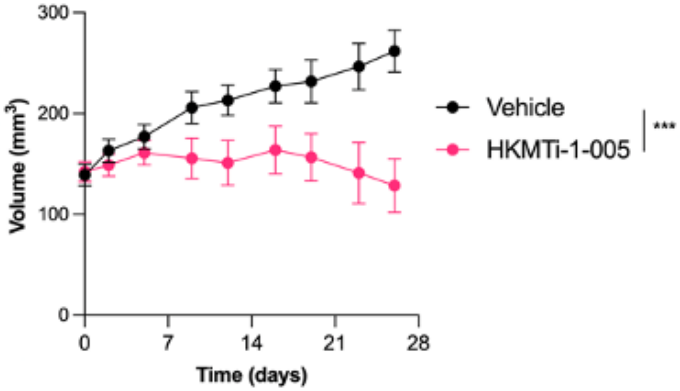


Figure A.2 **Tumour growth inhibitory effect of HKMTi-1-005 in IGROV1 xenograft model.** Mean tumour volume of IGROV1 xenografts (subcutaneous) in BALBc nu/nu mice. Mice treated with twice daily HKMTi-1-005 at 40mg/kg, using intraperitoneal administration, on days 0–4 and 7–11 inclusive. Error bars are SEM. *** = $p < 0.001$.

Intersections	**Degree**	**Observed.Overlap**	**Expected.Overlap**	**FE**	**P.value**
ID8_DiffBind & ID8_SUBTRACT	2	12724	10288.08117	1.23677096	0
P1_DiffBind & P1_SUBTRACT	2	1689	712.7043687	2.36984657	0
P4_DiffBind & P4_SUBTRACT	2	1672	721.1661973	2.31846696	0
B1_DiffBind & B1_SUBTRACT	2	7140	3414.175863	2.09128067	0
B2_DiffBind & B2_SUBTRACT	2	964	437.1991614	2.20494476	1.967E-258
F3_DiffBind & F3_SUBTRACT	2	6046	3037.838377	1.99023096	0

Table A.4 **Statistical comparison of ATACseq approach results.** Statistical testing of the overlap between genes with increased accessibility calculated using the two different ATACseq analysis approaches (see 3.15), in the ID8 *in vivo* model, the PEO1 (P1) cell line, the PEO4 (P4) cell line, the BRCA1 (B1) ID8 cell line, the BRCA2 (B2) ID8 cell line and the F3 ID8 cell line. ATACseq results are either genes associated unique HKMTi-1-005 treatment peaks of accessibility (labelled "SUBTRACT", subtraction method) or genes associated with upregulated DACs (labelled "DAC_UP", DAC method).

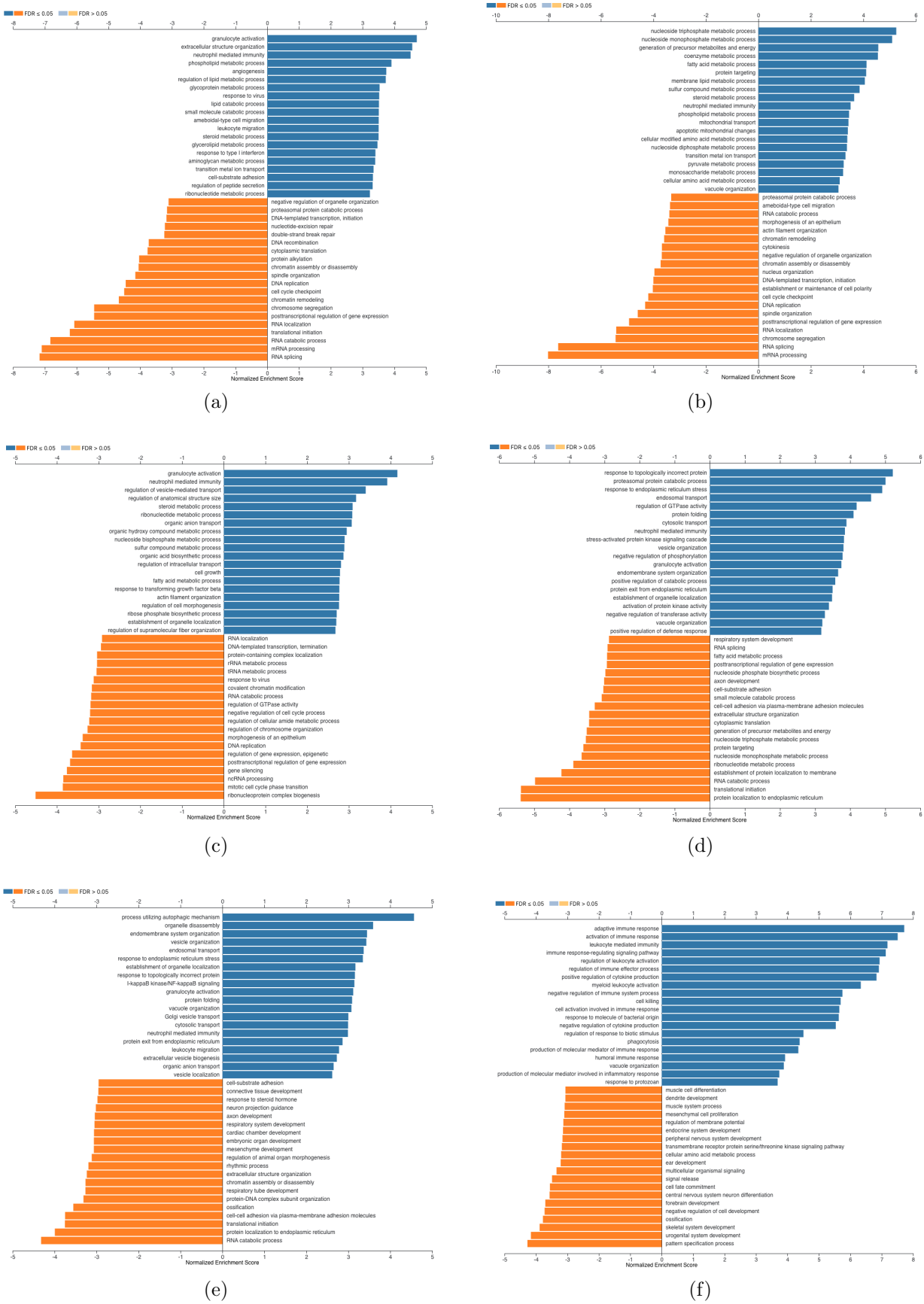


Figure A.3 **RNAseq GSEA, GOBP pathways: HGSC samples.** GSEA results from WebGestalt [402], showing the top 20 enriched and top 20 depleted GOBP pathways following HKMTi-1-005 treatment in HGSC samples. Samples include (left to right, top to bottom): PEA2, PEO1, PEO4, OVCAR5, OVCAR8 cell lines and ID8 *in vivo*.



Figure A.4 **RNAseq GSEA, Hallmark pathways: HGSC samples.** GSEA results from WebGestalt [402], showing the top 20 enriched and top 20 depleted Hallmark pathways following HKMTi-1-005 treatment in HGSC samples. Samples include (left to right, top to bottom): PEA2, PEO1, PEO4, OVCAR5, OVCAR8 cell lines and ID8 *in vivo*.

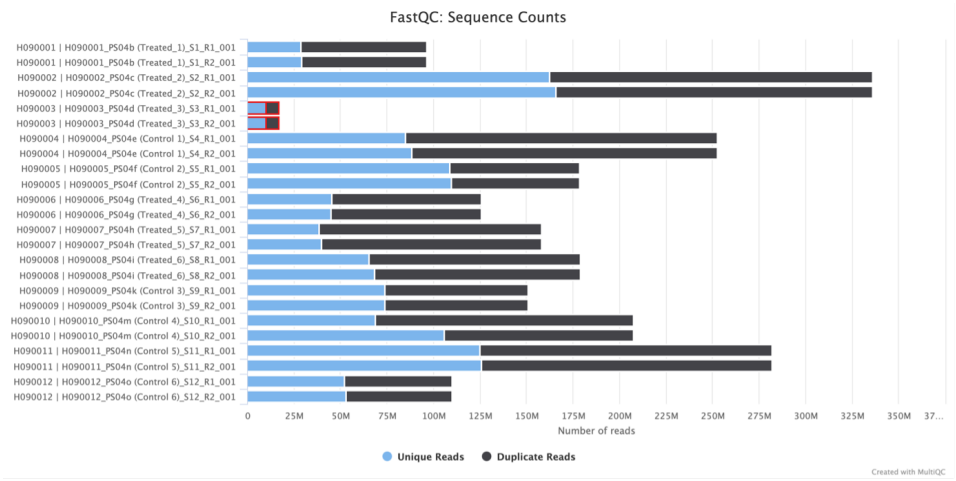


Figure A.5 **ATACseq FastQC sequence counts, before trimming: ID8 *in vivo* samples.** FastQC sequence count for each ID8 *in vivo* ATACseq sample, with an estimate of the proportion of duplicate reads. N=6 treated (PS04b, PS04c, PS04d PS04g, PS04h, PS04i) and n=6 untreated (PS04e, PS04f, PS04k, PS04m, PS04n, PS04o) replicates. One treated ID8 *in vivo* sample (PS04d, T3, highlighted in red) was removed prior to trimming based on these data, due to low read count (<25M).

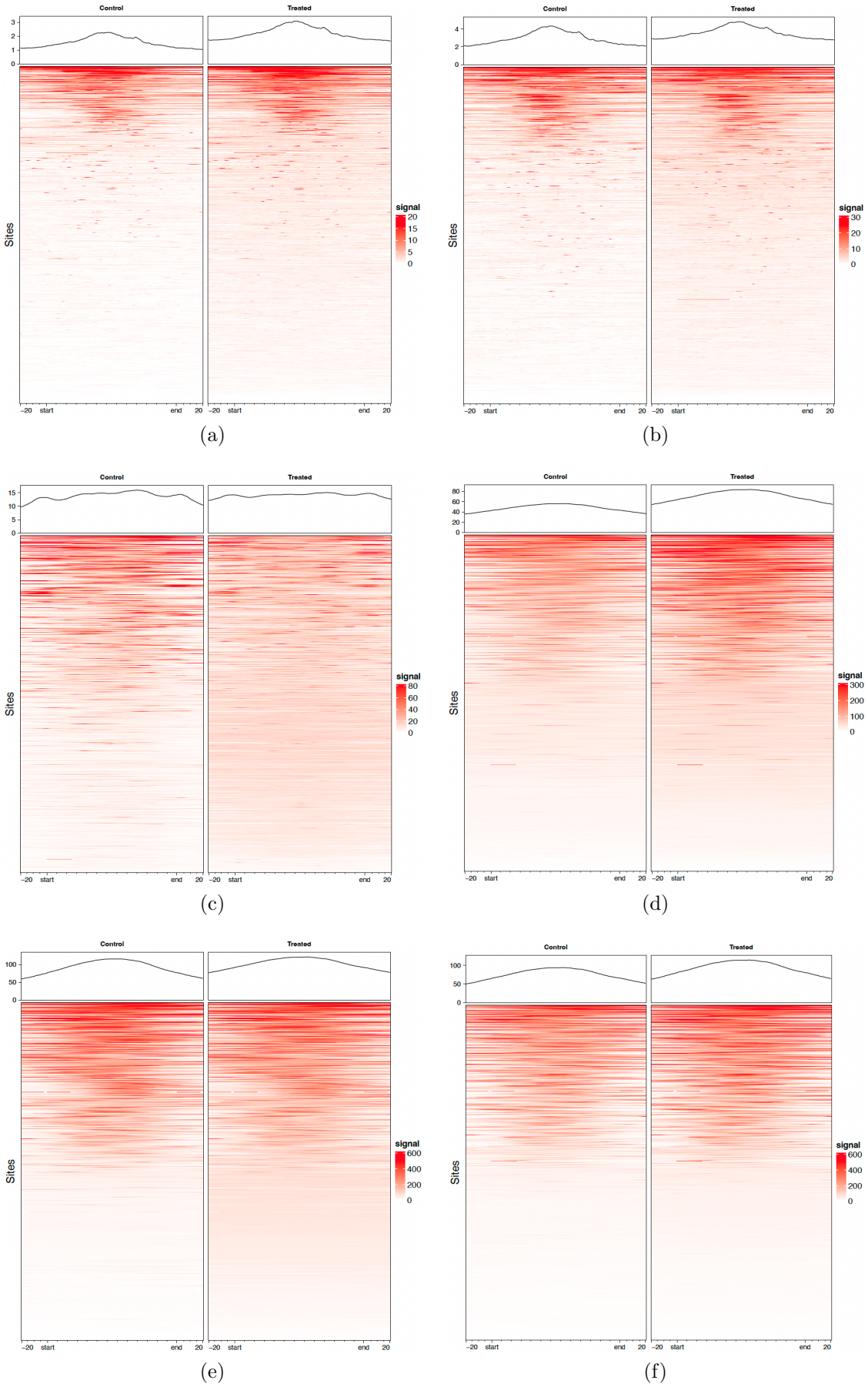


Figure A.6 **H3K9Ac accessibility heatmaps.** Heatmaps generated using DiffBind [390], displaying control versus HKMTi-1-005-treated chromatin accessibility at DNA associated with H3K9Ac histone marks. Samples include the (a) PEO1 cell line, (b) PEO4 cell line, (c) ID8 *in vivo* sample, (d) BRCA1 ID8 cell line, (e) BRCA2 ID8 cell line and (f) F3 ID8 cell line.

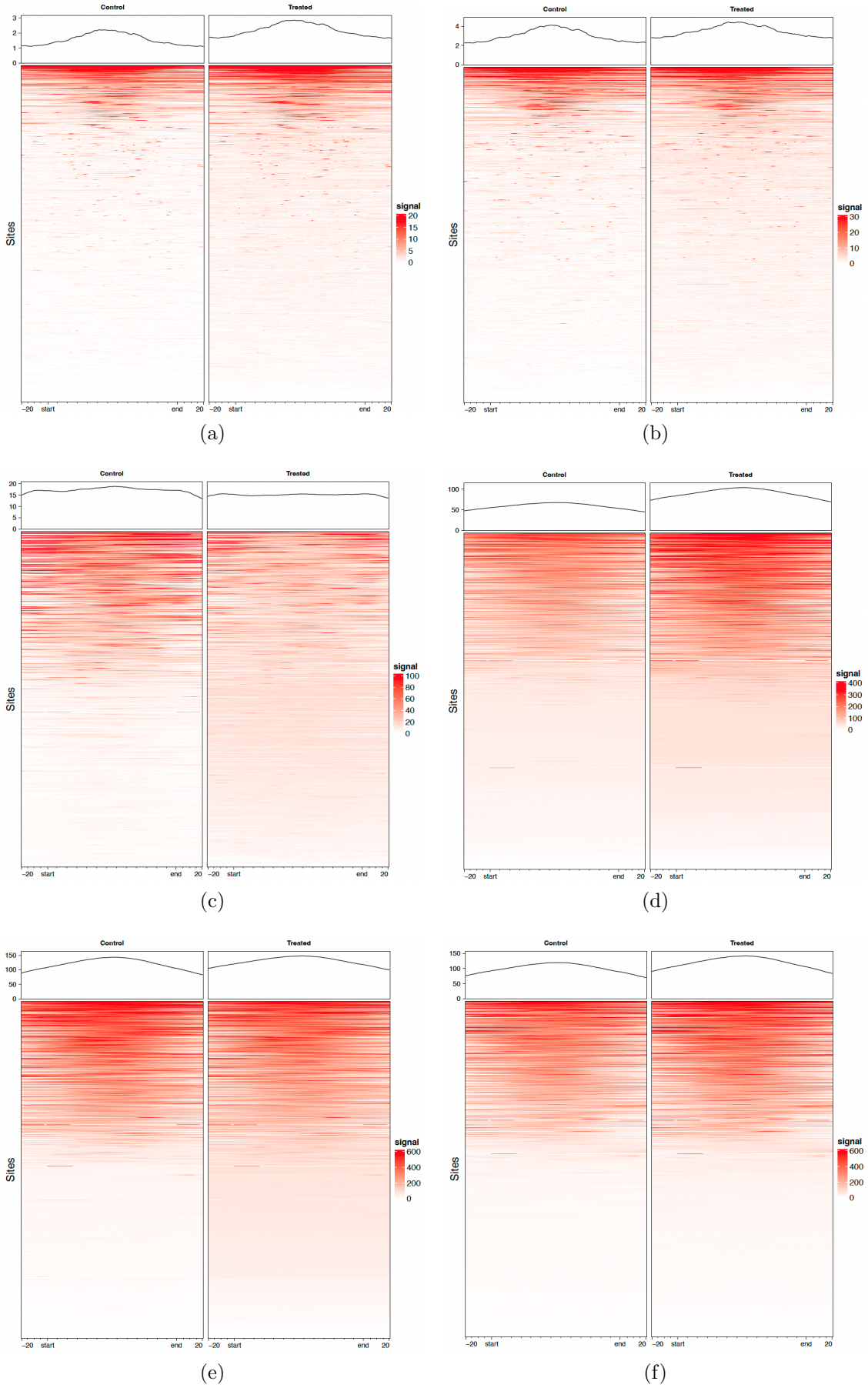


Figure A.7 **H3K27Ac accessibility heatmaps.** Heatmaps generated using DiffBind [390], displaying control versus HKMTi-1-005-treated chromatin accessibility at DNA associated with H3K27Ac histone marks. Samples include the (a) PEO1 cell line, (b) PEO4 cell line, (c) ID8 *in vivo* sample, (d) BRCA1 ID8 cell line, (e) BRCA2 ID8 cell line and (f) F3 ID8 cell line.

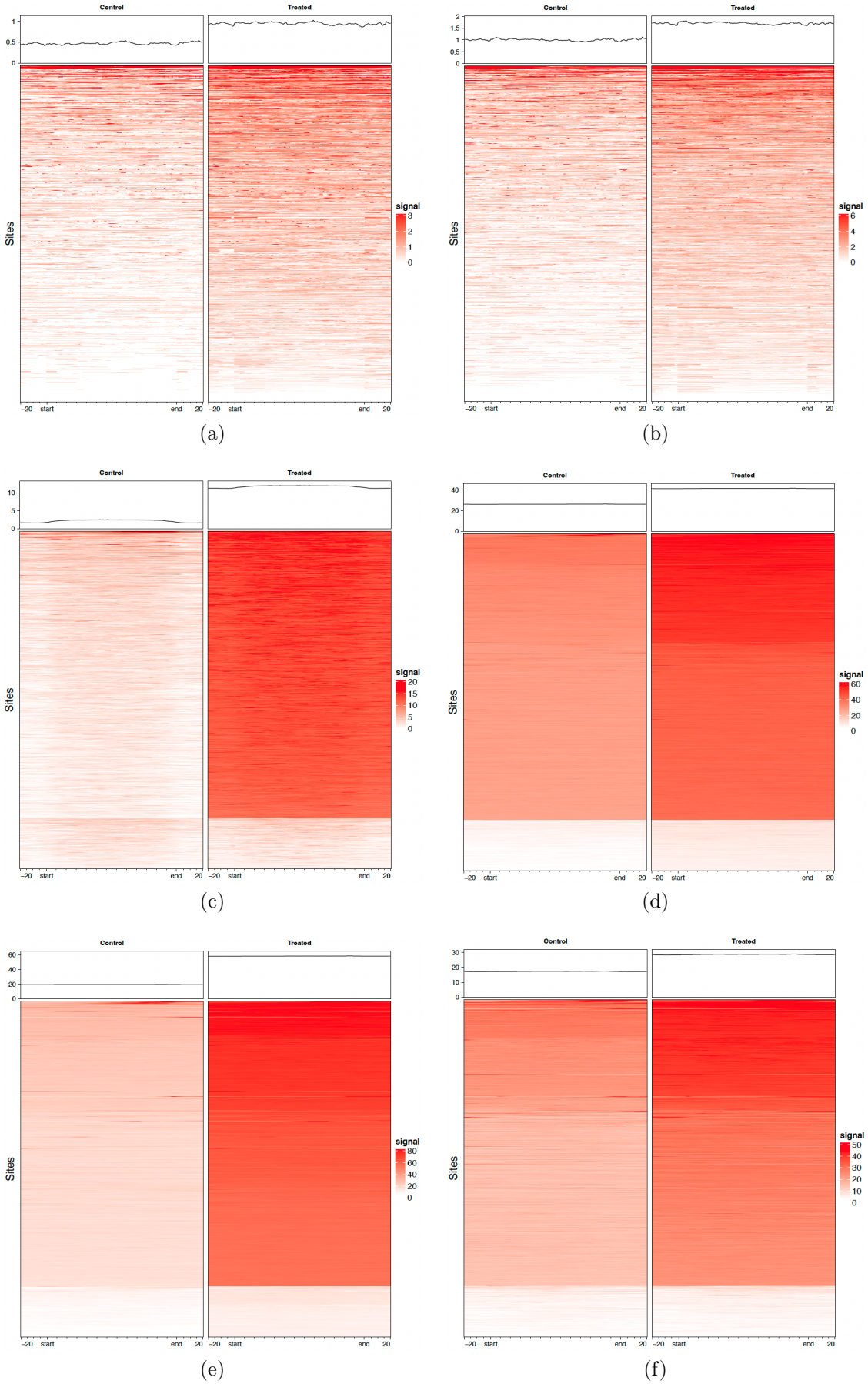


Figure A.8 **H3K36me3 accessibility heatmaps.** Heatmaps generated using DiffBind [390], displaying control versus HKMTi-1-005-treated chromatin accessibility at DNA associated with H3K36me3 histone marks. Samples include the (a) PEO1 cell line, (b) PEO4 cell line, (c) ID8 *in vivo* sample, (d) BRCA1 ID8 cell line, (e) BRCA2 ID8 cell line and (f) F3 ID8 cell line.

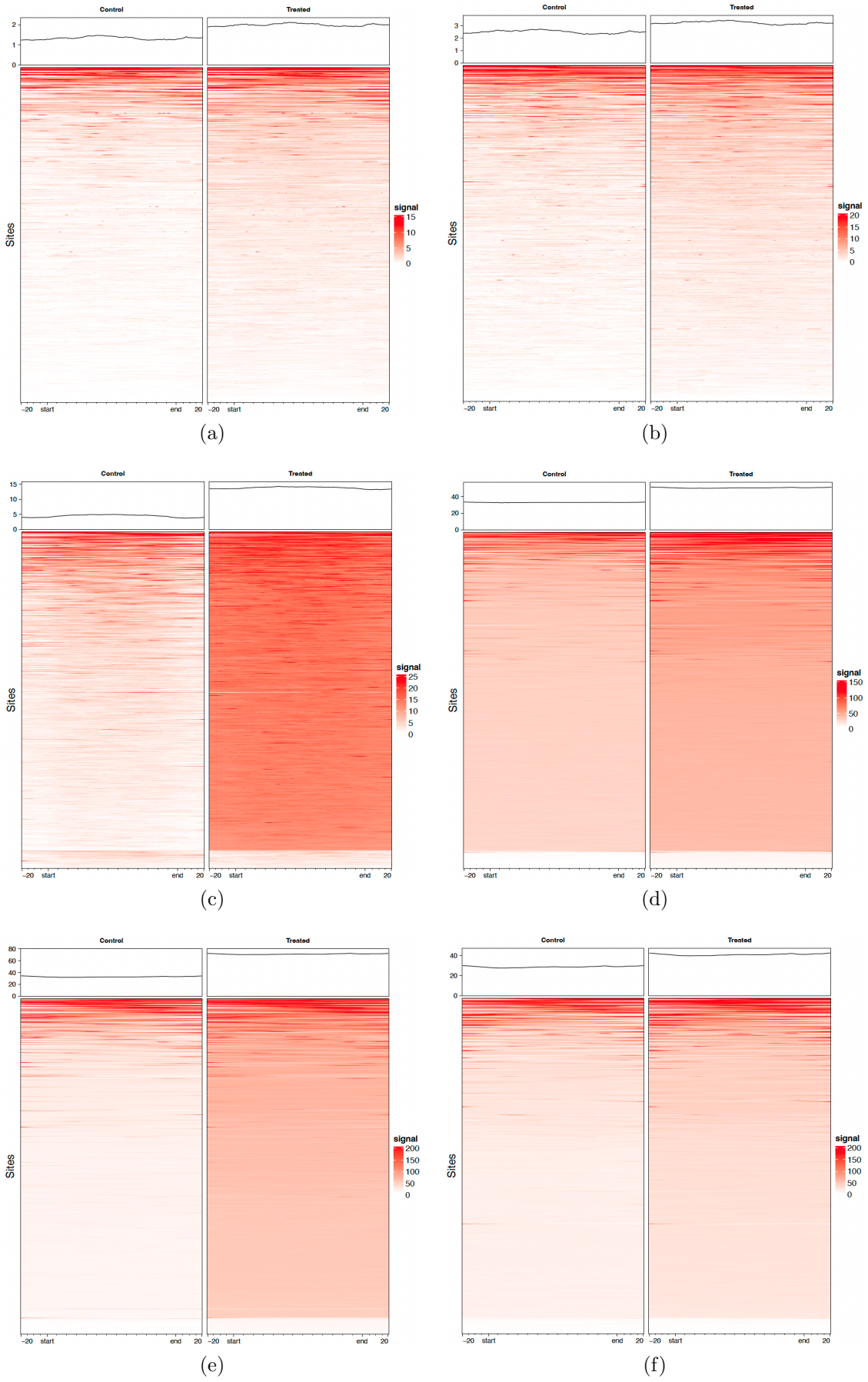


Figure A.9 **H3K4me1 accessibility heatmaps.** Heatmaps generated using DiffBind [390], displaying control versus HKMTi-1-005-treated chromatin accessibility at DNA associated with H3K4me1 histone marks. Samples include the (a) PEO1 cell line, (b) PEO4 cell line, (c) ID8 *in vivo* sample, (d) BRCA1 ID8 cell line, (e) BRCA2 ID8 cell line and (f) F3 ID8 cell line.

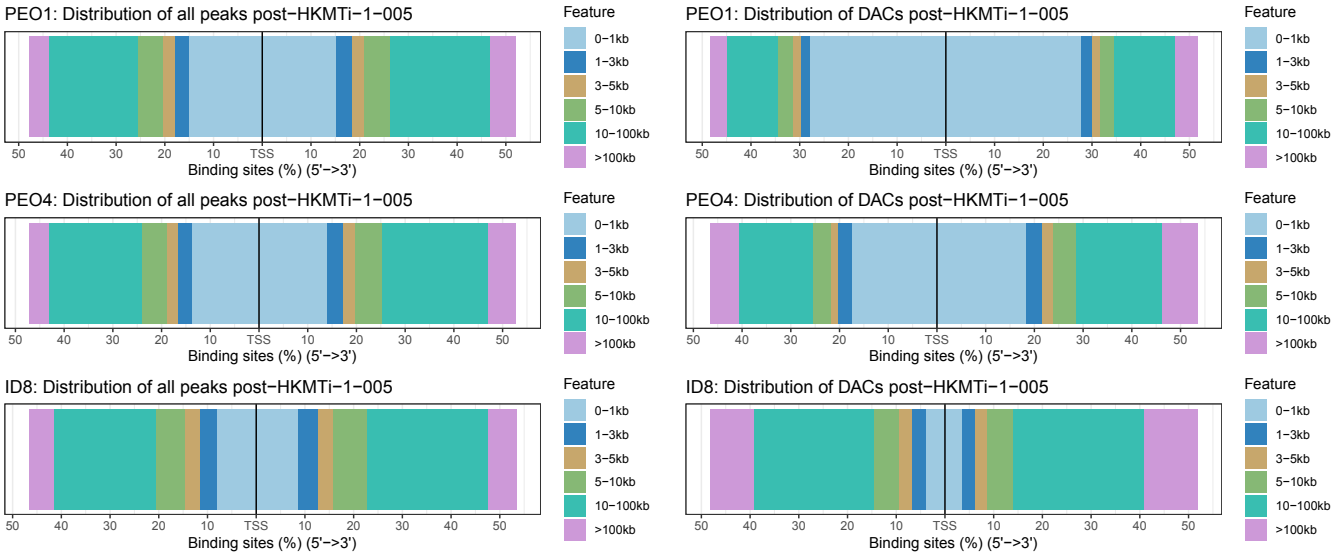


Figure A.10 **Accessibility changes versus TSS location, *in vitro* HGSC or ID8 *in vivo* samples.** Plot visualising the distribution of all peaks of accessibility (left), or differentially accessible chromatin regions (DACs) (right), compared to transcriptional start sites. ATACseq data derived from PEO1 (top), PEO4 (middle) and ID8 *in vivo* samples (bottom).

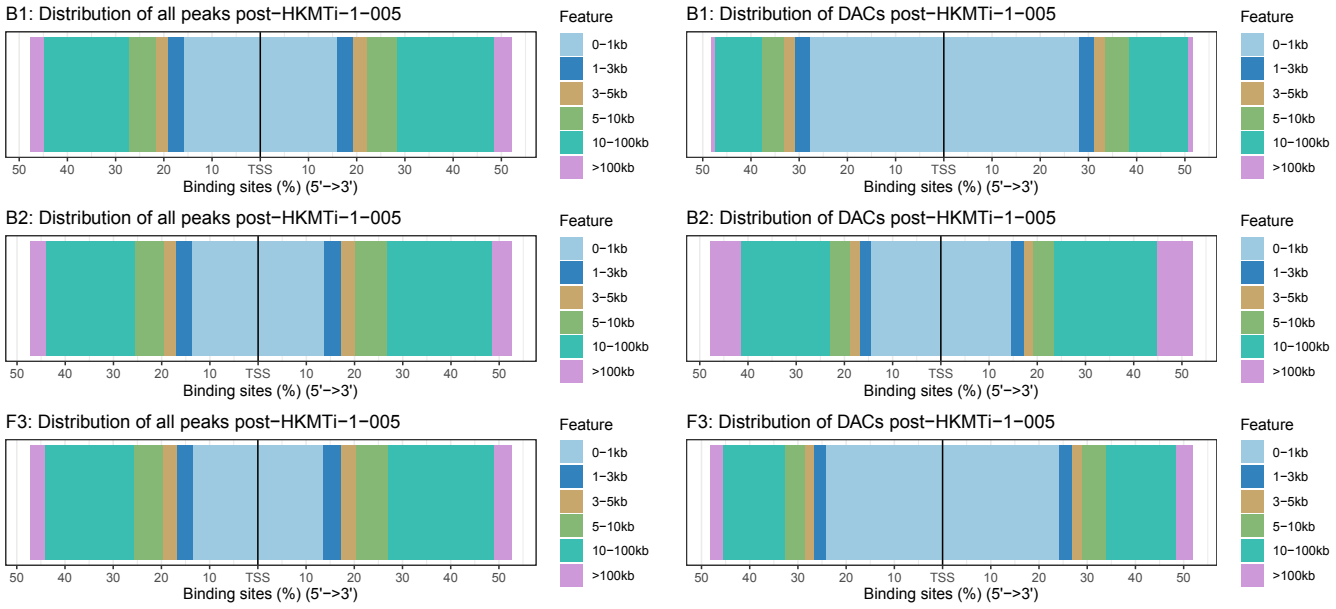


Figure A.11 **Accessibility changes versus TSS location, ID8 *in vitro* samples.** Plot visualising the distribution of all peaks of accessibility (left), or differentially accessible chromatin regions (DACs) (right), compared to transcriptional start sites. ATACseq data derived from *in vitro* BRCA1 (B1, top), BRCA2 (B2, middle) and F3 ID8 samples (bottom).

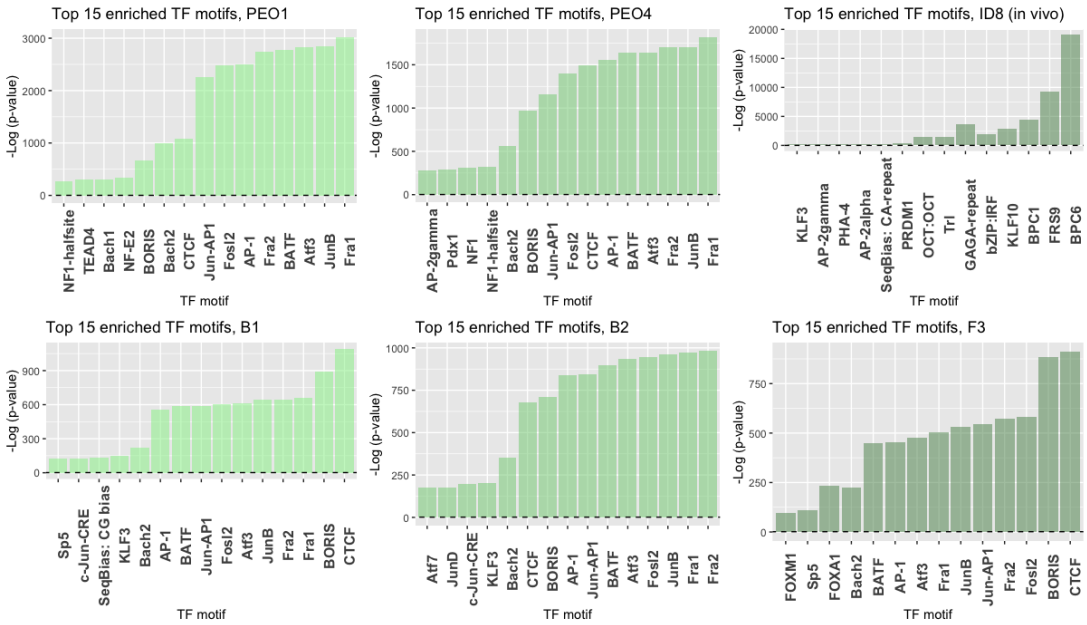


Figure A.12 **Motif enrichment analysis, ATACseq subtraction approach.** Motif analysis performed using HOMER [392], assessing enrichment of transcription factor motifs in unique HKMTi-1-005 treatment peaks of accessibility (subtraction approach) following treatment with HKMTi-1-005. Graphs show top 15 motifs enriched in the following samples from left to right, top to bottom: PEO1, PEO4, ID8 *in vivo*, *in vitro* BRCA1 (B1) ID8, *in vitro* BRCA2 (B2), ID8 and *in vitro* F3 ID8 samples.

Subtraction n = 24	DAC approach n = 47
AP-1	AP-1
AP-2alpha	Atf3
AP-2gamma	Bach1
Atf3	Bach2
Bach1	BATF
Bach2	BMYB
BATF	CTCF
EBF1	EHF
Fosl2	EKLF
Fox:Ebox	ELF1
Fra1	ELF3
Fra2	ELF5
Jun-AP1	Elk4
JunB	ERG
KLF10	ETS1
KLF6	ETV1
NF-E2	Etv2
NF1	EWS:ERG-fusion
Nrf2	EWS:FLI1-fusion
RUNX2	Fli1
Tlx?	Fosl2
Zfp809	Fra1
ZFX	Fra2
Zic3	Hoxc9
	Jun-AP1
	JunB
	KLF14
	KLF3
	Klf4
	KLF5
	KLF6
	MafA
	MafK
	NF-E2
	NF1
	Nrf2
	Pdx1
	RUNX
	RUNX-AML
	RUNX1
	RUNX2
	Sp5
	SPDEF
	TEAD
	TEAD2
	TEAD4
	Tlx?

Table A.5 **Overlap in transcription factor motif enrichment across HGSC samples.** Table of transcription factor motifs enriched in analysis of ATACseq data following HKMTi-1-005 treatment, across PEO1, PEO4, ID8 *in vivo*, *in vitro* BRCA1 (B1) ID8, *in vitro* BRCA2 (B2), ID8 and *in vitro* F3 ID8 samples. Results include motifs enriched by the subtraction (left) and DAC (right) ATACseq analysis methods.

Intersections	Degree	Observed.Overlap	Expected.Overlap	FE	P.value
PEO1_RNA_TEs & PEO1_ATAC_DAC_TEs	2	3	0.13294653	22.565463	0.000351906
PEO4_RNA_TEs & PEO4_ATAC_DAC_TEs	2	2	0.027929838	71.6080068	0.000376612
ID8_RNA_TEs & ID8_ATAC_DAC_TEs	2	30	32.63135954	0.91936102	0.702592596
PEO1_RNA_TEs & PEO1_ATAC_SUBTRACT	2	15	4.476116067	3.35111954	6.25146E-05
PEO4_RNA_TEs & PEO4_ATAC_SUBTRACT	2	5	0.616431921	8.11119579	0.000391658
ID8_RNA_TEs & ID8_ATAC_SUBTRACT	2	99	71.89528504	1.37700268	0.001194958

Table A.6 **Statistical testing of the overlap between more accessible TEs and transcriptionally upregulated TEs at the locus-specific level.** Statistical testing of the overlap between TEs with increased accessibility calculated using the two different ATACseq analysis approaches (see 3.15),in the ID8 *in vivo* model, the PEO1 cell line and the PEO4 cell line. ATACseq results are either TE associated unique HKMTi-1-005 treatment peaks of accessibility (labelled "SUBTRACT", subtraction method) or TE associated with upregulated DACs (labelled "DAC", DAC method).

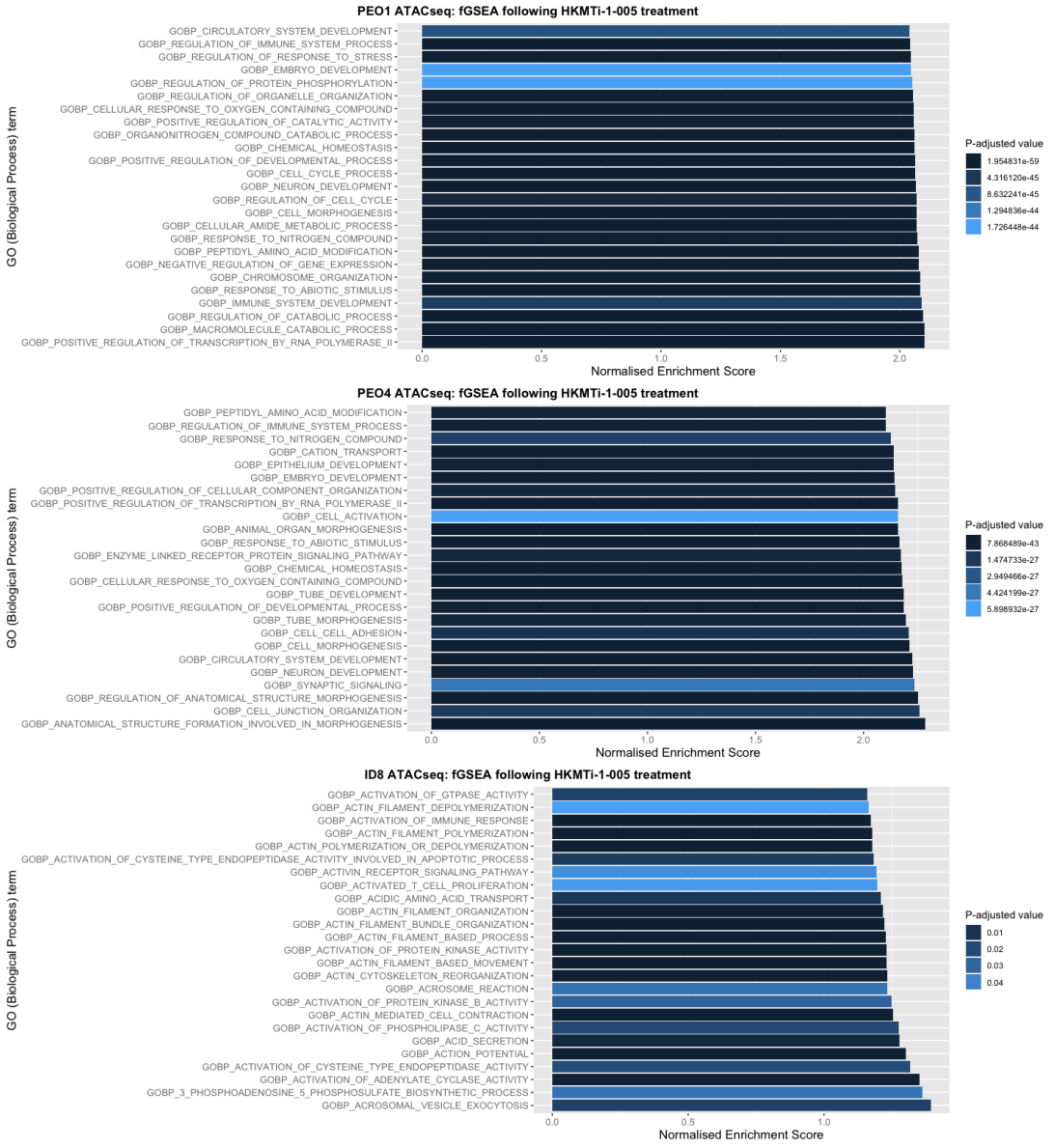
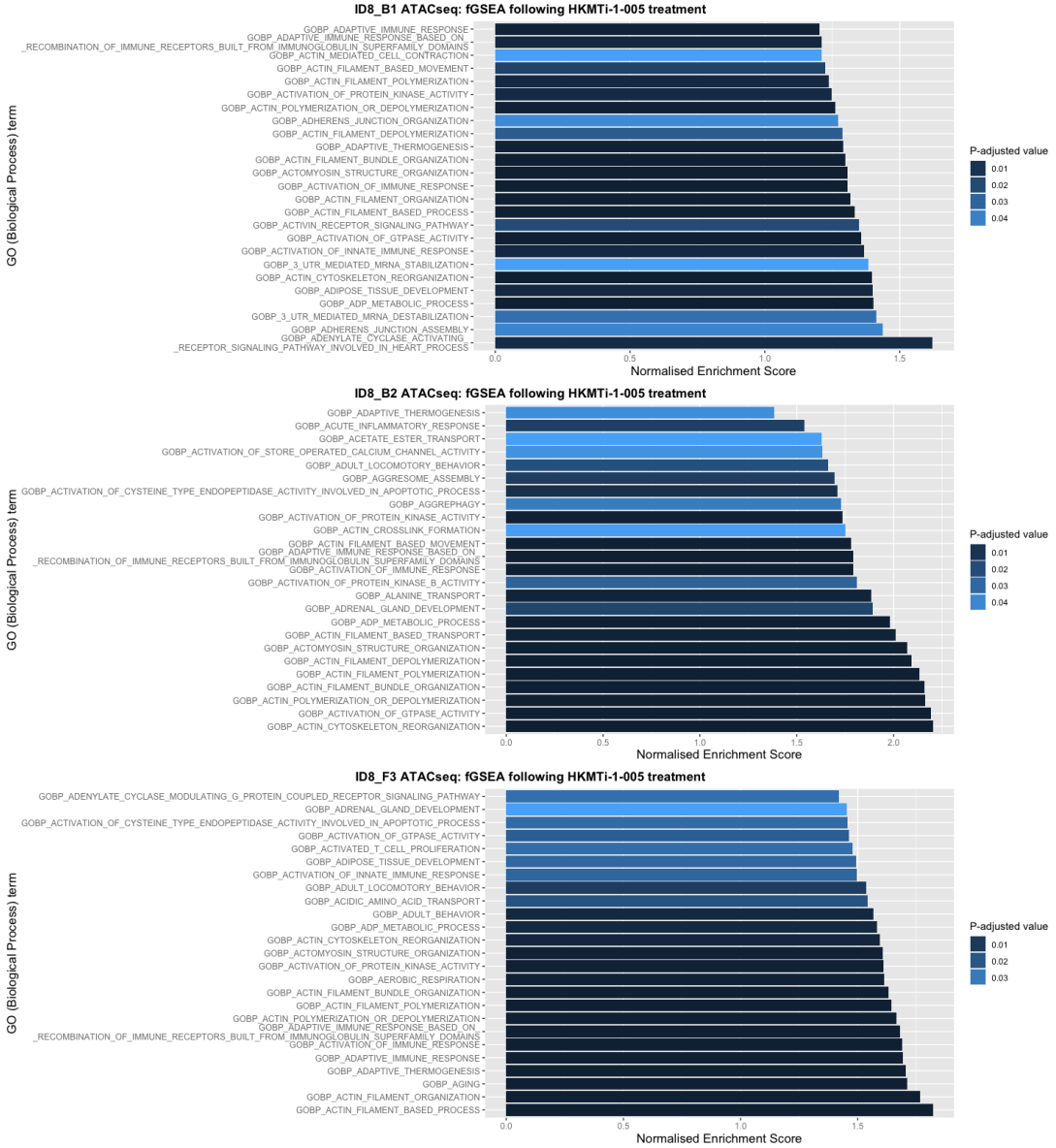


Figure A.13 **ATACseq fGSEA: GOBP database.**



(g)

Figure A.13 **ATACseq fGSEA: GOBP database.** fGSEA results showing the top 25 enriched GOBP pathways, based on DACs following HKMTi-1-005 treatment in (a) PEO1 cell line, PEO4 cell line and ID8 *in vivo*, and (b) *in vitro* BRCA1 (B1) ID8, *in vitro* BRCA2 (B2) ID8 and *in vitro* F3 ID8 cell lines. Lighter bar colour represents smaller p-adj (or greater -log10(p-adj)). DACs defined as p-adj < 0.05, absolute log2FC > 0, with an additional threshold of absolute log2FC of 1 for the ID8 *in vivo* sample.

Geneset	**MYC-amplified (n=65)**	**Non-amplified (n=500)**
HALLMARK_INTERFERON_ALPHA_RESPONSE	0.71 to 0.88	0.62 to 0.71
HALLMARK_INTERFERON_GAMMA_RESPONSE	0.77 to 0.91	0.66 to 0.75
INNATE_IMMUNE_RESPONSE	0.75 to 0.90	0.61 to 0.71
7ISG	0.73 to 0.89	0.58 to 0.69

Table A.7 **95% confidence intervals for correlation between EZH2 repression geneset scores and immune geneset scores, stratified by *MYC* status.** Table showing 95% confidence intervals for correlation between EZH2_REPRESSION_TARGETS scores and immune geneset scores (genesets listed in table) in *MYC*-amplified and non-amplified cell lines from the GDSC2 cell line panel, illustrating significant difference between the strength of correlation between the two cohorts.

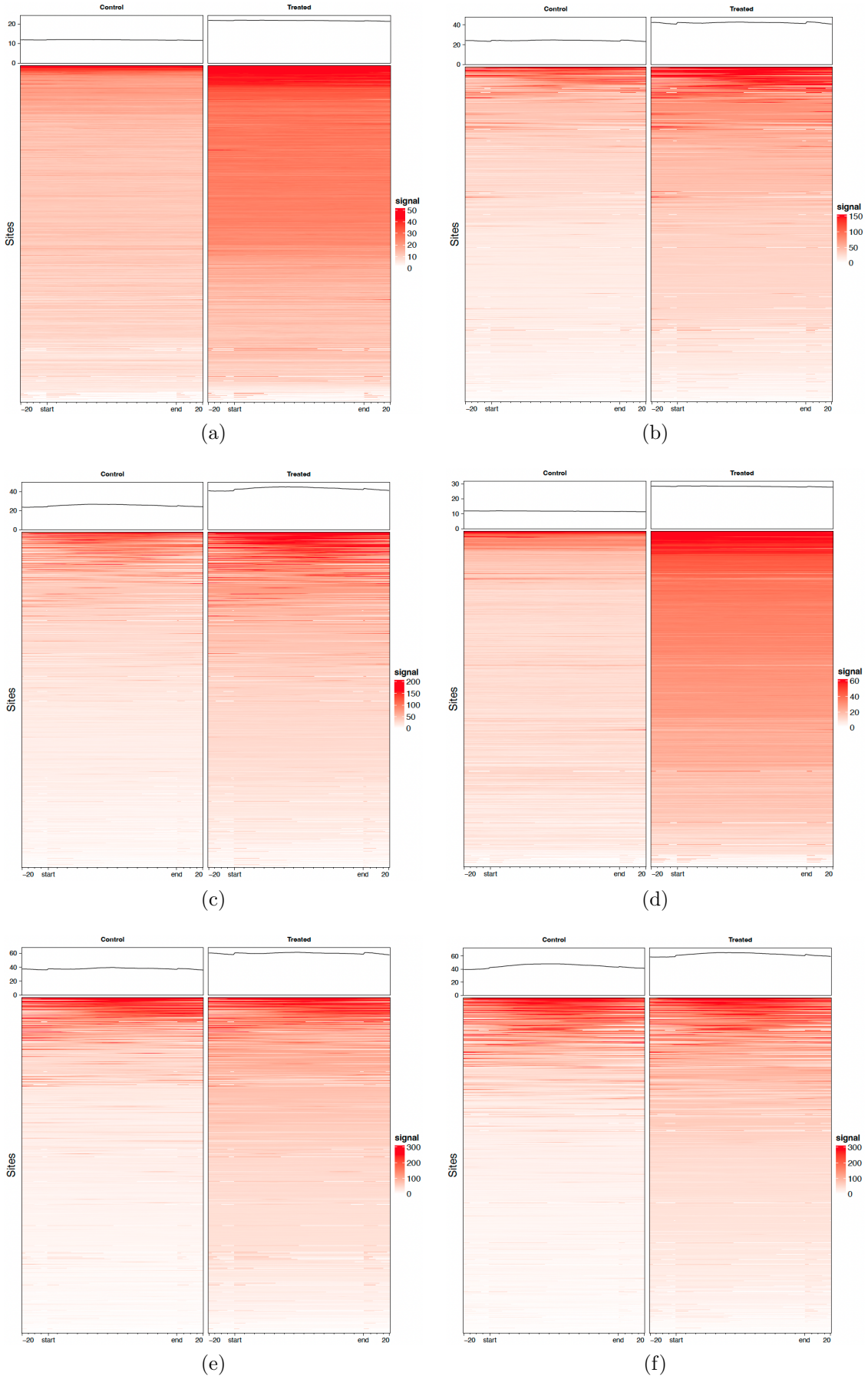
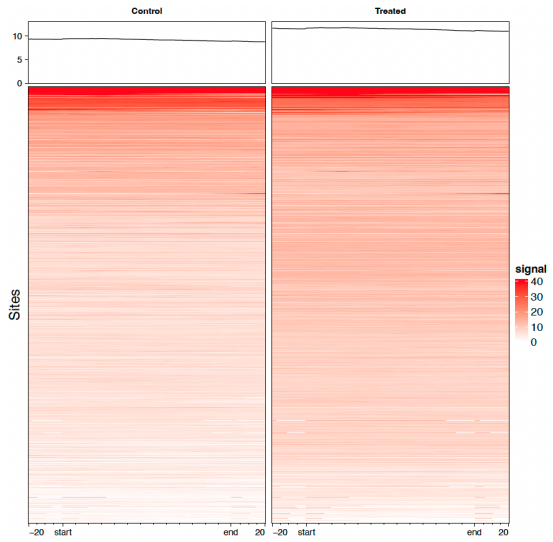
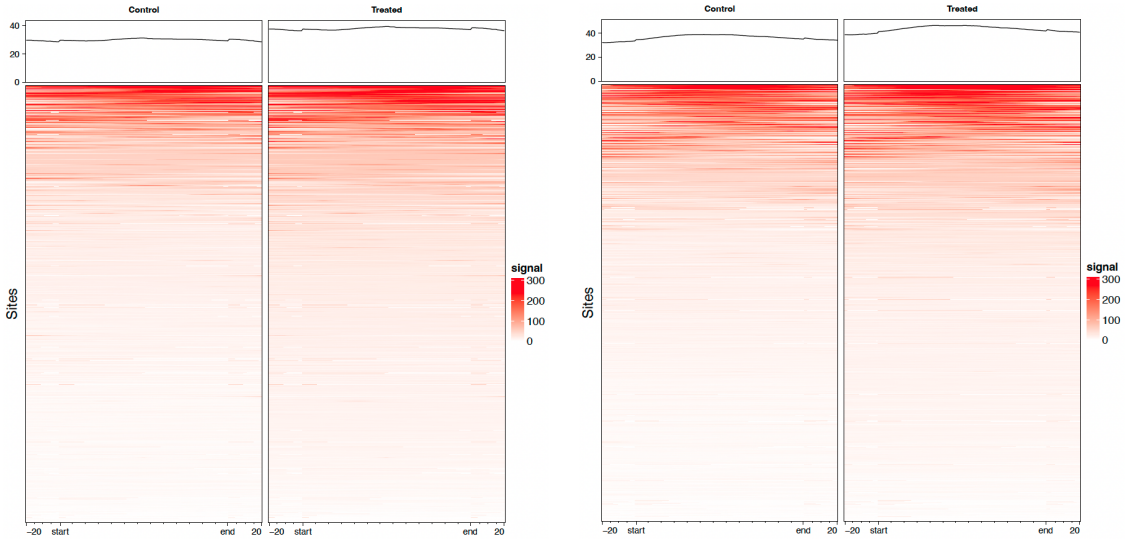


Figure A.14 **Bivalent region accessibility heatmaps.** Continued overleaf.



(g)



(h) (i)

Figure A.14 **Bivalent region accessibility heatmaps.** Heatmaps generated using DiffBind [390], displaying control versus HKMTi-1-005-treated chromatin accessibility at DNA associated with bivalent H3K9me3-H3K36me3 marks, bivalent H3K27me3-H3K4me1 marks and bivalent H3K27me3-H3K4me3 marks in the following the ID8 *in vitro* cell line samples (left to right, top to bottom): BRCA1 (a-c), BRCA2 (d-f) and F3 (g-i), Bivalent regions identified using overlap of coordinates defined for individual histone marks.

Geneset	**MYC-amplified (n=65)**	**Non-amplified (n=500)**
HALLMARK_INTERFERON_ALPHA_RESPONSE	0.43 to 0.75	0.41 to 0.55
HALLMARK_INTERFERON_GAMMA_RESPONSE	0.50 to 0.79	0.47 to 0.60
INNATE_IMMUNE_RESPONSE	0.40 to 0.74	0.39 to 0.54
7ISG	0.38 to 0.72	0.41 to 0.55

Table A.8 **95% confidence intervals for correlation between cMYC-G9a repression geneset scores and immune geneset scores, stratified by *MYC* status.** Table showing 95% confidence intervals for correlation between MYC_G9A_REPRESSION_TARGETS scores and immune geneset scores (genesets listed in table) in *MYC*-amplified and non-amplified cell lines from the GDSC2 cell line panel.

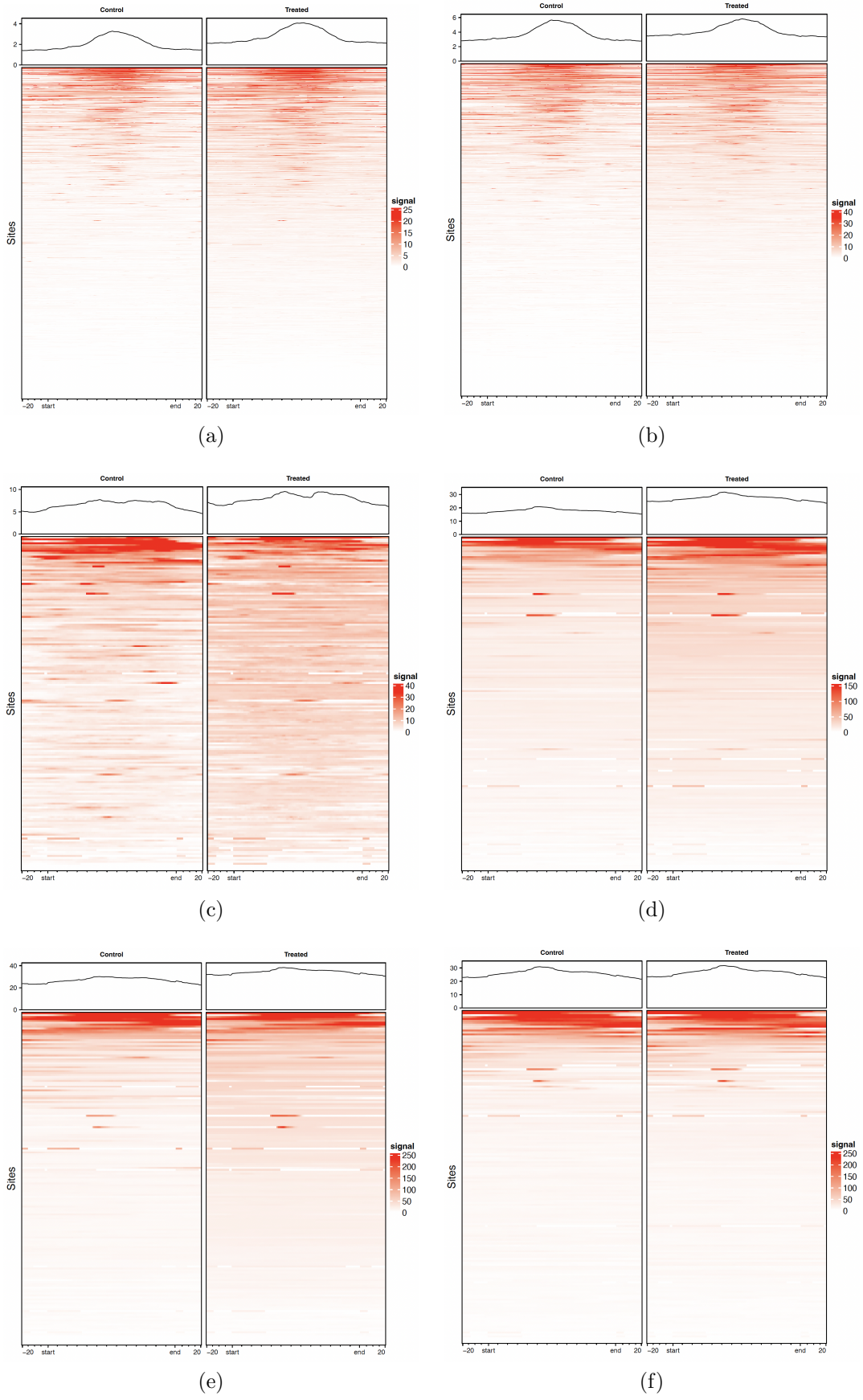
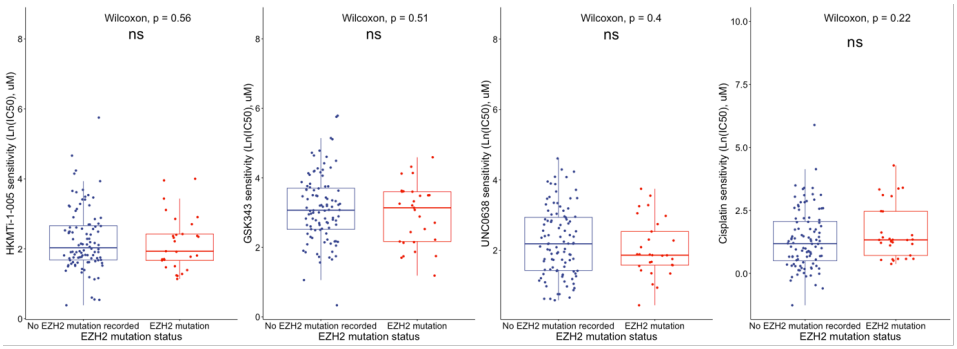
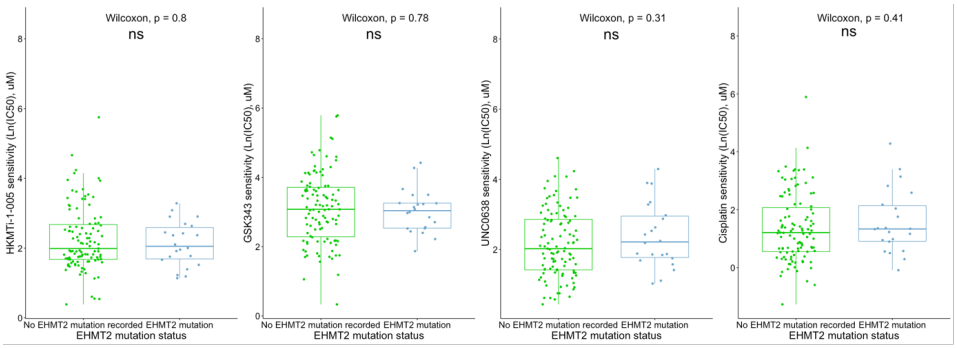


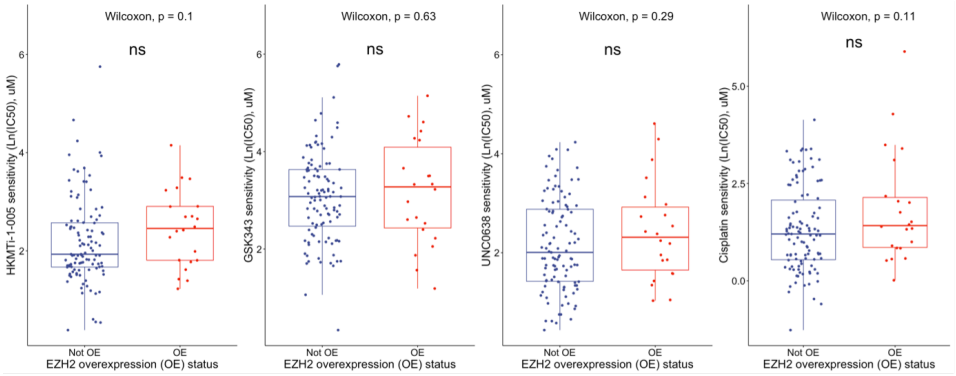
Figure A.15 **Bivalent H3K9me3-H3K4me3 accessibility heatmaps.** Heatmaps generated using DiffBind [390], displaying control versus HKMTi-1-005-treated chromatin accessibility at DNA associated with bivalent H3K9me3-H3K4me3 histone marks. Samples include the (a) PEO1 cell line, (b) PEO4 cell line, (c) ID8 *in vivo* sample, (d) BRCA1 ID8 cell line, (e) BRCA2 ID8 cell line and (f) F3 ID8 cell line.



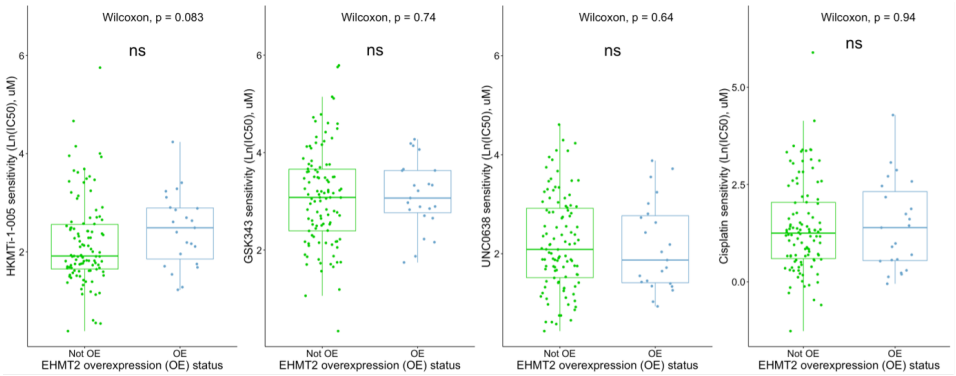
(a)



(b)



(c)



(d)

Figure A.16 **EZH2 or G9a mutation or OE versus sensitivity: haematological cell lines.** Relationship between (a) *EZH2* mutation or (b) *EHMT2* (encoding G9a) mutation, or (c) *EZH2* OE or (d) *EHMT2* (encoding G9a) OE and sensitivity to HKMTi-1-005, GSK343 (EZH2i), UNC0638 (G9ai) and cisplatin. Cell lines included are haematological cell lines from the GDSC2 panel with sensitivity data for each of the four compounds and gene expression data (n=129). OE was defined using a z-score of > 1.

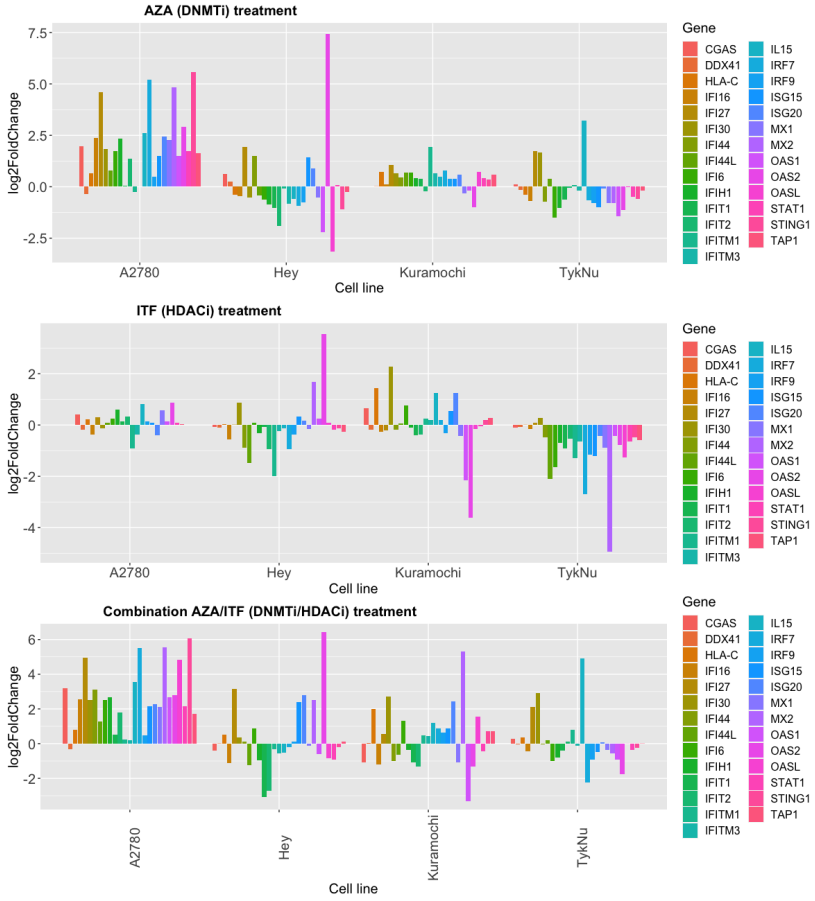


Figure A.17 **Comparison of ISG changes in response to DNMTi, HDACi or combination in 4 ovarian cancer cell lines.** Change in ISGs following AZA (DNMTi, top), ITF (HDACi, middle) or combination (bottom) treatment in ovarian cancer cell lines A2780, Hey, Kuramochi or TykNu. Reproduced based on data from [439] (reproduction of data plotted in supplementary figure S1 of [439]), data are available in GEO, accession GSE182430. Data represent single replicates for each cell line, hence no p-adjusted values are plotted (all p-adj > 0.9).

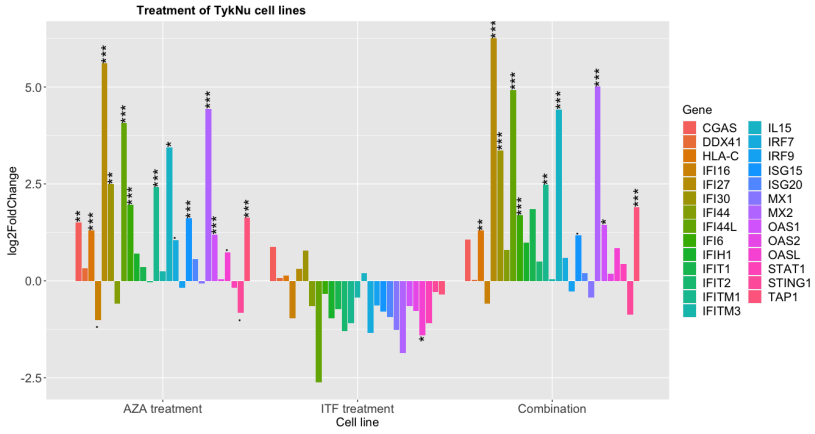


Figure A.18 **Comparison of ISG changes in response to DNMTi, HDACi or combination in the TykNu ovarian cancer cell line.** Change in ISGs following AZA (DNMTi, left), ITF (HDACi, middle) or combination (right) treatment in the TykNu ovarian cancer cell line. Reproduced based on data from [439], data are available in GEO, accession GSE182430. Data represent triplicate RNAseq runs. . = p-adj < 0.1, * = p-adj < 0.05, ** = p-adj < 0.01, *** = p-adj < 0.001.

Geneset	MYC_G9A_REPRESSION_TARGETS	EZH2_REPRESSION_TARGETS	HALLMARK_MYC_TARGETS_V1
INNATE_IMMUNE_RESPONSE	0.23 to 0.35	0.36 to 0.46	-0.06 to 0.06
HALLMARK_INTERFERON_GAMMA	0.19 to 0.31	0.28 to 0.39	0.03 to 0.15

Table A.9 **TCGA-BRCA cohort: 95% confidence intervals for correlation between cMYC-G9a/EZH2 repression geneset scores or canonical cMYC scores and immune geneset scores.** Table showing 95% confidence intervals for correlation between MYC_G9A_REPRESSION_TARGETS scores, EZH2_REPRESSION_TARGETS scores or HALLMARK_MYC_TARGETS_V1 scores and immune geneset scores in the TCGA-BRCA cohort, n= 1095.

Sample	Ens	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	Gene_symbol
PEO1	ENSG00000136997	11349.89382	-0.027248313	0.05084274	-0.5359332	0.59200469	0.77190321	MYC
PEO4	ENSG00000136997	6903.983044	-0.166681916	0.05588388	-2.9826477	0.00285767	0.05093332	MYC
PEA2	ENSG00000136997	1911.948292	-0.46797304	0.12349016	-3.7895572	0.00015092	0.00295403	MYC
OVCAR5	ENSG00000136997	3197.648959	-0.161105204	0.13812805	-1.1663467	0.24347429	0.56218403	MYC
OVCAR8	ENSG00000136997	6863.754066	-1.734595919	0.09207489	-18.838968	3.6195E-79	1.6712E-76	MYC
ID8 in vivo	ENSMUSG00000022346	1230.258203	0.108002433	0.14875491	0.72604283	0.46781252	8.09E-01	Myc

Table A.10 **Changes in expression of *MYC* following HKMTi-1-005 treatment in HGSC samples.**

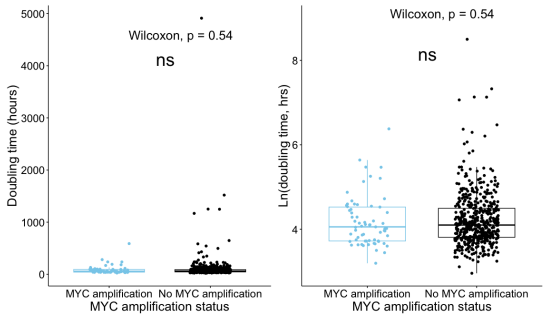


Figure A.19 ***MYC* amplification versus doubling time in the GDSC2 cell line panel.** Wilxocon test showing relationship between *MYC* amplification and doubling time. Doubling time shown in hours (left) and natural log of doubling time in hours (right) for easier visualisation. Cell lines included are those in the GDSC2 panel derived from any solid tumour type, with gene expression data and sensitivity data for each of the four compounds HKMTi-1-005, GSK343, UNC0638 and cisplatin (n=565). * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

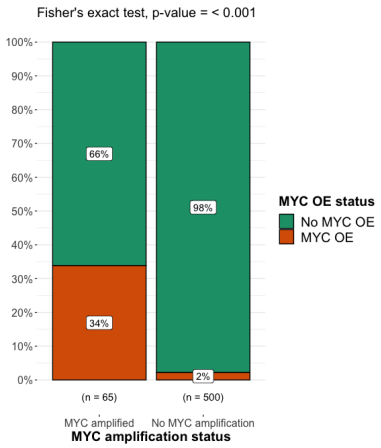


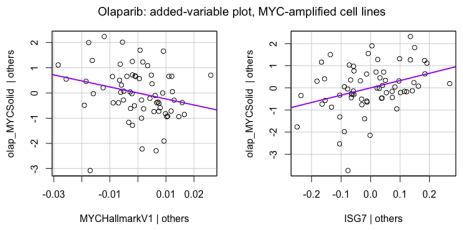
Figure A.20 ***MYC* amplification versus *MYC* OE in the GDSC2 cell line panel: Fisher's test.** Fisher's test comparing the overlap of cell lines in the GDSC2 cell line panel with a *MYC* amplification and those with *MYC* overexpression (as defined by a z-score > 1). Cell lines included are those in the GDSC2 panel derived from any solid tumour type, with gene expression data and sensitivity data for each of the four compounds HKMTi-1-005, GSK343, UNC0638 and cisplatin (n=565).

Geneset	HKMTi-1-005 sensitivity correlation (CIs)
7ISG	0.10 to 0.27
GOBP_INNATE_IMMUNE_RESPONSE	0.10 to 0.27
HALLMARK_INTERFERON_GAMMA	0.09 to 0.27
HALLMARK_INTERFERON_ALPHA	0.09 to 0.26

Table A.11 **95% confidence intervals for correlation between immune geneset scores and sensitivity to HKMTi-1-005.** Table showing the 95% confidence intervals for the correlation between genesets for the 7ISG (derived in chapter 4), innate immune response (INNATE_IMMUNE_RESPONSE), IFNA signalling (HALLMARK_INTERFERON_ALPHA_RESPONSE), IFNG signalling (HALLMARK_INTERFERON_GAMMA_SIGNALLING) and sensitivity to HKMTi-1-005 in *MYC*-amplified and non-amplified cell lines in the GDSC2 cohort. Correlation across cell lines in the GDSC2 panel with sensitivity data for each of the 12 compounds was assessed, including cell lines of any solid tumour type with gene expression data (n=515).

Geneset	MYC-amplified lines (n=55)	Non-amplified lines (n=460)
7ISG	0.17 to 0.62	0.06 to 0.24
INNATE_IMMUNE_RESPONSE	0.19 to 0.63	0.06 to 0.24
HALLMARK_INTERFERON_GAMMA_RESPONSE	0.15 to 0.61	0.06 to 0.24
HALLMARK_INTERFERON_ALPHA_RESPONSE	0.17 to 0.63	0.05 to 0.23

Table A.12 **95% confidence intervals for correlation between immune geneset scores and sensitivity to HKMTi-1-005, stratified by *MYC* status.** Table showing the 95% confidence intervals for the correlation between genesets for the 7ISG (derived in chapter 4), innate immune response (INNATE_IMMUNE_RESPONSE), IFNA signalling (HALLMARK_INTERFERON_ALPHA_RESPONSE), IFNG signalling (HALLMARK_INTERFERON_GAMMA_SIGNALLING) and sensitivity to HKMTi-1-005 in *MYC*-amplified and non-amplified cell lines in the GDSC2 cohort. Correlation across cell lines in the GDSC2 panel with sensitivity data for each of the 12 compounds was assessed, including cell lines of any solid tumour type with gene expression data (n=515: n=55 with *MYC* amplification and n=460 without).

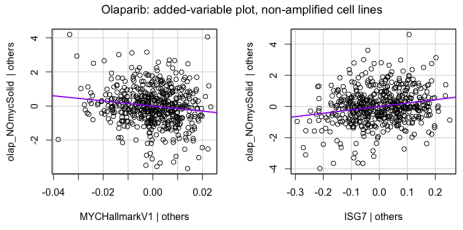


(a)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.209      4.888   2.702  0.00887 **
MYCHallmarkV1 -23.868     11.939  -1.999  0.04998 *
ISG7            3.305      1.145   2.886  0.00536 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.018 on 62 degrees of freedom
Multiple R-squared:  0.2466,    Adjusted R-squared:  0.2223
F-statistic: 10.14 on 2 and 62 DF,  p-value: 0.0001544
```

(b)



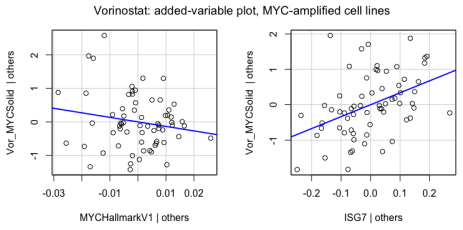
(c)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.0047     1.9088   5.241 2.36e-07 ***
MYCHallmarkV1 -15.0180     4.7072  -3.190  0.00151 **
ISG7            2.1781     0.4807   4.531 7.37e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.127 on 495 degrees of freedom
Multiple R-squared:  0.09476,   Adjusted R-squared:  0.0911
F-statistic: 25.91 on 2 and 495 DF,  p-value: 1.99e-11
```

(d)

Figure A.21 **Multiple linear regression: olaparib.** Multiple linear regression model examining olaparib sensitivity. (a) Added-variable plot and (b) summary of the results of the linear model in *MYC*-amplified cell lines (n=65). (c) Added-variable plot and (d) summary of the results of the linear model in non-amplified cell lines (n=498). Cell lines included are solid lines from the GDSC2 panel with sensitivity data for the four compounds assessed (HKMTi-1-005, UNC0638, olaparib and vorinostat, see text for details). MYCHallmarkV1=canonical cMYC signalling score (score for the HALLMARK_MYC_ TARGETS _V1 geneset). ISG7=score for the 7ISG geneset. Note added-variable plots are also called partial-regression plots.

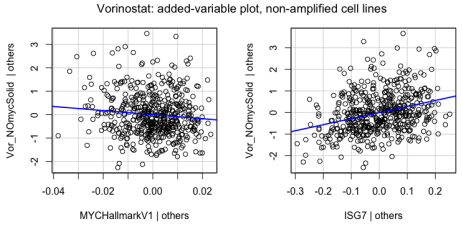


(a)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)     6.3406     3.8978   1.627 0.108875
MYCHallmarkV1 -13.5008     9.5211  -1.418 0.161202
ISG7            3.3746     0.9133   3.695 0.000467 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8119 on 62 degrees of freedom
Multiple R-squared:  0.2771,    Adjusted R-squared:  0.2538
F-statistic: 11.88 on 2 and 62 DF,  p-value: 4.281e-05
```

(b)



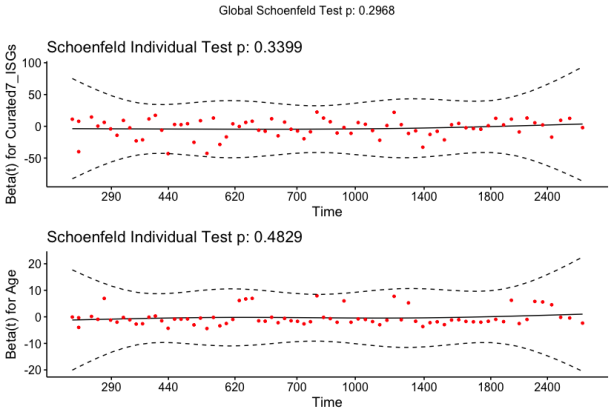
(c)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.7079     1.5615   3.015   0.0027 **
MYCHallmarkV1 -8.7253     3.8506  -2.266   0.0239 *
ISG7           2.8032     0.3932   7.129 3.61e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

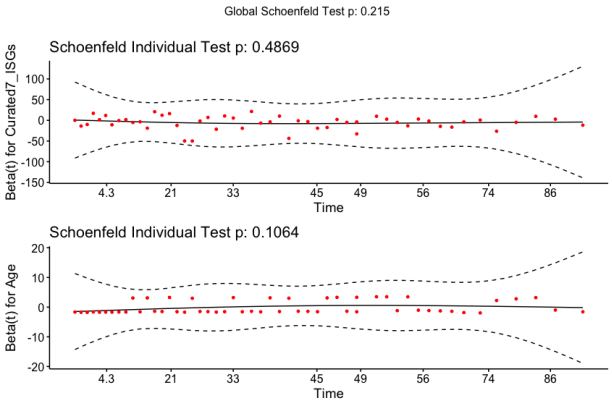
Residual standard error: 0.9222 on 495 degrees of freedom
Multiple R-squared:  0.1452,    Adjusted R-squared:  0.1418
F-statistic: 42.05 on 2 and 495 DF,  p-value: < 2.2e-16
```

(d)

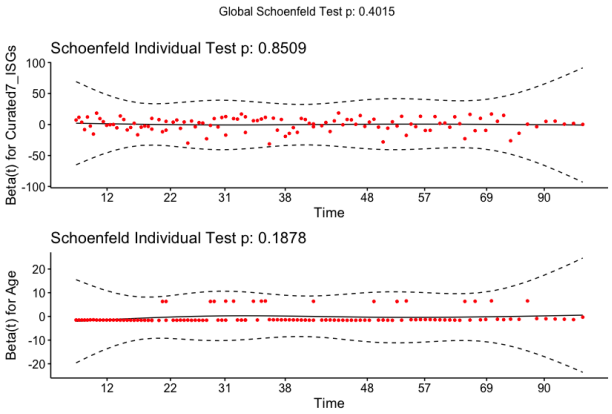
Figure A.22 **Multiple linear regression: vorinostat.** Multiple linear regression model examining vorinostat sensitivity. (a) Added-variable plot and (b) summary of the results of the linear model in *MYC*-amplified cell lines (n=65). (c) Added-variable plot and (d) summary of the results of the linear model in non-amplified cell lines (n=498). Cell lines included are solid lines from the GDSC2 panel with sensitivity data for the four compounds assessed (HKMTi-1-005, UNC0638, olaparib and vorinostat, see text for details). MYCHallmarkV1=canonical cMYC signalling score (score for the HALLMARK_MYC_ TARGETS _V1 geneset). ISG7=score for the 7ISG geneset. Note added-variable plots are also called partial-regression plots.



(a)



(b)



(c)

Figure A.23 **Schoenfeld residuals.** Schoenfeld residuals for (a) the ICGC-OV cohort, (b) the *MYC*-amplified TCGA-OV cohort and (c) the non-amplified TCGA-OV cohort. P values are > 0.05 in all cases, indicating proportional hazards assumptions are satisifed.

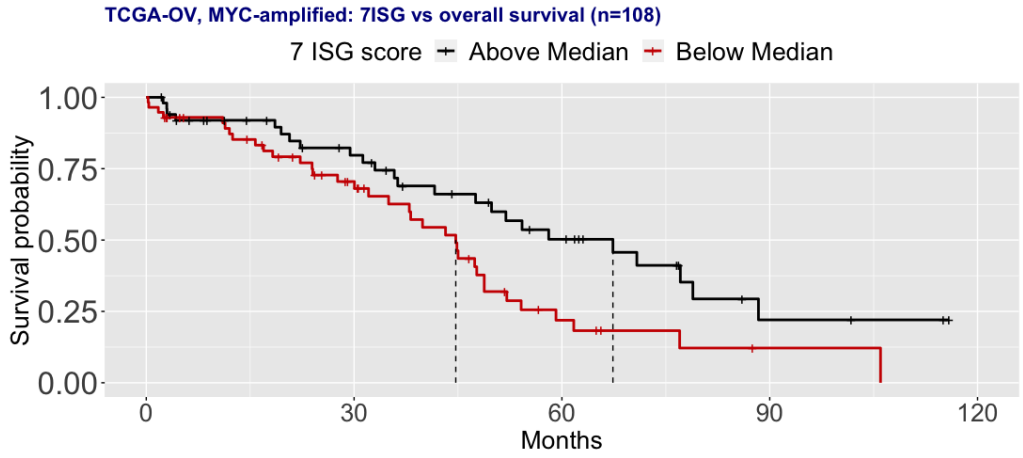


Figure A.24 **7ISG scores versus overall survival in the *MYC*-amplified TCGA-OV cohort.** Kaplan-Meier curve comparing overall survival between those with a high (above median) versus a low (below median) 7ISG score in the *MYC*-amplified TCGA-OV patient cohort. Note survival measured in months.

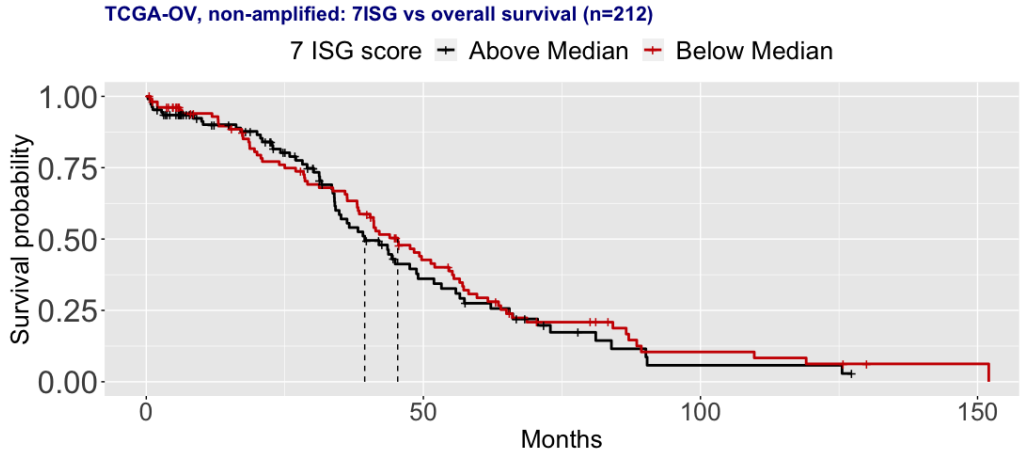


Figure A.25 **7ISG scores versus overall survival in the non-amplified TCGA-OV cohort.** Kaplan-Meier curve comparing overall survival between those with a high (above median) versus a low (below median) 7ISG score in the non-amplified TCGA-OV patient cohort. Note survival measured in months.

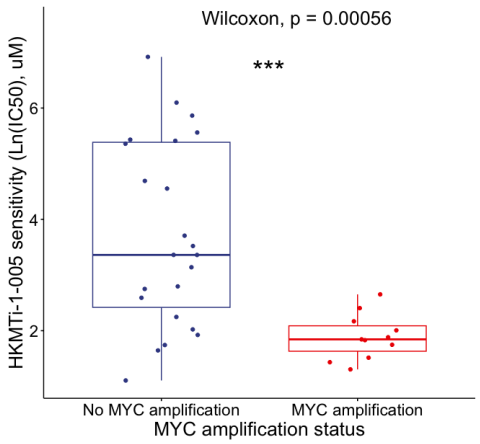


Figure A.26 ***MYC* amplification versus sensitivity in SCLC lines.** Comparison of sensitivity to HKMTi-1-005 between *MYC*-amplified and non-amplified SCLC cell lines in the GDSC2 cell line panel (n=34, of which n=11 are *MYC*-amplified).