Investigation of the oncogenic role of MAF in multiple myeloma

A thesis submitted for the degree of Doctor of Philosophy

Alexia Despoina Katsarou

August 2020

Hugh & Josseline Langmuir Centre for Myeloma Research

Centre of Haematology, Faculty of Medicine

Imperial College London

Statement of originality

I declare that the work presented below is my own, unless otherwise stated, and that any work performed or contributed to by others has been appropriately acknowledged, credited and referenced.

Alexia Katsarou

August 2020

Copyright declaration

The copyright of this thesis rests with the author and licensed under a Creative Commons Attribution Non-Commercial 4.0 International Licence (CC BY-NC). Under this licence, researchers may copy and redistribute the material in any medium or format on the condition that they credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, researchers must ensure they make the licence terms clear to others by naming the licence and linking to the licence text.

Acknowledgments

I would like to extend my sincerest gratitude to my supervisor Prof Anastasios Karadimitris for providing me with the opportunity to join his lab and welcoming me to work alongside his research team. This project, and with it the launch of my development as an academic clinician, would not have been possible without his guidance, experience and mentorship. I would also like to thank the senior members of our team Dr Valentina Caputo and Dr Aris Chaidos who's technical knowledge and advice on study design helped me navigate the difficulties posed by scientific research and opened my mind to new possibilities.

A very big thank you to Dr Nikolaos Trasanidis, for the advice and bioinformatic support he provided me with, without which this project would not have materialised. I am most grateful however for his endless and contagious enthusiasm, his ideas and inspirations, which helped us both persevere and focus, through many long hours and setbacks, cushioning us even at times when our motivation wavered.

I would also like to thank the many members of our team, for their advice, support and camaraderie; Dr Raju Ponnusamy for his guidance, Drs Steven Okoli and Deena Iskander for their sincerest advice, insights, support and humour as well as Dr Ioannis Kostopoulos and our students Foteini, Kieren and Bien for their contribution to this project. A special, cross-country, thank you to previous members of our lab, Drs Katerina Goudevenou and Kyriaki Petevi who guided me through my first uncertain steps in the world of science.

I dedicate this work whilst offering my eternal love and gratitude to my parents and sister, who's endless love and support have been my anchor throughout this and many other endeavours through the years. Their presence in my life has been the brightest guiding light, whose warmth has carried me and will continue carrying me throughout. Finally, a special thank you to my partner Saher who's patience, support and love have enabled us both to flourish in our joint path and provided a scaffold on which to build our future steps.

Funding

This work has been funded by The Kay Kendall Leukaemia Research Fund, the Imperial College Healthcare Charity and Imperial NIHR Biomedical Research Centre.

Abstract

Multiple myeloma (MM) is an incurable malignancy of bone marrow plasma cells (PC), affecting over 5,600 new patients per year in the UK. Translocation of the transcription factor MAF to the *IgH* enhancer, t(14;16), is a myeloma-initiating event in 3-5% of MM cases. As well as in t(14;16), MAF is over-expressed in another 30% of MM, mostly in association with the MMSET t(4;14) translocation. The mechanisms by which aberrant expression of MAF promotes myelomagenesis, including the oncogenic transcriptome it regulates as well as the regulatory regions it impacts are unknown. In this work, with reference to the normal PC, and using both primary myeloma cells and myeloma cell lines, I have applied a multi-omics approach to define the MAF regulome and determine the impact of ectopic MAF expression on MM cells.

First, analysis of the MAF transcriptome in primary myeloma, as compared to normal PC, revealed upregulated genes that are shared with other MM subgroups or are MAF-unique. Having established MAF as a t(14;16) myeloma cell dependency using knock down assays, I obtained and integrated the transcriptome following MAF depletion with the MAF cistrome using RNA-seq and ChIP-seq respectively. This identified 2,112 genes directly bound and regulated by MAF, the majority of which were predicted to be activated by the transcription factor. Intersecting these direct MAF targets with the genes differentially upregulated in MAF primary MM cells, revealed a core transcriptional MAF programme involved in cell cycle regulation and cell to cell interactions such as cell adhesion, migration, cytokine and chemokine signalling pathways.

Identification of the MAF cistrome revealed that MAF preferentially binds enhancers, prompting further exploration of the MAF-associated distal regulatory genome. Compared to normal PC, t(14;16) primary MM cells displayed enhanced chromatin accessibility as assessed by ATAC-seq, with 33% of those regions being over-accessible uniquely in the MAF MM subgroup. Indeed, the chromatin accessibility profile of MAF MM was able to distinctly cluster the samples separately from other genetic MM groups. Remarkably, a quarter of the over-accessible MAF regions were also bound by the MAF (in total 1,661 regions), consistent with

a role of MAF in the regulation of chromatin accessibility. Juxtaposing these regions against chromatin mark maps of the MAF translocated MM.1S cell line and normal PC revealed 44% (726/1,661) of these regions, acquire activatory marks in MAF-translocated cells having been inactive in normal PC and the majority also inactive throughout B cell development. This finding suggested the possible implication of MAF in the activation of these regions. Indeed, following overexpression of MAF in a MAF-negative myeloma cell line, over 50% of these regions acquired enhanced accessibility, demonstrating MAF as an activator of specific regulatory regions in t(14;16) MM. This enabled identification of MAF-activated enhancers, such as in the case of a *CCR1* enhancer peak, validated to be a *CCR1* regulatory region by CRISPRi.

Finally, using ChIP-seq data and chromatin proteomics I find IRF4 to be co-bound in 70-80% of MAF-bound genomic regions, suggesting possible cooperation as well as cross-regulation between the two transcription factors. In addition, several other proteins are identified as candidate partners of the MAF interactome.

In conclusion, this work has provided a comprehensive description of the role of oncogenic MAF in MM and given insights into events related to ectopic MAF expression in PC occurring during the process of myelomagenesis. It has revealed the full spectrum of directly MAF-regulated genes and the consequent cellular pathways affected in MM. In addition, it has established MAF's role in modifying the PC regulatory genome by activating specific regulatory regions, and has enabled the identification of enhancers regulating key MAF-dependent genes, critical for myeloma cell biology.

Table of contents

Chapter 1 Introduction 18

1.1 Multiple Myeloma 19

1.1.1 Definition and Demographics 19

1.1.2 Diagnosis, natural history and clinical features 19

1.1.3 Current treatment and unmet clinical needs 20

1.1.4 Molecular characteristics and cellular origin of multiple myeloma 22

1.2 c-MAF transcription factor 26

1.2.1 Structure and Function 26

1.2.2 MAF: physiological role 29

1.2.3 MAF in the haematopoietic system 30

1.3 MAF as an oncogene 31

1.4 MAF in multiple myeloma 33

1.4.1 Incidence and mechanisms of MAF overexpression in myeloma 33

1.4.2 Clinical features and prognosis of MAF t(14;16) in myeloma patients 34

1.4.3 Molecular profiling of MAF 37

1.5 Aims 41

Chapter 2 Materials and Methods 42

2.1 Human Samples 43

2.1.1 Sample Collection 43

2.1.2 Immunomagnetic CD138-positive cell selection 43

2.1.3 Healthy donor bone marrow mononuclear cell isolation 44

2.1.4 T cell and monocyte depletion of healthy donor bone marrow mononuclear cells 44

2.2 Cell lines 45

2.3 Flow cytometry 46

2.3.1 Surface marker staining for primary myeloma samples 46

2.3.2 Surface marker staining for FACS sorting of primary healthy donor plasma cells 46

2.3.3 Intracellular staining for MAF expression 47

2.3.4 Cell cycle and apoptosis assay 48

2.4 Cloning 48

2.4.1 Production of lentiviral MAF targeting shRNA vectors for knockdown experiments 48

2.4.2 Production of retroviral MIGR1 MAF vector for overexpression experiments 50
2.4.3 CRISPRi lentiviral cloning 51
2.5 Virus production and transduction 55
2.5.1 Lenti- and Retro-viral production 55
2.5.2 Viral transduction 56
2.6 Nucleic acid techniques 57
2.6.1 RNA extraction 57
2.6.2 cDNA synthesis 58
2.6.3 qPCR 58
2.7 RNA sequencing 60
2.8 Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq) 61
2.8.1 Transposase reaction 61
2.8.2 ATAC-seq library preparation 61
2.9 Chromatin Immunoprecipitation (ChIP) 62
2.9.1 Chromatin preparation 62
2.9.2 Immunoprecipitation and DNA extraction 63
2.9.3 DNA library preparation 64
2.9.4 ChIP re-ChIP 64
2.10 Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins (RIME) 66
2.11 Bio-informatic analysis 67
2.11.1 RNA-seq 67
2.11.2 ChIP-seq 68
2.11.3 Epigenomic annotation 69
2.11.4 ATAC-seq 69
2.11.5 Data from other public datasets 70
2.12 Statistical analysis 70
2.13 Other software 70
Chapter 3 MAF regulome in primary myeloma cells and myeloma cell lines 71
3.1 Introduction 72
3.2 Pattern of MAF expression in myeloma and normal PC 73
3.2.1 MAF expression in myeloma cell lines 73
3.2.2 MAF expression in primary myeloma cells 76
3.3 Transcriptome of MAF-translocated primary myeloma PC 79

3.4 Functional validation of MAF targets 85
3.4.1 shRNA MAF knockdown 85
3.4.2 MAF-regulated transcriptional programme 90
3.5 MAF cistrome in MAF-expressing myeloma cell lines 98
3.6 Identification of the MAF directly regulated targets 105
3.6.1 Integration of MAF cistrome with differential gene expression upon MAF knockdown 105
3.6.2 Integration of MAF directly regulated targets with genes differentially expressed in MAF primary samples 108
3.7 Discussion 110
3.7.1 MAF transcriptome profile in primary t(14;16) MM 110
3.7.2 Effects of MAF depletion 112
3.7.3 Defining the MAF cistrome 114
3.7.4 Predicting the direct gene targets of MAF 115
3.7.5 The MAF core transcriptional programme 116
Chapter 4 MAF associated regulatory genome 119
4.1 Introduction 120
4.2 Chromatin accessibility in primary MAF-myeloma plasma cells 121
4.3 Defining the MAF regulatory genome 127
4.3.1 Defining the "MAF-associated" regions 127
4.3.2 "MAF-associated" regions across development 129
4.4 MAF as an activator of specific regulatory regions 133
4.5 Validation of the *CCR1* MAF-activated enhancer 141
4.6 Identification and characterisation of the *CCND2* enhancer 152
4.7 Discussion 156
4.7.1 Chromatin accessibility profile in MAF-translocated MM 157
4.7.2 MAF as an activator of specific regulatory regions 158
4.7.3 The example of the *CCR1* enhancer, a validated "MAF-activated" region 160
4.7.4 Limitations of using a myeloma cell line 161
4.7.5 Possible mechanisms underpinning MAF-induced epigenetic activation 163
4.7.6 Other MAF regulatory regions – the *CCND2* enhancer 165
Chapter 5 MAF cooperating factors 167
5.1 Introduction 168
5.2 IRF4 co-occupancy 169

5.3 IRF4 and MAF self- and cross-regulation 177
5.4 Interaction of MAF with IRF4 and other cooperating factors investigated by chromatin proteomics 180
5.5 Discussion 186
Chapter 6 Conclusions and Future work 191
References 197
Appendices 205
Appendix A 205
Appendix B 208
Appendix C 214
Appendix D 223
Appendix E 231

List of Figures

Figure 1.1 Schematic of B cell development 23
Figure 1.2: Translocations in multiple myeloma 24
Figure 1.3: Myeloma disease evolution 25
Figure 1.4: c-MAF isoforms 27
Figure 1.5: c-MAF structure 27
Figure 1.6: MAF t(14;16) translocation 33
Figure 1.7: Survival per molecular myeloma subgroup 35
Figure 2.1: pLKO.1 vector used for MAF shRNA-mediated knockdown 50
Figure 2.2: Map of inducible lentiV2 dCAS9 KRAB vector, used for CRISPRi experiments 52
Figure 3.1: MAF and target gene mRNA expression in MMCLs 74
Figure 3.2: MAF protein expression in MMCLs 75
Figure 3.3: Pre- and post-selection flow cytometry of t(14;16) myeloma sample 76
Figure 3.4: Healthy donor plasma cell flow cytometry 77
Figure 3.5: MAF expression in primary myeloma plasma cells 78
Figure 3.6: Differential gene expression in MAF primary myeloma 80
Figure 3.7: Gene set enrichment analysis of MAF upregulated genes in primary samples 82
Figure 3.8: Hierarchical clustering on differentially expressed genes in myeloma PC 83
Figure 3.9: Comparison of MAF differentially expressed genes across myeloma genetic subgroups . 84
Figure 3.10: Viability of cells transduced with MAF shRNA 86
Figure 3.11: MAF and target gene expression upon MAF knockdown 87
Figure 3.12: MAF protein expression upon MAF knockdown 88
Figure 3.13: Cell cycle analysis upon MAF knockdown 89
Figure 3.14: Differential gene expression upon MAF knockdown 92
Figure 3.15: Comparison of anti-MAF shRNA downregulated genes with MF subgroup public datasets 93
Figure 3.16: Pathway enrichment analysis of downregulated genes upon MAF knockdown 94
Figure 3.17: Gene expression analysis following MAF knockdown in JJN3 cells 96
Figure 3.18: MAF transcriptional signature identified following MAF knockdown in MM.1S and JJN3 cells 97
Figure 3.19: Peak distribution of MAF binding across three myeloma cell lines 99
Figure 3.20: MAF binding in MM.1S 100
Figure 3.21: Histone mark maps corresponding to MAF binding sites 101
Figure 3.22: MAF ChIP-seq in JJN3 102
Figure 3.23: MAF ChIP-seq in H929 103
Figure 3.24: MAF binding on target genes 104
Figure 3.25: Integration of MAF cistrome with transcriptome following MAF depletion 106
Figure 3.26: MAF directly regulated targets 107
Figure 3.27: Pathway enrichment analysis on MAF activated genes 108
Figure 3.28: MAF core gene targets 109

Figure 3.29: Diagram of analysis workflow followed to identify MAF core regulated genes 118
Figure 4.1: Chromatin accessibility in MAF primary samples 122
Figure 4.2: Enrichment analysis of MAF differentially accessible regions 124
Figure 4.3: Comparison of differentially accessible regions across myeloma genetic subgroups 125
Figure 4.4: Hierarchical clustering on differential chromatin accessibility in myeloma PC 126
Figure 4.5: MAF bound and differentially open regions, the "MAF-associated regions" 128
Figure 4.6: Characteristics of the "MAF-associated" regions 129
Figure 4.7: MAF expression across normal haemopoietic cells 131
Figure 4.8: Comparative chromatin states of "MAF-associated" regions across B cell development 131
Figure 4.9: Putative MAF-activated regions 132
Figure 4.10: MAF overexpression in U266 cells 135
Figure 4.11: Transcriptome of MAF-overexpressing U266 136
Figure 4.12: Comparative analysis of MAF overexpression in U266 with MAF transcriptome 138
Figure 4.13: Accessibility of MAF-overexpressing U266 139
Figure 4.14: Validated MAF-activated promoters and enhancers 140
Figure 4.15: *CCR1* predicted enhancer 142
Figure 4.16: CRISPRi schematic representation 143
Figure 4.17: CCR1 putative enhancer sgRNA target sites 145
Figure 4.18: *CCR1* expression following CRISPRi targeting of enhancer peaks 146
Figure 4.19: *CCR1* expression across B cell maturation 147
Figure 4.20: *CCR1* accessibility and expression across MM subgroups 148
Figure 4.21: *TLR4* accessibility peak 151
Figure 4.22: *CCND2* enhancer 154
Figure 4.23: *CCND2* expression across myeloma subgroups 155
Figure 4.24: Schematic of epigenomic analysis and proposed MAF activatory function 166
Figure 5.1: IRF4 motif the second most enriched motif in MAF translocated cells 171
Figure 5.2: MAF and IRF4 cistrome 172
Figure 5.3: IRF4 binding on key MAF target genes 173
Figure 5.4: IRF4 binding in MAF-associated regions 174
Figure 5.5: Binding of IRF4 on *CCR1* enhancer 175
Figure 5.6: Binding of IRF4 on *TLR4* enhancer 175
Figure 5.7: Distance between MAF and IRF4 binding 176
Figure 5.8: ChIP re-ChIP assays for MAF and IRF4 177
Figure 5.9: IRF4 binding on *MAF* promoter 178
Figure 5.10: MAF binding on *IRF4* enhancer 179
Figure 5.11: *IRF4* expression upon MAF knockdown 179
Figure 5.12: *IRF4* expression across MM subgroups 180
Figure 5.13: Schematic of Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins (RIME) 181
Figure 5.14: MAF RIME results 183
Figure 5.15: String network of MAF interactors 184

Figure 5.16: Biological functions of MAF interactors.....184
Figure 5.17: Molecular functions of MAF interactors.....185

List of Tables

Table 2.1: Monoclonal antibodies used in flow cytometry panels 47
Table 2.2: shRNA sequences (5'-3') used for MAF knockdown experiments 49
Table 2.3: Primers used to for pLKO.1-eGFP and MIGR1 cloning 51
Table 2.4: sgRNAs used in CRISPRi experiments 54
Table 2.5: Constituents of transfection mixture for lentiviral and retroviral production 56
Table 2.6: List of Taqman probes used 59
Table 2.7: List of primer sequences used ChIP qPCR 60
Table 2.8: Antibodies used in ChIP experiments 64

List of publications

Katsarou, A., et al., Oncogenic MAF in Co-Operation with IRF4 Confers Extensive Chromatin Re-Arrangement in Plasma Cells and Generates 'Neo-Enhancers' That Regulate Genes Critical for Myeloma Biology. Blood, 2019. 134(Supplement_1): p. 3783-3783

Trasanidis*, N., Alvarez-Benayas*, J., Katsarou*, A., et al., Distinct Chromatin Accessibility Changes, Aberrant Transcription Factor Networks Combined with Novel Oncogenic Enhancers Characterise Myeloma-Initiating Genetic Events. Blood, 2019. 134(Supplement_1): p. 1769-1769.

Trasanidis, N., Katsarou, A., et al., PBX1 Co-Operates with FOXM1 to Regulate Myeloma Cell Proliferation and to Define an Ultra High-Risk chr1q Gain Myeloma Patient Subgroup. Blood, 2019. 134(Supplement_1): p. 3760-3760.

List of Abbreviations

ABLIM1	Actin Binding LIM Protein 1
AITCL	Angioimmunoblastic T cell lymphoma
APOBEC	Apolipoprotein B mRNA editing enzyme catalytic subunit
ATAC-seq	Assay for Transposase-Accessible Chromatin followed by sequencing
ATP13A2	ATPase type 13A2
BM	Bone marrow
bp	Base pairs
BSA	Bovine serum albumin
bZIP	Basic leucine zipper domain
CCND1	Cyclin D1
CCND1 MM	CCND1-translocated multiple myeloma
CCND2	Cyclin D2
CCND3	Cyclin D3
CCR1	C-C Motif Chemokine Receptor 1
ChIP	Chromatin immunoprecipitation
ChIP-qPCR	ChIP followed by real time polymerase chain reaction
ChIP-seq	ChIP followed by sequencing
CLL	Chronic Lymphocytic Leukaemia
CRISPR(i)	Clustered regularly interspaced palindromic repeats (inhibition)
CRUK	Cancer Research UK
CSR	Class switch recombination
CX3CR1	C-X3-C Motif Chemokine Receptor 1
dCAS9	Dead CAS9
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DPYSL2	Dihydropyrimidinase-like 2
DTT	Dithiothreitol
EDTA	Ethylenediamine Tetraacetic Acid
ENCODE	Encyclopaedia of DNA Elements Consortium Project
EV	Empty vector
FACS sorting	Fluorescence Activated Cell Sorting
FBS	Fetal Bovine Serum
FC	Fold change
FISH	Fluorescence in situ hybridisation
FLI1	Friend Leukaemia Integration 1 transcription factor
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GC	Germinal centre
GEP	Gene expression profiling
GFP	Green Fluorescent Protein
GSEA	Gene Set Enrichment analysis software
H3K4me1	Monomethylation of histone H3 at the 4^{th} lysine residue
H3K4me3	Trimethylation of histone H3 (4^{th} lysine residue)
H3K27ac	Acetylation of histone H3 (27^{th} lysine residue)
HMGB2	High mobility group protein B2
HBS	HEPES-buffered saline
HY	Hyperdiploid myeloma
IgH	Immunoglobulin heavy chain
IgL	Immunoglobulin light chain
IGV	Integrative Genomes Browser
IMDM	Iscove's Modified Dulbecco's Medium
IRF4	Interferon regulatory factor 4
ITGB7	Integrin β7
Kb	Kilobases
KRAB	Krüppel associated box domain

LCP2	Lymphocyte Cytosolic Protein 2
MARE	MAF recognition element
mB	Memory B cell
MF group	MAF molecular group as identified in GEP studies
MFI	Median fluorescence intensity
MGUS	Monoclonal Gammopathy of Undetermined Significance
MM	Multiple myeloma
MMCL	Multiple myeloma cell line
MMRF	Multiple Myeloma Research Foundation
MMSET	Nuclear Receptor Binding SET Domain Protein 2
MMSET MM	MMSET-translocated multiple myeloma
ND	Healthy donor normal plasma cells
NS	Non-significant
NTRK2	Neurotrophic tyrosine kinase receptor type 2
NUAK1	Nuak family kinase 1
PAM	Protospacer Adjacent Motif
PBS	Phosphate Buffered Saline
PC	Plasma cells
PI	Propidium Iodide
PMSF	Phenylmethylsulfonyl fluoride
PRDM1	PR Domain Containing 1, With ZNF Domain
PRMT1	Protein Arginine Methyltransferase 1
PYGB	Glycogen phosphorylase brain form
qPCR	Quantitative real time PCR
RASSF4	Ras Association Domain Family Member 4
RIME	Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins
RIN	RNA integrity number
R-ISS	Revised International Staging System
RNA-seq	RNA sequencing
RND3	Rho Family GTPase 3
RPMI1640	Roswell Park Memorial Institute Medium
RT	Room temperature
scrbl	Scrambled control
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SEM	Standard Error of the Mean
sgRNA	single guide RNA
shRNA	Short Hairpin RNA
SYNDIG1	Synapse Differentiation Inducing 1
TF	Transcription factor
Tfh	T follicular helper cells
Th cells	T-helper cells
TLR4	Toll-like receptor 4
TSS	Transcription start sites
VCAM-1	Vascular cell adhesion protein 1
XBP1	X-Box Binding Protein 1

Chapter 1

Introduction

1.1 Multiple Myeloma

1.1.1 Definition and Demographics

Multiple myeloma (MM) is an incurable malignancy, characterised by expansion of clonal plasma cells (PC), the terminally differentiated immunoglobulin (Ig)-secreting B cells resident in the bone marrow. The disease accounts for 2% of all new cancer diagnoses, having over 5,600 new cases per year in the UK (CRUK data, 2014-2016), with a continuously rising incidence over the past decades[1]. MM is more common in individuals of Afro-Caribbean decent and slightly commoner in men than in women. Although the disease prognosis has improved over the years, it remains disappointing; patients have a 47% 5-year overall survival, with MM accounting for over 3,000 annual deaths in the UK[1].

1.1.2 Diagnosis, natural history and clinical features

Multiple myeloma is defined by the presence of ≥10% PC bone marrow infiltration and the presence of end organ damage, as defined by the CRAB features that classically characterise the disease: hypercalcemia, renal dysfunction, anaemia and bone lesions[2]. More recently, the International Myeloma Working Group have revised the diagnostic criteria to include heavy bone marrow involvement with ≥60% clonal PC cell infiltration, marked elevation of the involved light chain (involved to uninvolved serum free light chain ratio ≥100) and ≥1 bone lesion seen on magnetic resonance imaging[3]. These features, despite the absence of clinical symptoms, are associated with an 80% risk of disease progression in 2 years, and therefore, including them in the myeloma defining events enabled patients to be treated prior to the development of clinically overt organ damage.

Multiple myeloma has a well characterised natural history. It is preceded by Monoclonal Gammopathy of Undetermined Significance (MGUS), a pre-malignant clonal disease of the plasma cells, characterised by PC infiltration of the bone marrow of <10% and a serum

monoclonal immunoglobulin <30g/l in the absence of any clinical symptoms[4]. MGUS affects 3% of the population over the age of 50, rising to 5% over 70 years of age[5], and carries an overall risk of progression of roughly 1% per year. Smouldering myeloma, the asymptomatic pre-malignant phase of the disease, is defined by a 10-60% PC infiltration, and carries a much higher risk of progression, in the order of 10% per year in the first 5 years[6]. In the final stages, multiple myeloma can progress to plasma cell leukaemia, with circulating tumour cells in the peripheral blood (≥20% of circulating leukocytes)[4], or extramedullary disease, where clonal plasma cells disseminate to form tumours at multiple sites distant to the bone marrow[7].

The clinical features associated with MM are numerous and can significantly impact the quality of life of patients. Renal dysfunction arises from light chain cast nephropathy, the deposition of light chains in the renal tubules, whereas anaemia, leading to fatigue and lack of energy, is the result of a combination of renal impairment and bone marrow infiltration by the clonal PC[4]. Other cytopenias, including neutropenia and thrombocytopenia, can also result from bone marrow disease infiltration. Hypercalcaemia and osteolytic bone lesions, leading to skeletal fractures and pain, sometimes intractable, are a result of plasma cell-induced imbalance between the activity of osteoblasts and osteoclasts. Other clinical features include susceptibility to infections, secondary to neutropenia and hypogammaglobulinemia, and occasionally compression symptoms, such as in the spinal cord, resulting from plasmacytomas, the extramedullary tumours formed by clonal PC[4].

1.1.3 Current treatment and unmet clinical needs

In the past, the mainstay of MM treatment included corticosteroids, alkylating agents (melphalan, cyclophosphamide) and anthracyclines (doxorubicin), whereas in the late 1990s the first immunomodulatory agent thalidomide was trialled[8]. Since the advent of thalidomide, our armoury of drugs against myeloma has expanded significantly, to include proteasome inhibitors, such as bortezomib, carfilzomib and ixazomib, 2nd and 3rd generation immune modulators, such as lenalidomide and pomalidomide, monoclonal antibodies such

as elotuzumab (anti-SLAMF7; CD319), isatuximab and daratumumab (anti-CD38), panobinostat, a histone deacetylase inhibitor, and venetoclax, a Bcl2 inhibitor[9].

Patients today are typically treated with a triplet agent combination, such as VD (bortezomib, dexamethasone), combined with thalidomide, lenalidomide or cyclophosphamide (VTD, VRD, VCD)[9], with less intense regimens reserved for frail patients. In fit patients, this is followed by melphalan-conditioned autologous haematopoietic stem cell transplantation and maintenance treatment. A multitude of options are becoming available at relapse, depending on funding arrangements in each country and availability of clinical trials, such as combinations with carfilzomib (KRd, KCd) daratumumab (D-VCd, D-Vd, D-Rd, D-VMP), pomalidomide (PVd, KPd), as well as izaxomib and elotuzumab combinations[10]. These treatments have led to an ongoing improvement in patients' prognosis[11]. In addition, after the discovery of Chimeric Antigen Receptor-T (CAR-T) cells for the treatment of B-cell malignancies, immune therapies, such as CAR-T cells and bi-specific T cell engagers, have also entered early-phase studies in MM, offering some hope to multiply relapsed patients[10].

Despite recent advances in therapeutics, MM is still incurable, ultimately becoming a terminal disease, with patients having a median overall survival of 6-7 years[9]. The Revised International Staging System (R-ISS) was designed to help clinicians identify patients who have poor risk disease, and who would benefit the most from intensive treatments[12]. This stratifies the patients into three groups (R-ISS I-III) according to their levels of lactate dehydrogenase, albumin, β2-microglobulin and presence or not of high-risk cytogenetic features (del(17p), translocation t(4;14) or t(14;16); see section 1.1.4 for further details). Patients in the high-risk group had a 5-year overall survival of just 40%. Therefore, there is an urgent clinical need to improve the outcomes in this cohort of patients. In addition, treatments with curative potential remain elusive and are the subject of intense drug discovery research.

1.1.4 Molecular characteristics and cellular origin of multiple myeloma

During B cell development, progenitor B cells in the bone marrow, undergo immunoglobulin heavy (*IgH*) and light (*IgL*) chain gene rearrangement, by recombining the V, D and J sections of the genes in order to form a battery of diverse functional B cell receptors[13] (**Fig. 1.1)**. These mature naïve B cells enter the circulation and upon antigen stimulation, develop into short lived IgM plasma cells or enter the germinal centre of the lymph node, where they undergo extensive proliferation and somatic hypermutation as centroblasts. The process of somatic hypermutation, involving multiple point mutations, insertions and deletions in the variable segments of the immunoglobulin genes, increases the range of the cells' affinity to the antigen[14]. Centrocytes are subsequently selected on the basis of high antigen affinity and undergo IgH class switching (class switch recombination, CSR), changing their expression of the *IgH* constant region, in order to produce different classes of antibodies[15]. Recently however, a study has suggested that CSR may mostly occur in the first few days post antigen stimulation, prior to the B cells entering the germinal centre[16]. Post germinal centre B cells are destined to either become memory B cells or differentiate into plasmablasts and long-lived PC, migrating to the bone marrow where they reside. This process of PC differentiation requires the activation of molecular pathways controlled by the upregulation of transcription factors critical for PC biology such as IRF4 and PRDM1/BLIMP1, which in turn downregulate PAX5 and upregulate XBP1[15].

Multiple myeloma arises from post germinal centre bone marrow PC, which have previously undergone a succession of somatic hypermutation and IgH class switching events as germinal centre B cells[14, 17]. Programmed double stranded DNA breaks mediate these genomic rearrangements, making developing B cells vulnerable to mutations. Indeed, primary genetic events occurring in MM can divide the disease roughly in two groups. Half of MM cases are

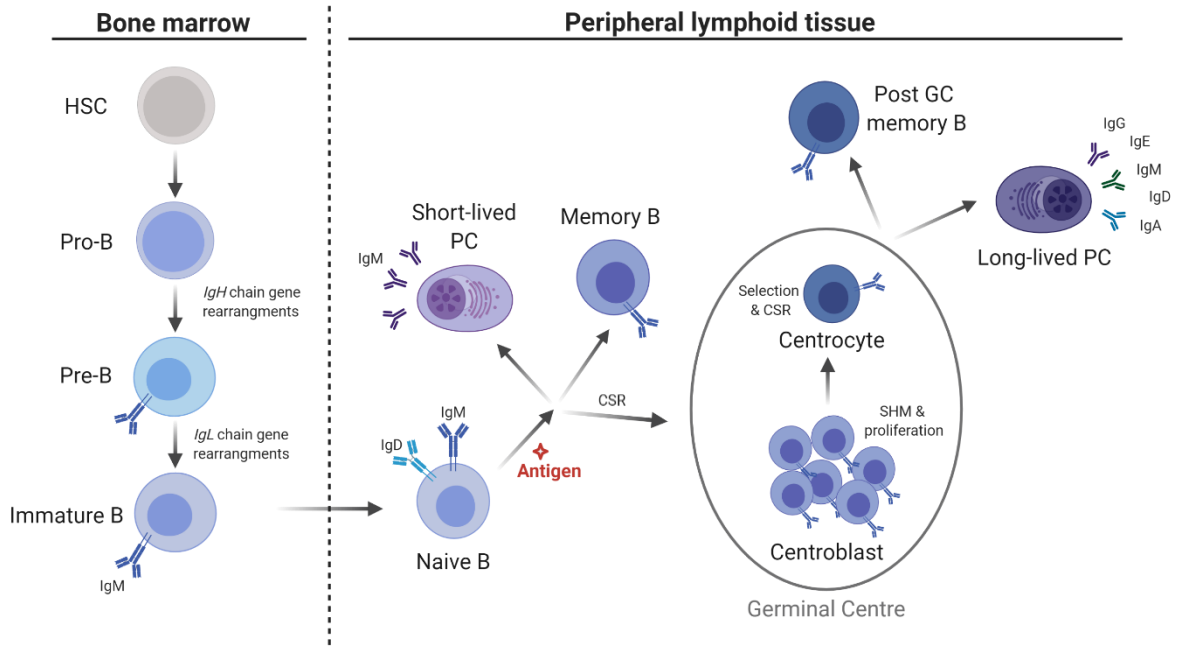


Figure 1.1 Schematic of B cell development

Schematic depicting normal B cell development and maturation. Immature B cells develop in the bone marrow following immunoglobin gene rearrangement. Naïve mature B cells upon antigen stimulation enter the germinal centre (GC), where they undergo somatic hypermutation (SHM) and class-switch recombination (CSR). Subsequently, post-GC B cells develop into memory B cells or long-lived plasma cells (PC).

associated with hyperdiploidy, involving trisomies of odd numbered chromosomes, such as 3, 5, 7, 9, 11, 15, 19 and 21. Conversely, oncogene translocations to the *IgH* (14q32) enhancer locus, and less frequently to *IgL* (κ, 2p11 or λ, 22q11) are implicated in 30-50% of cases, and involve *FGFR3* and *MMSET* [t(4;14); 11-15%], *CCND3* [t(6;14); <1%], *CCND1* [t(11;14); 14-15%], *c-MAF* [t(14;16), 3-5%] and *MAFB* [t(14;20); 1.5%][18-20] (**Fig. 1.2**). Break points, thought to occur during class switch and somatic hypermutation, occur between the JH and switch regions of the immunoglobulin gene and juxtapose the oncogenes to the powerful *IgH* enhancers[14, 18, 20]. The translocations mentioned above can also be found at the premalignant MGUS stage of the disease in up to 50% of cases[14, 21]. Given the timing of their occurrence and their presence in MGUS, these genetic alterations are considered myeloma-initiating events, consistently and stably present throughout the disease and in all subclones[19, 22] (**Fig. 1.3**).

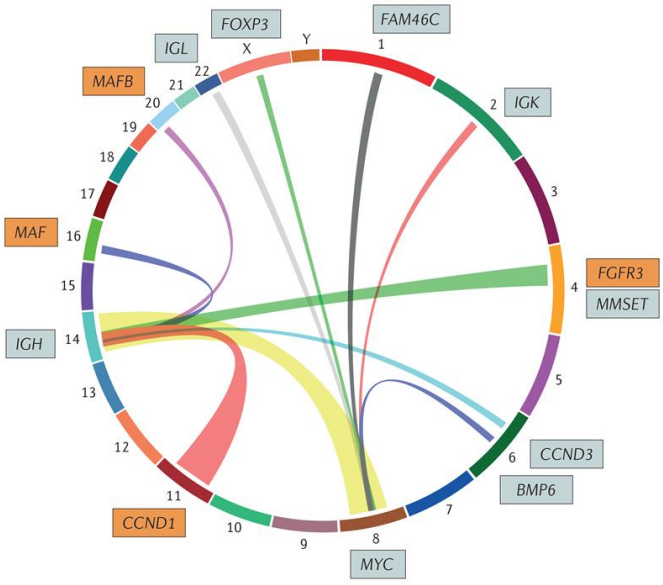


Figure 1.2: Translocations in multiple myeloma

Circus plot depicting translocations found in MM. Genes sometimes also found to be mutated in orange (From Manier, S., et al., Nat Rev Clin Oncol, 2017. **14**(2): p. 100-113[23]).

Following the initial genetic event, mutations affecting various molecular pathways are acquired, transforming the disease into a fully malignant entity (**Fig. 1.3**). Secondary translocations such as the *MYC* translocation t(8;14) are found in 20% of patients and copy number alterations such as gain (1q), del (1p) and del (17p), occur in 30-40%, 20-30% and 10% of patients respectively[18, 22]. Furthermore, myeloma is characterised by an average of 1.6 mutations per Mb, with each patient on average having 5 detectable subclones[23]. Whole genome and exome sequencing studies conducted on large cohorts of patients have identified numerous recurrent mutations, including mutations affecting the MEK-ERK signalling pathway (KRAS, NRAS, BRAF), the NF-kB pathway (CYLD, TRAF3), cell cycle regulation (RB1, CDKN2C), epigenetic regulators (HIST1H1E, ARID1A , TET2, KDM5C, DNMT3A, IDH1, and IDH2) and tumour suppressors such as TP53 and DIS3[18, 22-25]. Genetic aberrations accumulate with disease progression, with successive rounds of treatments providing a survival advantage to the resistant subclones, which in turn are further supported by a permissive bone marrow microenvironment, in a process that has been termed "branching evolution"[22, 23]. Transcriptome profiles from single cell RNA-seq studies in multiple myeloma further exemplify intra-tumour diversification and complexity[26].

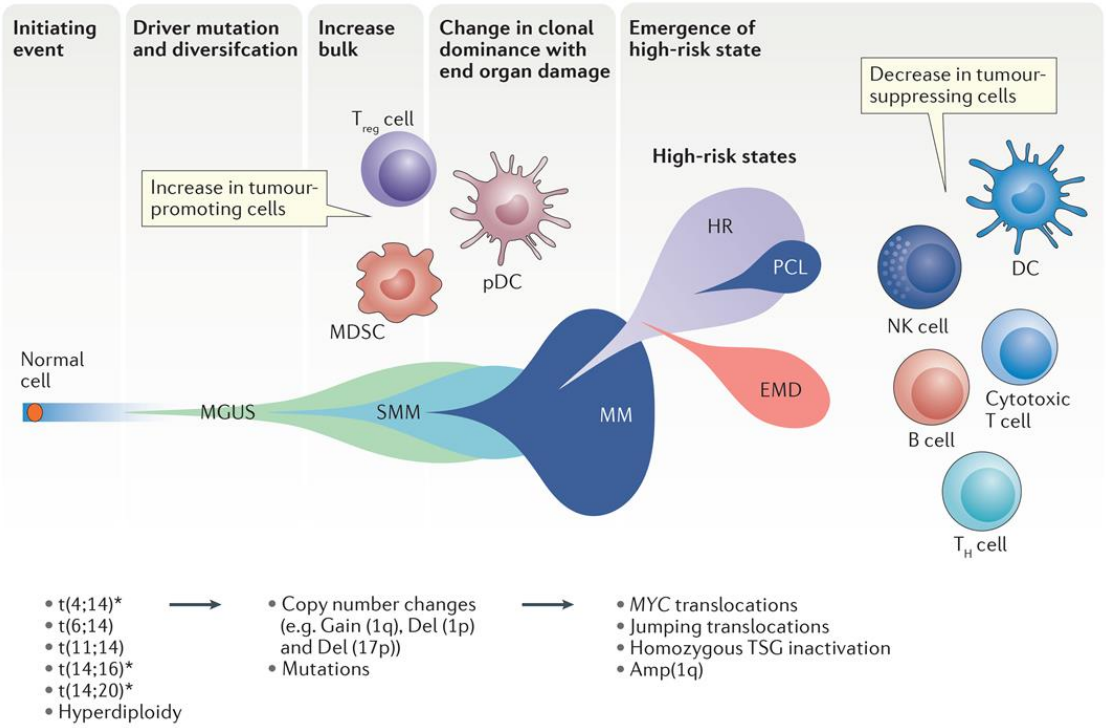


Figure 1.3: Myeloma disease evolution

Schematic depicting multiple myeloma disease evolution with successive acquisition of molecular events. (Adapted from Pawlyn, C. and Morgan, G.J. Nat Rev Cancer, 2017. 17(9): p. 543-556[22])

Multiple myeloma is therefore a disease of extensive clonal heterogeneity and clonal selection upon exposure to various treatments eventually leads to drug resistant disease[18, 23, 25]. As a result, the prospect of targeted therapy against a particular mutation or genetic driver being effective in the long term is challenging, with emerging subclones presenting an ever-moving target. Effective treatment would instead either include a combination of therapies directed at all the subclones present at a given time, or instead target a lesion present throughout the clones[25]. Primary genetic events exist stably in the entirety of the myeloma clone, thus presenting themselves as ideal targets for precision medicine. The c-MAF translocation, t(14;16), is one of the primary disease-initiating genetic events, and hence is the focus of this project.

1.2 c-MAF transcription factor

1.2.1 Structure and Function

c-MAF is a member of the Muscular Aponeurotic Fibrosarcoma, MAF, group of transcription factors (TF), belonging to the Activated Protein 1 (AP1) family of basic leucine zipper domain (bZIP) proteins, which also include JUN, FOS and ATF families. The first MAF to be described was the avian v-MAF in 1989, following isolation of an oncogenic retroviral gene in chicken fibrosarcoma[27]. Since that time, 7 transcription factors of the same family have been described in humans, divided based on their size and structure into the large MAFs: MAFA, MAFB, c-MAF, NRL (Neural retinal leucine zipper), and small MAFs: MAFG, MAFK, MAFF[27].

c-MAF is encoded on chromosome 16 (16q23.2) and has two known isoforms produced via alternative splicing: a short isoform consisting of 373 amino acids (38kDa) and a long isoform consisting of 403 amino acids (42kDa) (**Fig. 1.4**). A third variant, computationally predicted but not experimentally validated consists of 383 amino acids. Large MAFs, including c-MAF, have a basic domain, which binds DNA in the major groove and an Extended Homology Region (EHR) which allows them to bind longer DNA sequences than other bZIP transcription factors[28] (**Fig. 1.5**). The leucine zipper, located at the C-terminus, contains a leucine residue every seventh amino acid, and facilitates homo- or heterodimerisation with other proteins[28, 29], stabilising DNA binding. At the N-terminus c-MAF contains an acidic transactivation domain, which is not present in small MAFs, and permits transcriptional activation by recruiting activating complexes to the genome[28, 30]. Between the transactivation domain and the extended homology region, lies a linker domain, rich in glycine and histidine residues[30], allowing the protein flexibility[31].

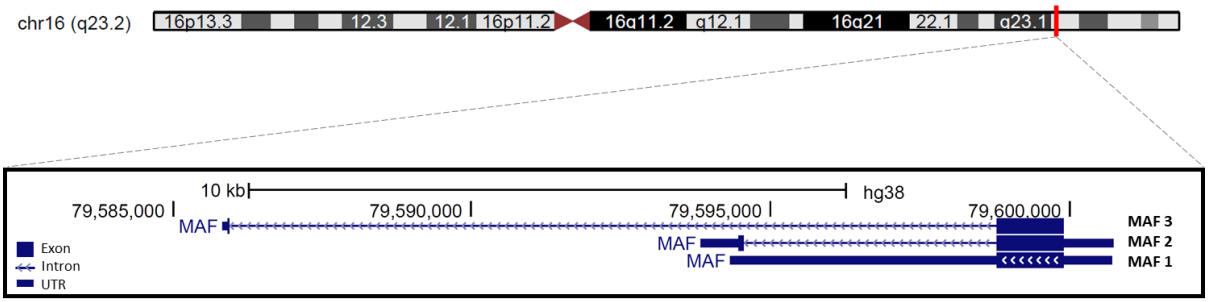


Figure 1.4: c-MAF isoforms

Schematic depicting the three c-MAF isoforms: the short isoform with one exon (MAF 1), long isoform with two exons (MAF 2) and a computationally derived isoform (MAF 3). Arrows indicate direction of transcription. Schematic created using the UCSC (University of California, Santa Cruz Genomics Institute) browser, using the GRCh38/hg38 genome.

c-MAF binds DNA by recognising a 13-14 bp palindromic sequence termed MARE, MAF recognition element, with the presence of the extended homology region on c-MAF allowing it to be longer than the recognition sites of other bZIP transcription factors. c-MAF's basic domain binds the core component of MARE, which comprises of a TRE (phorbol 12-Otetradecanoate-13-acetate (TPA)-responsive element) or CRE (cyclic AMP-responsive) element. TGC sequences flanking this core are bound by c-MAF's extended homology region[31].

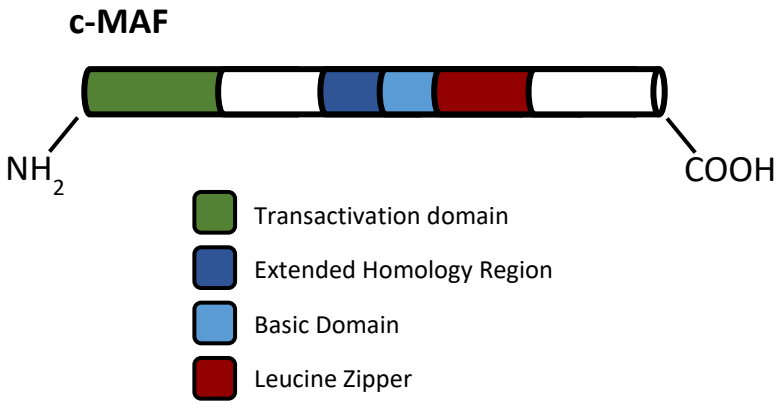


Figure 1.5: c-MAF structure

Schematic depicting the structure of the c-MAF protein. In red the leucine zipper domain, which allows the protein to homo- or heterodimerise. In light blue and dark blue the basic domain and extended homology region which bind DNA and in green, the transactivation domain, which permits transcriptional activation.

The MAF recognition element is thought to be bound by c-MAF homodimers or heterodimers[28, 31], however it has been shown that c-MAF can bind half the MARE sequence as a single protein[32]. Some of these interacting proteins, mostly within the AP-1 family of transcription factors, have previously been described[28], though have only been validated *in vitro*, for example by using electrophoretic mobility shift assays of *in vitro* synthesised proteins. Synthesised v-MAF was found to heterodimerise with synthetic forms of v-Jun and v-FOS[33]. c-MAF has been shown to interact with c-MYB in CD4 T-cells, downregulating Bcl-2 and promoting apoptosis[34]. Similarly, c-MAF has been shown to bind c-MYB early during the myeloid cell development regulating transcription[35]. Finally, c-MAF was shown *in vitro* to bind ATF4, BACH1, NFE2L1, MAFB and c-MAF, when tested against other bZIP transcription factors using protein microarrays[36]. The strongest interactions were seen in the context of c-MAF-MAFB heterodimer and a c-MAF homodimer. However, most of these interactions have not been validated *in vivo* or in the context of multiple myeloma.

The function of c-MAF is regulated by post-translational modifications promoting the protein's degradation. c-MAF is phosphorylated by glycogen synthase kinase 3β (GSK3β) leading to the transcription factor's degradation, and inhibition of the kinase has been shown to increase c-MAF in MM cell lines[37]. Ubiquitin-conjugating enzyme UBE20 mediates c-MAF ubiquitination and consequent degradation, with co-transfection assays demonstrating dose-dependent reduction in c-MAF protein[38]. In addition, deubiquitinase USP5 (ubiquitin specific peptidase 5) reduces ubiquitination of c-MAF leading to the protein's stabilisation[39]. Both phosphorylation- and ubiquitination-induced degradation are thought to be mediated via the proteasome, as c-MAF degradation is abolished upon exposure to a proteasomal inhibitor leading to stabilisation of the protein[37, 38].

1.2.2 MAF: physiological role

c-MAF (hereafter referred to as MAF) has a role in cellular development and differentiation across a variety of tissues, including in the development of the lens, chondrocytes[28], liver, kidneys, pancreatic islet cells[30, 40] and keratinocytes.

In the lens, MAF enables the differentiation of the lens fibre cells by controlling the expression of crystallin genes[40]. MAF knockout mice develop microphthalmia with loss of expression of these genes[31]. A missense mutation in the extended homology region of MAF was described in a family affected by congenital cataracts, with the mutation failing to activate the expression of crystallin genes[41]. Other mutations in the basic domain and extended homology region are also associated with cataracts and microcornea, whereas mutations in the N-terminus lead to syndromes of deafness, cataracts, seizures and learning difficulties[41].

MAF and another large MAF transcription factor, MAFB, are highly expressed during epidermal differentiation. In a set of experiments conducted on keratinocytes, MAF/MAFB knockdown reduced the expression of epidermal differentiation genes and resulted in increased proliferation of the cells[42]. By contrast, ectopic expression of the transcription factors in immature epidermal cells resulted in their forced differentiation, demonstrating their requirement for the development of epidermal cells[42].

In the nervous system, MAF is expressed early in the differentiation of various sensory neurons, such as mechanoreceptors, sensing touch, proprioceptors, detecting position, and nociceptors, the pain fibres[43]. MAF mutant mice showed impaired morphology of these neurons with thinner axons, and impaired function with reduced conduction velocity and abnormal firing, whereas some of the genes with reduced expression following MAF mutations encoded for potassium channels[44].

1.2.3 MAF in the haematopoietic system

MAF is not expressed in normal bone marrow[45] or tonsillar plasma cells[46]. Indeed, very little is known about MAF expression, if any, and its function in the B cell lineage.

By contrast, MAF plays a significant role in the development and cytokine production in various CD4+ T cell lineages. MAF is required for the differentiation of T-helper cells (Th cells) into the Th2 subset, were it is selectively upregulated upon stimulation of the T cell receptor in cells destined to develop into Th2 instead of Th1 cells[47]. MAF upregulation is necessary for IL-4 expression in Th2 cells, where it induces *IL-4* expression by direct binding to its promoter[47, 48]. In addition, MAF is expressed in T follicular helper cells (Tfh) where it regulates cytokine production and Tfh cell differentiation from naïve CD4+ cells. It is upregulated early on in the differentiation process and appears to be critical for Tfh cell development; T-cell compartment MAF knockout mice fail to acquire Tfh cell markers and have reduced numbers of Tfh cells upon antigen stimulation[49]. MAF knockout T cells also had significantly reduced IL-21 secretion, a cytokine required for the differentiation of CD4+ cells into Tfh cells[49, 50]. The mice were also unable to produce high affinity IgG antibodies upon antigen stimulation, implicating MAF further in the development and function of Tfh cells[49]. MAF also stimulates the secretion of IL-4 by Tfh by directly binding to the *IL-4* locus[51], as happens in the Th2 cells, as well as the expression of CXCR5[50] which mediates migration to the germinal centre. Finally, MAF was also found to be critical for certain Treg (T regulatory cells) subsets development, following induction by IL-6[52].

MAF has also been found to be essential in mouse fetal erythropoiesis, where its absence leads to defective interaction between otherwise normally-developing erythroblasts and macrophages, leading to embryonically lethal anaemia[53]. The failure in erythroblastic islands was due to downregulation of adhesion molecules such as VCAM-1 (vascular cell adhesion protein 1) in the MAF deficient macrophages[53].

Overall, in normal haematopoietic cell development and function, MAF regulates a transcriptional programme which mainly has a role in cell-cell interaction, whether this is by

direct interaction or signalling molecule-mediated, and regulates cellular communication with the microenvironment in a context dependent manner.

1.3 MAF as an oncogene

Contrary to its role in cellular differentiation, MAF can also function as a potent oncogene. The first MAF family member to be described, v-maf, was discovered as a potent retroviral transforming oncogene in a fibromyosarcoma in chickens[27]. Since then, MAF, and indeed other large MAFs such as MAFA and MAFB, have been found to play a major role in various cancers.

Pouponnot et al. demonstrated the transforming capacity of MAF (and MAFA) in primary cultures of chicken fibroblasts[54]. MAF-transfected cells displayed the ability to form colonies, increase their proliferation rate and had reduced rates of apoptosis[54].By contrast, this did not happen in neuroretinal cells, causing the authors to conclude that the transforming ability of MAFs depends on the cellular context.

In human cancers, MAF has been associated with bone metastasis in breast cancer: patients whose tumours had chromosome 16q23 gain had a higher incidence of bone metastases and a poorer overall survival, but was not associated with higher rates of other solid organ metastases[55]. Tumours from these patients expressed high levels of MAF. Mice with breast cancer tumours overexpressing MAF, had a similar tumour size to control mice, but a higher rate of bone metastases. This effect was MAF dependent, as mice injected with a MAF knockdown breast cancer cell line displayed reduced rates of bone metastases compared to controls[55].

MAF is expressed highly in hairy cell leukaemia and approximately half of T and NK-T cell lymphomas, in the absence of a MAF translocation, as assessed by immunohistochemistry [46]. It is also detected in 60% of angioimmunoblastic T cell lymphomas[56]. Indeed, transgenic mice

expressing MAF in the T cell compartment developed T cell lymphoma[56]. Transgenic mice were created by using a vector expressing murine MAF under the control of the *CD2* promoter. After 50 weeks of age the mice developed tumours, with enlarged lymph nodes, splenomegaly and hepatomegaly, with organs infiltrated by abnormal CD3+ T cells, suggestive of lymphoma[56]. It was noted that mice with lower transgene copies developed fewer tumours, prompting the authors to hypothesise that the level of expression of MAF regulates its oncogenic capacity.

MAF has also been implicated directly in myelomagenesis using a transgenic mouse model expressing MAF in the B cell compartment, created using constructs expressing MAF cDNA under the control of the *IgH* promoter and enhancers[57]. Splenic B cells from the mice overexpressed MAF, in contrast to wild type B cells, but otherwise showed no difference with respect to numbers or cell surface markers. However, the animals proceeded to develop tumours, with splenic enlargement, infiltration of CD138+ plasmacytoid cells and evidence of a paraprotein. In addition, the mice had evidence of renal cast nephropathy, which classically afflicts myeloma patients. The disease was characteristic of a B cell lymphoma with plasmacytoid features; however, these sets of experiments demonstrate the direct role of MAF in myelomagenesis.

1.4 MAF in multiple myeloma

1.4.1 Incidence and mechanisms of MAF overexpression in myeloma

In the context of MM, MAF is primarily overexpressed via the translocation t(14;16), which juxtaposes *MAF* to the *IgH* enhancer. The MAF breakpoint on 16q23 occurs centromeric to the *MAF* gene, within the intron of the oxidoreductase gene *WWOX*, resulting in its mono-allelic inactivation[14]. On the *IgH* locus, breakpoints occur within the JH or more commonly in the switch regions, juxtaposing MAF to the *IgH* enhancers[14, 58], sometimes up to 1Mb away[20]. (**Fig. 1.6**). The translocation can be found in MGUS patients and occurs at an incidence of 3-5% of MM cases [4, 9, 17-20, 23]. In addition there is a higher incidence of t(14;16) in plasma cell leukaemia patients[59], with the translocation detected in 13% of cases[60].

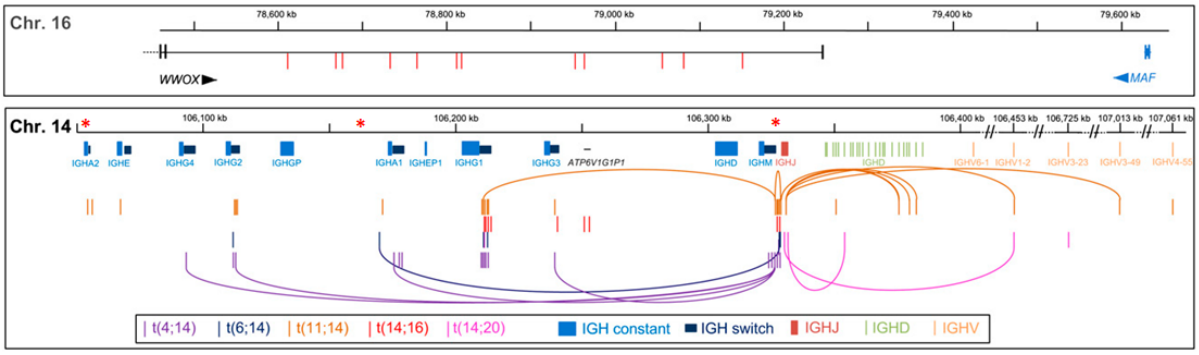


Figure 1.6: MAF t(14;16) translocation

Illustration of breakpoints involved in the MAF t(14;16) myeloma translocation. *MAF* breakpoints occur on chr.16 and respective *IgH* breakpoints are indicated with red lines on chromosome 14. Arcs indicate breakpoints occurring in the sample. Stars indicate approximate locations of *IgH* enhancers. Adapted from Walker, B.A., et al., Blood, 2013. 121(17): p. 3413-9[58].

MAF is also overexpressed in another 30% of myeloma cases [18, 45], and has been detected by immunohistochemistry in 30.5%, [61], a rate significantly higher than that of the translocation alone. This was found to be associated with the presence of the MMSET translocation, t(4;14). Primary samples and t(4;14) myeloma cell lines typically overexpressed MAF, though to a lesser degree than the MAF-translocated samples, and had upregulated c-

MAF gene targets[62] (see section 1.4.3). MAF overexpression in the context of the MMSET translocation was found to be due to activation of the MEK-ERK (mitogen activated protein kinase (MEK) – extracellular signal regulated kinase (ERK)) pathway, downstream of MMSET, which in turn upregulated FOS, an AP-1 transcription factor, to increase MAF expression by direct binding to the *MAF* promoter[62]. MAF, in the absence of a t(14;16) translocation, has also been found to correlate with Transmembrane activator and CAML interactor (TACI) expression, one of the receptors for B cell activating factor, BAFF, and a proliferation inducing ligand, APRIL[63]. Stimulation by BAFF and APRIL appeared to increase MAF expression, and the converse happened upon their inhibition. However, the exact pathways behind MAF upregulation in this context remain to be elucidated.

1.4.2 Clinical features and prognosis of MAF t(14;16) in myeloma patients

The presence of MAF translocation has been considered one of the high risk cytogenetic features in MM, with a median overall survival of approximately 3 years[9]. The t(14;16) forms part of the high risk cytogenetics considered in the R-ISS scoring system, along with del(17p) and t(4;14). Patients in the high risk group, R-ISS stage III, have a median overall survival of 24.5months[12]. Early studies using FISH, identified the presence of t(14;16) as a poor risk feature, with reduced progression-free and overall survival[64]. However, this was not the case in MGUS[64] or smouldering myeloma[9], where MAF was not associated with increased risk of progression, suggesting that perhaps the translocation is a clonal-initiating event, which however is not sufficient for full malignant transformation and progression of disease. In a combined analysis of the genetic features in 1905 newly diagnosed patients participating in the Myeloma IX and XI trials, MAF translocation was consistently associated with poor progression-free and overall survival[65]. This finding persisted in the subgroup analysis of young, fit patients participating in the trials. Gene expression profiling studies in purified myeloma plasma cells from patients participating in the HOVON-65/GMMG-HD4 trial used to identify a cluster of patients with less than 2 year overall survival, found the MAF translocation to be over-represented in the high risk group[66]. Finally, gene expression profiling studies conducted on large cohorts of patients have clustered patients based on their

molecular signature, identifying a subgroup of patients with a MAF gene expression signature, the "MF group"[67, 68]. In a study of intensively treated patients, the MF group was distinctly associated with poor event-free and overall survival[69] (**Fig. 1.7**). In addition, the t(14;16) has been associated with gain of 1q and del17p [18, 64, 70], both of which are independent predictors of poor survival.

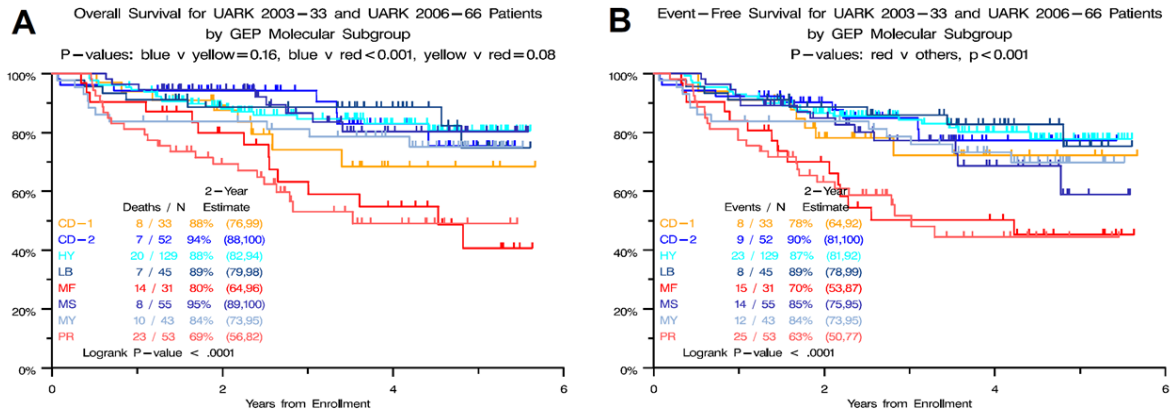


Figure 1.7: Survival per molecular myeloma subgroup

Overall survival **(A)** and event-free survival **(B)** in patients treated in the Total Therapy 3 trials 2003-33 and 2006-66 per gene expression profiling molecular subgroup (Figure from Nair, B., et al, Blood, 2010. 115(21): p. 4168-73[69]). Abbreviations refer to the molecular subgroups: CD-1, CD2; CCND1 and CCND2, HY; Hyperdiploid, LB; low bone, MF; MAF, MS, MMSET, MY; myeloid, PR; proliferative.

Casting doubt to the poor risk associated with MAF, the IFM 99-02 and 99-04 studies who treated patients with tandem autografts following chemotherapy found no significant difference in the overall survival of the small subset of t(14;16) patients[59]. This may suggest that a tandem autograft abrogates the poor prognosis conferred by the presence of the translocation. Similarly, the ISS-MUT score, derived from combining the ISS with whole exome sequencing from the NCRI Myeloma XI trial, did not include t(14;16) as a poor risk feature[71].

Discrepancies in these findings could be explained by the different methods used to define the MAF subgroup (gene expression profiling vs FISH), the different treatments received in each trial and the overall small number of patients affected by this translocation. In addition, MAF expression detected by immunohistochemistry was not associated with a reduced

overall survival, suggesting that MAF protein expression in the absence of a translocation, which leads to a much lower level of MAF expression, may not have a clinical course similar to t(14;16)[61].

t(14;16) is associated with an increased mutational burden which may be partly driving the high risk association. A study integrating RNA-seq, whole genome sequencing and clinical data from 450 newly diagnosed patients participating in the Multiple Myeloma Research Foundation CoMMpass trial, defined 37 modules of gene co-expression and their associated genetic events[72]. In this network model of myeloma, the MAF cluster was associated with an excess of mutations compared to other clusters. The mutational profile observed is that of an APOBEC-induced (apolipoprotein B mRNA editing enzyme catalytic subunit) mutational signature, an enzyme which modifies single stranded DNA[22]. APOBEC signature mutations are significantly enriched in t(14;16) and t(14;20), MAFB, myeloma compared to other subgroups and account for just under 29% of mutations in the former group. In addition, APOBEC3A and APOBEC3B enzymes are significantly overexpressed in t(14;16) samples and MAF knockdown resulted in a decrease in APOBEC3B expression[73]. Interestingly, *MAF* itself is mutated in 12.5% of t(14;16) cases, with missense mutations in the leucine zipper domain predicted to be activatory[73] whilst no mutations in *MAF* are detected in other myeloma subgroups[24]. However, these mutations were only detected in 50% of the t(14;16) clones, suggesting they are a secondary event, occurring after the translocation event[73]. Finally, t(14;16) has also been associated with an excess of mutations in *BRAF*, *TRAF2*, *ATM*, and *DIS3*[24].

Clinically, t(14;16) has been associated with higher β2-microglobulin levels[74], elevated lactate dehydrogenase (LDH)[68], thrombocytopenia[68], lower incidence of bone disease [75] and higher rates of renal failure[76]. In addition, MAF protein overexpression in patients, as detected by immunohistochemistry, is associated with hypoalbuminaemia and high β2-microglobulin levels [61].

The MAF translocation has been associated with resistance to bortezomib, with t(14;16) patients maintaining their poor prognosis despite addition of bortezomib to treatment regimens[69]. Resistance to the proteasome inhibitor by cell lines carrying t(14;16) is thought

to be mediated through reduction of proteasome-mediated degradation of MAF, which is dependent MAF phosphorylation by glycogen synthase kinase 3b (GSK3b) in the proteasome[37]; consistent with this, treatment with both bortezomib or carfilzomib increased MAF expression. Indeed, MAF overexpression in a t(14;16) negative background resulted in increased resistance to bortezomib, whereas conversely, reducing MAF expression increased cell death in response to treatment[37]. By contrast, MAF translocated cells, appear sensitive to the effects of dexamethasone, a glucocorticoid used in myeloma treatment, possibly associated with increased levels of NR3C1 (Nuclear Receptor Subfamily 3 Group C Member 1), a glucocorticoid receptor, highly expressed in MAF-translocated myeloma plasma cells[77]. Indeed, exposure to glucocorticoids reduces the expression of c-MAF protein in myeloma cell lines by increasing its ubiquitin mediated degradation[78]. However, t(14;16) cells could acquire resistance to dexamethasone *in vitro* by prolonged exposure to the drug, by reducing NR3C1 levels[77].

1.4.3 Molecular profiling of MAF

Studies conducted on cell lines have investigated to a certain degree the downstream molecular targets of MAF in multiple myeloma.

The first study conducted on myeloma cell lines by Hurt et al. identified cyclin D2 (*CCND2*), integrin β7 (*ITGB7*) and C-C motif chemokine receptor 1 (*CCR1*) as MAF targets[45]. The researchers first observed that these three genes along with MAF are differentially overexpressed in MAF-expressing cell lines as opposed to MAF-negative cells, and confirmed these findings in primary myeloma samples. Ectopic expression of MAF in MAF-negative cell lines resulted in the increased expression of the three putative target genes, and conversely expression of a dominant negative form of MAF (expressing MAF with an acidic DNA binding domain instead a basic DNA binding region) in MAF-positive cell lines, reduced their expression. Focusing on CCND2, a cell cycle protein, the researchers identified the MAF DNA binding motif, MARE, in the *CCND2* promoter but did not demonstrate direct binding of MAF to it by chromatin immunoprecipitation. MAF overexpression resulted in increased DNA

replication and cell proliferation and conversely expression of the dominant negative MAF resulted in reduced cell proliferation *in vitro* and delay of tumour formation *in vivo*. ITGB7, an adhesion molecule, heterodimerises to bind to E-cadherin, expressed on bone marrow stromal cells. Ectopic MAF expression increased the ability of myeloma cells to adhere to E-cadherin and to bone marrow stromal cells. This effect was abrogated when incubating the same cells with antibodies against ITGB7 or E-cadherin. In addition, when allowed to bind to bone marrow stromal cells, MAF-transduced cells produced VEGF at much higher levels than MAF-negative cells. Taking these data together, the authors concluded that MAF promotes myeloma cell survival by increasing their proliferation and adhesion to bone marrow stroma.

Along the same lines, subsequent studies have found that increased ITGB7 was associated with MAF translocation and overexpression[79]. ITGB7 knockdown reduced myeloma cell adhesion to bone marrow stromal cells, E-cadherin and fibronectin, an effect which was recapitulated when using an anti-ITGB7 antibody. This effect could be rescued upon expressing ectopic ITGB7. In addition, ITGB7 knockdown led to reduced cell transwell migration as well as the ability of myeloma cells to migrate through Matrigel, which was used as a surrogate for the endothelial basement membrane of the bone marrow sinuses. Although myeloma tumour growth was not reduced in transgenic mice injected with ITGB7-knockdown cells, the number of cells observed in the animals' bone marrow was reduced, suggesting the capacity of the cells to home in to the bone marrow was affected by the reduction in ITGB7. Furthermore, in support of the previous findings, the investigators observed that ITGB7-knockdown myeloma cells produced less VEGF, IL-1β and C-C Motif Chemokine Ligand 4 (CCL4) when co-cultured with bone marrow stromal cells[79]. These cells also had reduced survival upon exposure to bortezomib as compared to control cells with intact ITGB7 expression, suggesting another potential mechanism for the proteasome resistance associated with MAF.

As described above, CCR1, a chemokine receptor, was also found to be a putative target of MAF, which increased upon MAF overexpression and reduced upon MAF knockdown[45]. In the same study researchers found that MAF ectopic expression in myeloma cells increased chemotaxis towards CCL3, the ligand for CCR1. In addition to chemotaxis, CCR1 enhances cell proliferation and adhesion, by upregulating ITGB1 expression and ITGB1-mediated adhesion

to stromal cells upon CCL3 stimulation, a process which can be inhibited with CCR1 or CCL3 blockade[80]. This binding induces IL-6 secretion by bone marrow stromal cells[80], which in turn stimulates myeloma cell proliferation. CCR1 expression correlates with plasma cell infiltration in the bone marrow and number of circulating plasma cells in the peripheral blood[81], further supporting the role of CCR1 in both myeloma cell adhesion and migration. CCR1 is also involved in mediating osteoclast proliferation and differentiation and has been implicated in lytic bone disease[82, 83], although paradoxically, MAF has been associated with reduced bone disease. A recent abstract published in the American Society of Haematology Conference (ASH) also implicated CCR1 in myeloma cell migration *in vitro* and dissemination of the malignant cells *in vivo*, implicating overexpression of the molecule in myeloma dissemination[84], although these findings have not yet been peer-reviewed.

Finally, another key target of MAF in MM has been reported to be NUAK1/ ARK5[85], a serine/threonine-protein kinase which has been shown to interact with p53[86] and has been associated with tumour metastases and cellular invasion[87]. NUAK1 was upregulated upon MAF ectopic expression, and this was abrogated by mutations induced in the MARE DNA binding MAF motif in the *NUAK1* promoter[85]. In the same study, NUAK1 was shown to promote cellular invasion of Matrigel in the presence of Insulin-like growth factor 1(IGF-1), leading to the hypothesis that NUAK1 promotes myeloma cell invasion in the context of t(14;16) MM.

Gene expression profiling studies conducted on smaller or larger number of myeloma patients have clustered MM into molecular subgroups and identified their distinct expression profiles. Zhan et al., using samples from 414 newly diagnosed patients, clustered MM in 7 subgroups based on the molecular profile, and noted that the groups clustered based on the overexpression of specific cluster-defining genes, including MAF and MAFB which defined the MF subgroup[67]. Each of the clusters was distinguished by a set of uniquely over or under-expressed genes; the top 50 distinguishing genes was used to define each molecular subgroup. In this study again, the MF molecular subgroup was identified as one of the poor risk groups. Using the same methodology, Broyl et al., performed gene expression profiling on samples from 320 newly diagnosed patients and clustered myeloma into 10 distinct molecular subgroups, with the MF subgroup being one of them, and identified the top 50

upregulated and top 50 downregulated defining genes per subgroup[68]. As was the case in the previous study, the MF subgroup was characterised by the presence of a *MAF* gene translocation, however, the translocation was not detected by FISH in all patients clustering in this subgroup. It is worth noting that the top 50 gene discriminators for the MF subgroup differed to a certain extent between studies: for example, *ITGB7* and *NUAK1* were found to be significant MF subgroup distinguishing genes in both studies, whereas *CCND2* was only found to be a discriminator for the MAF subgroup in the latter study. This could be due to the differences in the patient populations (for example in Zhan et al., a quarter to a fifth of the patients were over 65 years old, whereas no patients over 65 years old were included in Broyl et al.), different numbers of patients per cluster and relatively small number of patients per cluster. Finally, Mattioli et al., analysed DNA microarray data from 52 patients with MGUS, myeloma and plasma cell leukaemia and found that gene expression clustering was mainly determined by presence or absence and the type of IgH translocation[88]. Comparing the expression panels between groups the MF group was identified on the basis of 18 upregulated and 1 downregulated gene, mostly involving genes regulating chemotaxis and cell adhesion.

By combining the above expression profiling data associated with the MAF molecular group[67, 88], Eychene et al., have distilled a core programme of MAF regulated targets consisting of 12 genes: *CCND2*, *NUAK1*, *ITGB7*, chemokine (C-X-C) ligand 12 (*CXCL12)*, chemokine (C-X3-C) receptor 1 (*CX3CR1)*, neurotrophic tyrosine kinase receptor type 2 (*NTRK2)*, Toll-like receptor 4 (*TLR4)*, Dihydropyrimidinase-like 2 (*DPYSL2)*, Actin binding LIM protein 1 (*ABLIM1)*, ATPase type 13A2, *ATP13A2*, Glycogen phosphorylase brain form (*PYGB)*, and LOC400451 with no known encoded protein[28]. Based on these, they concluded that MAF affects the interface between myeloma and bone marrow microenvironment, promoting cell proliferation, adhesion and migration. These cellular processes, however, are not part of the differentiation profile MAF regulates in normal tissues, leading the authors to hypothesise that other factors such as different, PC-specific proteins partners or microenvironment stimulation promote the shift to a malignant MAF-transcriptional programme[28].

1.5 Aims

The overall aim of the project was to determine the impact of oncogenic MAF on the regulatory genome of myeloma PC.

Specifically, to:

1. Determine the genes and genomic regulatory areas that are directly regulated by MAF by defining and comparing the
    a. Genome-wide MAF binding profile
    b. Chromatin accessibility
    c. Transcriptome

    in MAF-translocated myeloma cell lines, primary myeloma cells and normal plasma cells.

2. Determine the role of MAF in establishing enhancers promoting the MAF oncogenic transcriptional programme.

3. Identify the transcription factors and other chromatin associated proteins cooperating with MAF as part of the MAF oncogenic interactome.

Chapter 2

Materials and Methods

A list of reagents and kits used is supplied in **Appendix A**.

2.1 Human Samples

2.1.1 Sample Collection

Bone marrow samples were collected after informed consent, following local ethics committee approval (Research Ethics Committee reference 11/H0308/9). Multiple myeloma patients were attending the haematology services either at the Hammersmith Hospital, Imperial College Healthcare NHS Trust in London, or at the AHEPA University Hospital of Thessaloniki in Greece. Normal bone marrow plasma cells were obtained from healthy adults donating bone marrow for the purpose of haematopoietic stem cell transplantation. Patient details are outlined in **Appendix B Table S1**. Bone marrow aspirate samples were placed in Ethylenediamine Tetraacetic Acid (EDTA)-containing tubes to prevent coagulation. All samples were anonymised and given an identifying code upon receipt.

2.1.2 Immunomagnetic CD138-positive cell selection

Myeloma bone marrow aspirates were processed within 24 hours of collection and downstream applications were performed the same day. One-7mls of bone marrow aspirate was passed through a 40μm cell strainer and subjected to red cell lysis, (BD Pharm Lyse, BD Biosciences) incubating for 30 minutes at room temperature (RT). Cells were then washed twice with Phosphate Buffered Saline (PBS; Sigma-Aldrich) and counted, using the trypan blue exclusion test method and a Neubaeur haemocytometer. A small aliquot of the sample was taken for flow cytometry, to assess pre-selection bone marrow disease infiltration (described in section 2.3.1). Plasma cells were purified by performing CD138 positive immunomagnetic bead selection using anti-CD138 MicroBeads (Miltenyi Biotec) and MACS columns (Miltenyi Biotec) as per manufacturer's instructions. Briefly, cells were resuspended in ice cold RoboSep buffer (Stem Cell Technologies) and FcR blocking reagent (Miltenyi Biotec) was

added at the appropriate amount, incubating for 5 minutes. Subsequently, the sample was incubated with the calculated amount of CD138 selection beads for 15 minutes, at 4°C. Cells were then washed and resuspended in cold RoboSep buffer and passed twice through the selection column placed in a magnetic field. LS columns were used without taking into account the number of cells. The purity of post-selection plasma cells was assessed using flow cytometry (section 2.3.1) and the sample was immediately processed for ATAC-seq and RNA-seq as described below (sections 2.6, 2.7, 2.8). Remaining sample was preserved by resuspending in cryopreservation medium containing Fetal Bovine Serum (FBS; Thermofisher Scientific) and 10% dimethyl sulfoxide (DMSO; Sigma-Aldrich) Samples were stored at -80°C for 48 hours and subsequently transferred to liquid nitrogen.

2.1.3 Healthy donor bone marrow mononuclear cell isolation

Healthy donor bone marrow aspirates were processed on the day of collection. Up to 30mls of aspirate were passed through a 40µm cell strainer and diluted at a ratio of 1:3 with PBS. The diluted sample was then layered over histopaque (Sigma-Aldrich) at a volume of 2:1, sample to histopaque. This was then centrifuged for 40 minutes at 400g RT, with acceleration set at 5 and deceleration at 0. The bone marrow mononuclear cell layer was then aspirated and washed twice with PBS with each centrifugation set at 300g for 10 minutes, after which cells were counted. Samples were subsequently depleted of T cells and monocytes, by CD3 and CD14 negative selection as described below.

2.1.4 T cell and monocyte depletion of healthy donor bone marrow mononuclear cells

Healthy donor bone marrow mononuclear cells were subjected to sequential T cell and monocyte depletion, using the EasySep Human CD3 and CD14 positive selection kits (Stem Cell Technologies) respectively, following manufacturer's instructions. For CD3+ cell depletion, cells were resuspended in RoboSep buffer at a final concentration of 1 x 10^8cells/ml and 100µl/ml of the Selection Cocktail was added to the sample, incubating for 3 minutes RT.

Then 60µl/ml of RapidSpheres were added, and the mixture was incubated for 3 minutes at RT. Subsequently, RoboSep buffer was added to 2.5mls and the sample was placed on an EasySep^{TM} Magnet. The flow through, containing the cells of interest, was poured off into a falcon tube, whilst CD3 labelled cells were held in the magnet. This process was repeated twice. Cells contained in the flow throughs were counted and proceeded to CD14+ cell depletion (using the EasySep CD14 selection kit), for which the same process was followed, as per manufacturer's instructions. Following this, samples were immediately processed for flow cytometry and fluorescence-activated cell sorting (FACS sorting) of plasma cells.

2.2 Cell lines

Human multiple myeloma cell lines MM.1S, NCI-H929 (ATCC, Manassas, VA, USA), JJN3 and U266 (DSMZ, Germany) and the Human Embryonic Kidney 293 cells (293T; ATCC, Manassas, VA, USA) were cultured at 37°C, in a 5% CO_2, humidified atmosphere. MM.1S, NCI-H929 and U266 were maintained in Roswell Park Memorial Institute 1640 medium (RPMI1640; Sigma-Aldrich), supplemented with 20% FBS, 1% penicillin/streptomycin (Sigma-Aldrich), 2mM L-Glutamine (Sigma-Aldrich), 0.1mM non-essential amino acids (Sigma-Aldrich) and 1mM sodium pyruvate (Sigma-Aldrich). JJN3 were grown in 40% Dulbecco's Modified Eagle Medium (DMEM; Sigma-Aldrich), 40% Iscove's Modified Dulbecco's Medium (IMDM; Sigma-Aldrich), 20% FBS and 1000U/ml penicillin/streptomycin. 293T cells were grown in DMEM supplemented with 10% FBS and 1% penicillin/streptomycin. Cell lines were maintained at a concentration of 0.4-1x10^6/ml as required. Mycoplasma testing was performed monthly by PCR.

2.3 Flow cytometry

2.3.1 Surface marker staining for primary myeloma samples

Flow cytometry was performed on primary myeloma samples pre- and post-CD138+ cell selection (section 2.1.2), in order to assess pre-selection marrow infiltration and post-selection sample purity. Cells were washed twice with PBS and a third wash was performed with 1ml 0.1% PBS / Bovine Serum Albumin solution (PBS/BSA). Cells were resuspended in 100µl 0.1% PBS/BSA and incubated with 1µl FcR blocking agent (Miltenyi Biotec) for 5 minutes at RT. Cells were then stained with fluorophore-conjugated antibodies for CD138, CD38, CD45, CD56 and CD19 (**Table 2.1**) for 20minutes at 4°C, protected from light. Samples were washed with PBS, resuspended in 300µl of PBS and analysed on a BD LSR Fortessa. Subsequent analysis was performed using the FlowjoX platform. Only samples attaining >80% plasma cell purity were used for downstream experiments (see **Appendix B Table S1** for details).

2.3.2 Surface marker staining for FACS sorting of primary healthy donor plasma cells

Healthy donor bone marrow samples were processed as described in sections 2.1.3 and 2.1.4 and subsequently stained with fluorescent markers to allow FACS-sorting of normal plasma cells. Cells were resuspended in 200-300µl of PBS and 10µl of FcR blocking reagent was added allowing for a 5-minute incubation. Cells were then stained using a "healthy donor" panel of fluorophore-conjugated antibodies (**Table 2.1**), by incubating for 20 minutes at 4°C, shielded from light. Cells were finally washed and resuspended in 3mls of RoboSep buffer. A small aliquot was used to assess the quality of the sample on a BD LSR Fortessa. Plasma cells were purified from the rest of the sample using FACS-sorting on a BD FACSAria II. Cells were then immediately processed for ATAC-seq and RNA-seq (described in sections 2.6, 2.7 and 2.8).

Table 2.1: Monoclonal antibodies used in flow cytometry panels

Surface marker	Fluorophore	Manufacturer	Clone
Healthy Donor Normal Plasma Cell Panel			
CD138	FITC	BD Biosciences	MI15
CD319	PE	eBioscience	162
CD27	PECy7	Biolegend	0323
CD3, CD14, CD16	Alexa647	Biolegend	HIT39, HCD14, 3G8
CD2	Alexa647	Bio-Rad	0310
GPA	Alexa647	Bio-Techne	R10
CD19	APC eFluor780	eBioscience	HIB19
CD45	Alexa700	Biolegend	2D1
CD38	eFluor450	eBioscience	HIT2
Myeloma Sample Plasma Cell Panel			
CD138	FITC	BD Biosciences	MI15
CD38	BV421	BD Biosciences	HIT2
CD45	Alexa700	Biolegend	2D1
CD19	APC efluoro780	eBioscience	HIB19
CD56	PECy7	eBioscience	CMSSB

2.3.3 Intracellular staining for MAF expression

To assess the level of MAF protein expression by flow cytometry, intracellular staining for MAF or isotype control was performed using the eBioscience Foxp3 / Transcription Factor Staining Buffer Set (Thermofisher Scientific), following manufacturer's instructions with some optimisations. Cells were washed and resuspended 100µl of PBS. 1ml of the Fixation/Permealisation working solution was added whilst vortexing and samples were incubated on ice and shielded from light, for 1 hour. Each sample was subsequently split into two (for anti-MAF or isotype control staining), washed twice with 2mls of cold 1x Permealisation Buffer, by centrifuging at 600g at 4°C, and resuspended in 100µl of the buffer. 0.1µg of either anti-MAF (eBioscience) or mouse isotype (eBioscience) eFluor660 conjugated monoclonal antibody was added and incubated at RT in the dark for 2 hours. Samples were washed as before, resuspended in 300µl of PBS and analysed on a BD LSR Fortessa. Results were analysed on the FlowjoX platform. Median Fluorescence Intensity Ratio was calculated as the ratio of MAF median fluorescence intensity over Isotype median fluorescence intensity.

2.3.4 Cell cycle and apoptosis assay

For cell cycle and apoptosis assays, the AnnexinV apoptosis detection kit (eBioscience) was used, in order to bind and stain phosphatidylserine a cell surface marker of apoptosis, in conjunction with Propidium iodide (PI; Sigma-Aldrich), a marker for necrosis, and a Hoechst 33342 dye (Abcam) to stain DNA. Cells were resuspended in fresh media and incubated with 10μM final concentration of Hoechst dye for one hour at 37°C. Following washing with AnnexinV buffer (eBioscience), cells were resuspended in 100μl of the buffer and 2μl of AnnexinV (eBioscience) were added. Samples were incubated for 15 minutes on ice, protected from light, after which 3μl of PI were added. Data acquisition was performed on a BD LSR Fortessa, and results were analysed using the FlowjoX software.

2.4 Cloning

2.4.1 Production of lentiviral MAF targeting shRNA vectors for knockdown experiments

Short hairpin RNAs (shRNA) against MAF were used to knockdown gene expression, as previously described[89]. The 3rd generation lentiviral pLKO.1 vector (Addgene Plasmid #8453) was used, modified previously in the lab to replace puromycin resistance with eGFP gene (pLKO.1-eGFP). shRNA 21 bp sequences (**Table 2.2**) were designed against the first exon of the gene, in order to target all three known transcripts, using the online Probe function of NCBI (https://www.ncbi.nlm.nih.gov/probe). Candidate shRNA sequences were then screened based on available literature. The absence of predicted off target effects was confirmed using the eRNAi online tool, provided by the German Cancer Research Centre. Sense and antisense sequences were designed with appropriate intervening bases to allow for looping of the shRNA construct and cloning into the pLKO.1 vector, as per company's protocol. A scrambled (scrbl), non-targeting sequence was used as a control.

The vector was first digested using Fast Digest EcoRI and AgeI restriction enzymes (Thermofisher Scientific), incubating at 37°C for 2 hours, and was dephosphorylated using FastAP Thermosensitive Alkaline Phosphatase (Thermofisher Scientific). The linearised vector was subsequently separated on a 0.8% agarose gel and gel extracted using the Agarose Gel Extraction kit (Jena Bioscience), as per manufacturer's instructions. DNA was quantified using the NanoDrop 2000 (Thermofisher Scientific). shRNA oligos were phosphorylated and then annealed by mixing 20µM of forward and reverse sequences with T4 Polynucleotide Kinase (New England Biolabs) in 1X T4 DNA ligase reaction buffer (New England Biolabs) and incubating for 30 minutes at 37°C, for phosphorylation, and then sequentially for 5 minutes at 95°C, 10 minutes at 70°C and gradual cooling to RT. Finally, the digested vector and annealed oligos were ligated at a 1:2 insert to vector molarity ratio, using T4 DNA ligase and buffer (New England Biolabs) and incubating at room temperature for 3-4 hours. The final ligated vector is shown in **Fig. 2.1**. A negative control ligation, with the digested vector alone, was also performed to assess extent of vector self-annealing or incomplete digestion. The ligation product was then transformed into NEB 10-beta competent E.coli cells (New England Biolabs). Colonies were screened by colony PCR using pLKO.1-eGFP specific primers (**Table 2.3**). Positive for the insert colonies were grown and plasmids were purified using the GeneJET Plasmid Miniprep Kit (Thermofisher scientific). Correct sequences were confirmed by Sanger sequencing.

Table 2.2: shRNA sequences (5'-3') used for MAF knockdown experiments

shRNA	Forward	Reverse
MAF sh1[73]	TGGAAGACTACTACTGGATGA	TCATCCAGTAGTAGTCTTCCA
MAF sh4	TGTTAATGACTTCGATCTGAT	ATCAGATCGAAGTCATTAACA
Scrbl control	GGCCCTCCATCACAGTTAT	ATAACTGTGATGGAGGGCC

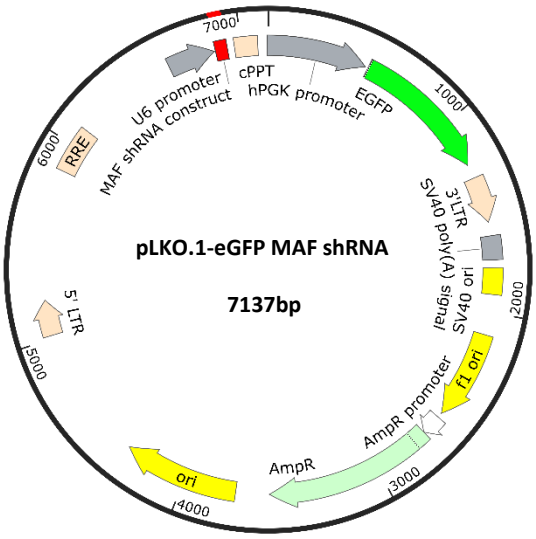


Figure 2.1: pLKO.1 vector used for MAF shRNA-mediated knockdown

Map of pLKO.1 eGFP vector containing MAF shRNA construct.

2.4.2 Production of retroviral MIGR1 MAF vector for overexpression experiments

For overexpression experiments, MIGR1 vector expressing eGFP was used (Addgene, #27490). MAF cDNA, designed using the gene sequence from the Ensembl Genome Database as a template, was ordered from GenScript, cloned into a pUC57 plasmid. The full sequence of the cDNA can be found **in Appendix B Table S2**. Forward and reverse primers used to amplify the cDNA (**Table 2.3**) were designed to contain the restriction sites for KspAI and EcoRI at the 5' and 3' end respectively.

First, the MIGR1 vector was digested using FD KspAI and EcoRI (Thermofisher scientific) restriction enzymes and the linearised vector was separated on a 0.8% agarose gel and extracted using the Agarose Gel Extraction kit (Jena Bioscience) as described above. MAF cDNA was amplified using the Phusion High-Fidelity PCR Master Mix with GC Buffer (New England Biolabs) as per manufacturer's instructions. The PCR product was then digested with the same restriction enzymes and purified from an agarose gel as above. Purified cDNA and linearised vector products were quantified using NanoDrop 2000 (Thermofisher Scientific) and subsequently ligated using T4 DNA ligase and buffer (New England Biolabs) as described above,

at a molarity ratio of 1:2 vector to insert. Ligation products were transformed into NEB 10-beta competent E.coli cells (New England Biolabs). Successful cloning was confirmed by colony PCR and Sanger sequencing using primers outlined in **Table 2.3**.

Table 2.3: Primers used to for pLKO.1-eGFP and MIGR1 cloning

Name	Sequence (5'-3')	PCR product size
pLKO.1 shRNA sequencing		
Forward 5' pLKO.1	TGGACTATCATATGCTTACCGTAAC	250bp
Reverse 3' pLKO.1	GTATGTCTGTTGCTATTATGTCTA	
MIGR1 MAF sequencing		
F_SEQ_MIG	CCTAAGCCTCCGCCTCCTCTTC	1354bp
R_SEQ_MIG	GAAGACAGGGCCAGGTTTCCGG	
MAF cDNA amplification		
F_MAF_MIG	TCGAGGTTAACATGGCATCAGAACTGGCAATG	1212bp
R_MAF_MIG	GGGCGGAATTCTCATTTTGTGAACACACTGGTAAG	
CRISPRi sgRNA sequencing		
CRISPRi KRAB F	GAGATCCAGTTTGGTTAATTAAGG	400 bp
CRISPRi KRAB R	TTCGCGGCCAAAGTGAATTC	

2.4.3 CRISPRi lentiviral cloning

In order to silence genomic regulatory regions, a CRISPRi method was employed, using an inducible lentiCRISPR-dCas9-KRABv2 vector kindly donated by Dr Feldhahn's lab, Imperial College London (**Fig. 2.2**). The lentiviral lentiCRISPRv2 vector (Addgene #52961) was modified to express inactive, "dead" CAS9 nuclease (dCAS9) fused to the Krüppel associated box (KRAB) domain and eGFP, upon induction with doxycycline. The level of induction was indicated by green fluorescence. The vector also contained a constitutively-expressed puromycin resistance gene as a selection marker. Two 20bp-long single guide RNAs (sgRNA) were used per target region. sgRNAs were designed using the CRISPOR online tool[90] (http://crispor.tefor.net/), using an "NGG" Protospacer Adjacent Motif (PAM), required by

the Streptococcus pyogenes Cas9 enzyme. Guides were selected based on a high MIT specificity score[91, 92], a predictive score of how "unique" the guide sequence is in the genome, and lowest number of off-target loci. The exact sequence and details pertaining to each guide are outlined in **Table 2.4.** When off-target binding was predicted, the highest number of mismatches with the off-target sequences was preferred (> 2 mismatches), with off-target sequences preferentially present in intergenic or intronic regions rather than exons. Guides were excluded if the sgRNA had no mismatches with the off-target regions in the 12bps adjacent to the PAM. Guides used as part of a larger collaborative project (CCND2 sgRNAs) were designed using the CRISPR MIT tool (http://crispr.mit.edu) following the same principles. Oligos, obtained from Sigma-Aldrich, were designed to have overhangs compatible with the plasmid following digestion with BsmBI restriction enzyme. A non-targeting guide against the *S cerevisiae* transcription factor Gal4, was used as a negative control[93].

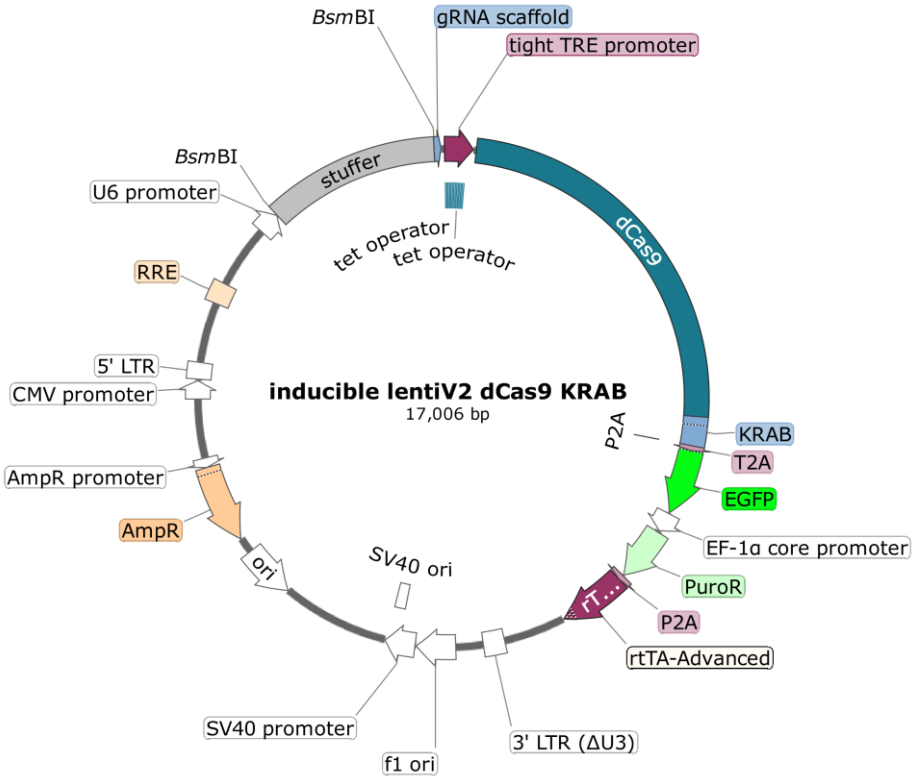


Figure 2.2: Map of inducible lentiV2 dCAS9 KRAB vector, used for CRISPRi experiments

Cloning was performed according to the Zhang Lab lentiCRISPRv2 and LentiGuide Oligo Cloning Protocol[94] with some modifications. The vector was first digested using the FastDigest Esp3I (BsmBI) restriction enzyme (Thermofisher Scientific), in the presence of Dithiothreitol (DTT) and FastAP Thermosensitive Alkaline Phosphatase (Thermofisher Scientific), incubating at 37°C for 2 hours. The linearised vector was separated on a 0.8% agarose gel and gel extracted using the Agarose Gel Extraction kit (Jena Bioscience), leaving behind the 2kb stuffer. Subsequently, 100µM of each of the forward and reverse oligos were phosphorylated and annealed as described in section 2.4.1. The annealed oligos were diluted at 1:200 with nuclease-free water and 1µl of the dilution was used to ligate with 50ng of the digested vector, using T4 DNA ligase and buffer (New England Biolabs) and incubating for 3-4 hours. Following transformation in NEB 10-beta competent E.coli cells (New England Biolabs), colony PCR and Sanger sequencing were used to confirm successful cloning, using the primers indicated in **Table 2.3**.

Table 2.4: sgRNAs used in CRISPRi experiments

Guide Name	Target	Sequence	PAM sequence	MIT specificity Score
CCR1 P.1	*CCR1* promoter	TGAGTACTTCCTACTTGGTC	AGG	84
CCR1 P.2		CAAGTTGAGTCAACATTGTC	AGG	84
CCR1 1.1	*CCR1* enhancer peak 1	GAGCACCTCTCAGTTTGATG	TGG	76
CCR1 1.2		GAGAAGACTCACTCCTCTGT	GGG	75
CCR1 2.1	*CCR1* enhancer peak 2	CGTTTGCAAGTTTCTAGTAC	TGG	91
CCR1 2.2		GTGATCTTCACAATAGACTG	AGG	78
CCR1 3.1	*CCR1* enhancer peak 3	TGGTTACACTCACGCAACTA	TGG	93
CCR1 3.2		ACACTCTATAATCGCAAAGT	AGG	92
TLR4 P.1	*TLR4* promoter	ATAGCGAGCCACGCATTCAC	AGG	96
TLR4 P.2		TCGAGCCGCCCCACAGCAAT	TGG	92
TLR4 1.1	*TLR4* enhancer	GTCTAGCACCACTAAAGGCA	GGG	86
TLR4 1.2		AATGGCAGGGGATGAATATC	TGG	76
gal4	nil	AACGACTAGTTAGGCGTGTA	N/A	N/A
CCND2 P.1	CCND2 promoter	TCTGAAGCGGTGACGCAAGC	TGG	91
CCND2 P.2		GCTAAATAGGGGGTTTTCGG	GGG	92
CCND2 1.1	CCND2 enhancer peak 1	TCTACCTGCCGCACCCACGG	AGG	78
CCND2 1.2		CCGTGAGGCCCCTGTGTATC	TGG	83
CCND2 2.1	CCND2 enhancer pear2	TAGCTGCACTCAGCGCTCGA	GGG	94
CCND2 2.2		GCTCAGGCACGACCGTCAGT	GGG	92
CCND2 3.1	CCND2 enhancer peak 3	CAAGATCTGCCACCGTCTGT	GGG	86
CCND2 3.2		AGCCTATTAGGGAGAACCCC	AGG	83
CCND2 4.1	CCND2 enhancer peak 4	GCACGTCTATTAAGCGTTGA	GGG	96
CCND2 4.2		GACCACGTCGAGAGCTCCCC	AGG	87

2.5 Virus production and transduction

2.5.1 Lenti- and Retro-viral production

Following successful cloning, plasmids were purified using the GeneJET Plasmid Maxiprep Kit (Thermofisher Scientific) and quantified using the Nanodrop 2000. Retroviral and lentiviral production was achieved using the calcium phosphate precipitation method[95] to transiently co-transfect the viral vector and helper plasmids into HEK293T cells. **Table 2.5** outlines the helper plasmids used for each recombinant virus and composition of the transfection mixture, containing Low TE (10ml 0.5M Tris-HCl pH 8, 100µl EDTA 0.5M), 2.5M $CaCl_2$ and 2x Alfa Aesar™ HEPES-buffered saline (2x HBS; Thermofisher Scientific). 2x HBS was added to the transfection mixture last, slow and drop-wise, whilst vortexing the mixture. 1ml of the mixture was added per 10cm^2 plate of HEK293T cells. 8-12 hours later the medium was changed and the cells were subjected to a glycerol shock using a 15% glycerol solution (Honeywell) with 2x HBS and sterile water. 48 and 72 hours post transfection the virus-containing supernatant was collected and concentrated by ultracentrifugation at 23,000rpm for 1 hour and 40 minutes at 4°C. Concentrated virus was gently resuspended in 200-300µl of FBS-free medium for a minimum of 2 hours at 4°C. Viral concentrates were then aliquoted into 50-300µl vials and snap frozen in liquid nitrogen, prior to storage at -80°C.

Table 2.5: Constituents of transfection mixture for lentiviral and retroviral production

Lentivirus		
COMPONENT		**QUANTITY**
Lentiviral plasmid	pLKO.1	10µg
	lentiCRISPR-dCas9-KRABv2	
Packaging plasmid for 3rd generation lentivirus	pRSV - REV	2.5µg
	pMDLg / pRRE	5µg
Packaging plasmid for 2nd generation lentivirus	psPAX2	7.5µg
Envelope plasmid	pMD2.G (VSV-G)	3µg
Low TE solution		To 450µL total volume
$CaCl_2$ 2.5M		50µL
2xHBS		500µL
TOTAL		1mL per 10cm^2 plate of HEK293T
Retrovirus		
COMPONENT		**QUANTITY**
Retroviral plasmid	MIGR1	12µg
Packaging plasmid	pCMV-Gag-Pol	12µg
Envelope plasmid	pVSV-G	3µg
Low TE solution		To 450µL total volume
$CaCl_2$ 2.5M		50µL
2xHBS		500µL
TOTAL		1mL per 10cm^2 plate of HEK293T

2.5.2 Viral transduction

100,000 target cells were plated in 500µl of media the day before the transduction. The following day, virus was transduced into the target cells using hexadimethrine bromide (polybrene; Sigma-Aldrich) at final concentration of 8µg/ml. Cells were subsequently spinoculated at 2000rpm at 32°C for 2 hours with acceleration set to 4 and deceleration to 1. Culture medium was replaced 16-20 hours later. For MAF knockdown experiments cells were

FACS sorted on a BD FACSAria II on day 2 and 3 post transduction and processed immediately for downstream experiments whereas for overexpression experiments, cells were FACS-sorted on day 5. For CRISPRi experiments, transduced cells were selected by culturing in the presence of puromycin (VWR life science) at a final concentration of 5μg/ml added 5 days post transduction. Cells alive at 48 hours post puromycin treatment, indicating successful integration of the virus, were FACS-sorted using an MA900 Multi-Application Cell Sorter (SONY). Cells were allowed to recover for 10-14 days, after which induction was performed using daily application of doxycycline hyclate (Sigma-Aldrich) at 1μg/ml final concentration. Doxycycline was reconstituted at 100μg/ml in PBS, fresh prior to each experiment, and stored at -20°C, protected from light. Four days post doxycycline induction, cells were FACS-sorted for downstream experiments.

2.6 Nucleic acid techniques

2.6.1 RNA extraction

Total RNA was extracted from sorted cells (or CD138 selected cells in the case of primary myeloma plasma cells). For RNA-seq, the Nucleospin RNA II kit (Macherey Nagel) was used as per manufacturer's instructions. Briefly, after washing with PBS, cells were lysed by resuspending in lysis buffer containing β-mercaptoethanol (Sigma-Aldrich), vortexing and storing in -80°C. Subsequently, lysate underwent column purification, including incubation with rDNAse. Following RNA quantification by NanoDrop 2000 (ThermoFisher Scientific), an aliquot was used for cDNA synthesis and qPCR, as described below, and the rest of the sample was stored in -80°C for subsequent library preparation (see section 2.7).

For samples for gene expression analysis by qPCR alone, the ReliaPrep RNA Cell Miniprep System kit (Promega) was used, as per manufacturer's instructions. Principles of the protocol followed were as described above, with the lysis buffer containing 2-propranolol (Sigma-Aldrich). RNA was quantified by NanoDrop 2000 and proceeded immediately to cDNA synthesis.

2.6.2 cDNA synthesis

The RevertAid RT kit (Thermofisher Scientific) was used for cDNA synthesis, following manufacturer's instructions. First, RNA was incubated at 65°C with random hexamer primers for 5 minutes, after which samples were brought at 4°C and kept on ice. Subsequently, dNTPs, reverse transcriptase and RiboLock RNAse inhibitor with the appropriate amount of buffer were added to the samples, which were incubated for 5 minutes at 25°C and then at 42°C for one hour. The reaction was terminated by bringing samples to 70°C for 10 minutes. Samples were subsequently diluted with nuclease free water by at least 3 times, and were used for qPCR, as described below. Leftover sample was stored at -80°C.

2.6.3 qPCR

For gene expression assays, the Taqman Universal PCR Master Mix (Applied Biosystems) was used, with predesigned Taqman Gene Expression Assays (Applied Biosystems). A list of taqman probes used can be seen in **Table 2.6**. Reactions were run on an AB StepOne Plus Real-Time PCR (Applied Biosystems), with the following programme:

Hold 1	50°C for 2 minutes
Hold 2	95°C for 10 minutes
PCR cycling (45 cycles)	95°C for 15 seconds
	60°C for 1 minute

Gene expression levels were calculated using the ΔCT method[96], after normalization to the housekeeping gene Glyceraldehyde 3-phosphate dehydrogenase (GAPDH).

Table 2.6: List of Taqman probes used

Gene	Probe code
CCND2	HS00153380_m1
CCR1	HS00174298_m1
GAPDH	HS03929097_g1
ITGB7	HS01565750_m1
MAF	HS04185012_s1
NUAK1	Hs00934234_m1
TLR4	Hs00152939_m1

For Chromatin Immunoprecipitation (ChIP) assays, in order to investigate enrichment of specific genomic loci, primers were designed using the Primer Blast function of NCBI ((https://www.ncbi.nlm.nih.gov/tools/primer-blast/). PCR product size was set at 60-300bp, with optimal primer melting temperature at 60°C, with a range allowed between 59-61°C. Primers were designed to have at least 4 mismatches to unintended off target sequences. 2-4 primer pairs per region were selected and tested for linear amplification and specificity by qPCR, using serial dilutions. The primer with best linear amplification was subsequently used (sequences in **Table 2.7**). For qPCR the SYBR Select Master Mix (Applied Biosystems), using the following programme:

Hold 1	50°C for 2 minutes
Hold 2	95°C for 2 minutes
PCR cycling (45 cycles)	95°C for 3 seconds
	60°C for 30 seconds
Melting curve	95°C for 15 seconds
	60°C for 1 minute
	95°C for 15 seconds

Enrichment of the amplicon was calculated using the ΔΔCT method relative to INPUT DNA (% INPUT) or to DNA immunoprecipitated by the IgG control antibody (Fold enrichment).

Table 2.7: List of primer sequences used ChIP qPCR

Genomic locus	Forward Primer (5'-3')	Reverse Primer (5'-3')
CCR1_promoter	AGTCAGCACTTTTCGGTGGG	CATGGGATCAGGCTTGCACT
CCR1_enhancer peak 2	GTTCCCTCCTGCTGAGGTTG	CCACAGAAATTGCCCACAGC
CCND2_gene body	ATGCGGTAGGCTCAGATGTC	GAGGATCTCCAAACAGCCTGG
IRF4_enhancer peak 1	TGTGGTGAGTAGGCCAGGAT	GACACTCCCTTTCCCCATGA
IRF4_enhancer peak 2	TTTCAGTGGTGAGACGGAGC	AAAGTGGGCACGGTCAGAAT
ITGB7_promoter	ATGGACAGGTGAGGATTCCG	AATTGGACGCCAAGATCCCA
TLR4_promoter	AGAATGCTAAGGTTGCCGCT	CTATCACCGTCTGACCGAGC
TLR4_enhancer peak 1	TGAGGAGTCTACTGCTCCCC	GAGCACAGCACTAGCCAGAT
MAF_promoter	TGCGAGATAACTCCGGGGAA	CGAGCGCGCAATATGAGTCT

2.7 RNA sequencing

Following RNA extraction as described in section 2.6.1, RNA quantity was determined using the Qubit RNA assay kit (Life Technologies). RNA quality was assessed using Agilent RNA 6000 Pico Kit and 2100 Bioanalyser (Agilent), with RNA integrity number (RIN) values > 8 considered of adequately good quality. Libraries were prepared using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs) and the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs), as per manufacturer's protocol. Samples were indexed using the NEBNext Multiplex Oligos for Illumina, Index Primers Set 1 and 2 (New England Biolabs) and all DNA purification steps were performed with Agencourt AMPure XP Beads (Beckman Coulter). Library DNA was quantified using the Qubit High Sensitivity DNA kit (Life Technologies) and library size was assessed using the Bioanalyser High Sensitivity DNA kit (Agilent). Libraries were diluted and pooled to 4nM-20nM, depending on available DNA quantities, and sequenced in the Imperial BRC genomics facility on an Illumina HiSeq 4000 platform to obtain 75bp paired-end reads.

2.8 Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq)

2.8.1 Transposase reaction

50,000 primary cells or transduced myeloma cell lines (MMCL) were used for ATAC-seq as previously described[97]. Briefly, cells were span down at 600g for 5 minutes at 4°C and washed once with 50µl of cold PBS. Cells were again span down, with care taken to remove all residual supernatant. Cells were then resuspended in 50µl of ATAC lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL CA-630) and immediately span down at 600g for 10minutes at 4°C. All supernatant was carefully removed and samples proceeded immediately to the transposase reaction. For this, cells were gently resuspended in a mixture of 25µl TD Tagment DNA Buffer (Illumina), 2.5µL TDE1 Tagment DNA enzyme (Illumina) and 22.5µL nuclease-free water and incubated at 37°C for 30 minutes. DNA was purified using the MinElute Reaction Cleanup Kit (Qiagen) and stored at -20°C.

2.8.2 ATAC-seq library preparation

Transposed DNA was thawed and PCR amplification (5 cycles) was performed using the NEBNext High Fidelity 2x PCR Master Mix (New England Biolabs) and Index Primers as previously described[97]. A 5µL aliquot of each sample was used to run a qPCR using SYBR Select Master Mix (Applied Biosystems), as per the protocol described in section 2.6.3. The qPCR amplification plot was used to estimate the additional PCR cycles required for the remaining samples (kept to <12 additional PCR cycles to avoid over-duplication). Amplified DNA was purified using a double 0.9-1X Agencourt AMPure XP Beads clean-up and eluted in 20µL of Tris-HCl. DNA library quality was determined using the Bioanalyser High Sensitivity DNA kit (Agilent). Libraries were quantified using the NEBNext Library Quant Kit for Illumina (New England Biolabs) as per manufacturer's instructions, on an AB StepOne Plus Real-Time PCR (Applied Biosystems). Samples were diluted to a minimum of 5nM and pools were

sequenced at the Imperial BRC genomics facility, on an Illumina HiSeq 4000 platform for 75bp paired-end reads.

2.9 Chromatin Immunoprecipitation (ChIP)

2.9.1 Chromatin preparation

Approximately $10x10^6$ cells were used per immunoprecipitation reaction, unless otherwise stated. Cells were harvested, washed once, resuspended in PBS and fixed by incubating in a 1% formaldehyde (Sigma-Aldrich) solution, at RT, with agitation for 15 minutes. The reaction was subsequently quenched by the addition of 1.25M Glycine (to a final concentration of 125mM) and placement of samples on ice. Cells were span down at 300g for 5 minutes at 4°C and washed twice with cold PBS containing 1:100 PMSF protease inhibitor (phenylmethylsulfonyl fluoride, Sigma-Aldrich). Cells were finally span down at 2600rpm for 5 minutes at 4°C and after carefully removing all supernatant, samples were snap frozen in liquid nitrogen and stored at -80°C.

Upon thawing, cells were lysed by resuspending in 100µl per $10x10^6$ cells of ChIP lysis buffer (1% SDS, 10mM EDTA, 50mM Tris-HCL pH 8, 1:100 protease inhibitor; Sigma-Aldrich), and incubating for 20 minutes on ice. Chromatin was subsequently sonicated at 4°C for 20 minutes in a 30 seconds on / 30 seconds off cycle, using the Bioruptor Pico Sonicaton System (Diagenode). Samples were subsequently span down at 13,000rpm for 10minutes at 4°C recovering the supernatant and discarding cell debris. A small aliquot was de-crosslinked overnight using NaCl 0.2M and heating to 65°C and was subsequently used to confirm efficiency of sonication at 250-500bp length on a 1.5% agarose gel. The rest of the sample proceeded to immunoprecipitation as described below.

2.9.2 Immunoprecipitation and DNA extraction

Chromatin immunoprecipitation was performed as previously described[98] with modifications. Chromatin was diluted 10 times with ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA, 16.7mM Tris-HCL pH 8, 167mM NaCl) containing 1:100 SIGMAFAST protease inhibitor (Sigma-Aldrich). Pre-clearing was performed by incubating the chromatin with BSA-blocked Dynabeads Protein A and Protein G (Thermofisher Scientific) for one hour at 4°C rotating. 1/10 of the sample was used as Input control and stored at -20°C whist the rest of the chromatin was immunoprecipitated with 5µg of antibody (or 2µg for $1x10^6$ cells), in an overnight incubation at 4°C, whilst rotating (see **Table 2.8** for a list of antibodies). Immunoprecipitated chromatin was then incubated with BSA-blocked Dynabeads Protein A and G for 3-4 hours at 4°C whilst rotating. The beads were subsequently washed twice sequentially for 5 minutes with each of Low Salt Buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl pH 8, 150mM NaCl), High Salt Buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl pH 8, 500mM NaCl), LiCl Buffer (0.25M LiCl, 1% IGEPAL, 1% sodium deoxycholate, 1mM EDTA, 10mM Tris-HCl pH 8) and TE (Tris-EDTA pH 8.0, Fisher BioReagents), whilst being kept cold at 4°C. Finally, immunocomplexes were eluted by incubating the beads overnight at 65°C with 75µl ChIP elution buffer (50mM Tris-HCl pH 8, 50mM NaCl, 1mM EDTA) and freshly added 1% SDS and 20µg/ml RNAse A (Thermofisher Scientific), whilst continuously agitated. At this point, input samples were also thawed and de-crosslinked with ChIP elution buffer to 200µl. Eluate from the beads was recovered the following day and a second elution was performed by incubating the beads for 2 hours with 25µl of chip elution buffer at 65°C whilst shaking. The two eluates were then merged and along with the input samples, incubated for 1 hour with 2µl Proteinase K (Thermofisher Scientific) at 45°C. DNA was purified using Agencourt AMPure XP Beads and eluted in 30µl of Tris-HCL.

Immunoprecipitated and input DNA were quantified using the Qubit High Sensitivity DNA kit and stored at -20°C. An aliquot of each sample was used for ChIP-qPCR using SYBR Select Master Mix as described in section 2.6.3. The rest of the samples proceeded to library preparation and sequencing as required.

Table 2.8: Antibodies used in ChIP experiments

Target Protein or isotype control	Host species / Clonality	Company	Catalogue Number	Clone
MAF	Rabbit polyclonal	Santa Cruz	sc7866	N/A
IgG isotype control	Rabbit polyclonal	Santa Cruz	sc2027	N/A
MAF	Rabbit monoclonal	Abcam	ab243901	BLR045F
IgG isotype control	Rabbit polyclonal	Cell Signalling Technology	CST 2729	N/A
IRF4	Rat IgG1 κ	Biolegend	646412	IRF4.3E4
IgG isotype control	Rat IgG1 κ	Biolegend	400401	RTK2071

2.9.3 DNA library preparation

DNA libraries were prepared using the NEBNext Ultra II DNA Library Prep kit for Illumina (New England Biolabs), as per manufacturer's instructions. Samples were indexed using the NEBNext Multiplex Oligos for Illumina (New England Biolabs) and all DNA purification steps were performed using Agencourt AMPure XP Beads. Library DNA was quantified using the Qubit High Sensitivity DNA kit and library quality and size was assessed using the Bioanalyser High Sensitivity DNA kit. Libraries were diluted to a minimum of 2nM and were sequenced in the Medical Research Council Genomics Facility on an Illumina HiSeq 2500 platform to obtain 50bp single-end reads.

2.9.4 ChIP re-ChIP

ChIP re-ChIP experiments were performed as previously described[99], with some modifications. Cells were cross-linked as outlined in section 2.9.1, with $10x10^6$ cells used per immunoprecipitation reaction. Cells were lysed with lysis buffer with the addition of 1mM PMSF protease inhibitor and sonicated as above. Chromatin was centrifuged at 13,000rpm for 10 minutes at 4°C, following which the supernatant was recovered and an aliquot for INPUT DNA was stored at -20°C. The rest of the chromatin was diluted with dilution buffer (1% Triton X-100, 2mM EDTA, 20mM Tris– HCl pH 8.1, 140mM NaCl) containing freshly added 1mM PMSF, and pre-cleared with BSA-blocked Protein A and G beads by incubating at 4°C on

continuous rotation for 1 hour. Chromatin was then divided into 4 aliquots; one for anti-MAF antibody (Abcam), one for anti-IRF4 antibody (Biolegend) and the other two for their respective IgG controls (**Table 2.8**). 5µg of each antibody was added to the chromatin aliquot and incubated overnight at 4°C, whilst rotating. The following day, 50µl of pre-blocked Protein A and G beads were added to each sample, and incubated for 3 hours at 4°C whilst rotating. For IRF4 reactions, the beads were then washed twice with each of dilution buffer, Wash Buffer A (50mM Hepes pH 7.9, 500mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Deoxycholate, 0.1% SDS), Wash Buffer B (0.25M LiCl, 0.5% NP40, 0.5% Deoxycholate, 1 mM EDTA, 20mM Tris–HCl pH 8.1) and TE (10 mM Tris–HCl pH 8.0, 2 mM EDTA) each containing 1mM PMSF and 1x SIGMAFAST protease inhibitor. For MAF reactions, wash buffers were as described in section 2.9.2. Beads were then resuspended in 75µl TE with 10mM DTT and eluted by incubating at 37°C for 30 minutes. Beads were then magnetically eluted and the eluate was then diluted 20 times with Dilution Buffer (+ 1mM PMSF). The IP reactions were now each split into 3 aliquots: one was stored at -20C to serve as INPUT DNA for the second reaction, one was incubated overnight at 4°C rotating with 5µg of the contralateral antibody, and the third, kept without antibody to serve as the control reaction. The third day, samples were incubated with Protein A and G beads and washed as the previous day, without the addition of protease inhibitor. Beads and INPUT DNA from day 1 were finally de-crosslinked by resuspending in ChIP elution buffer with freshly added 10µg/µ proteinase K and 10µg/µl RNase A and incubating at 65°C overnight. DNA was then purified using Phenol / Chloroform (Sigma-Aldrich) followed by ethanol precipitation, as per manufacturer's instructions. QPCR was performed using SYBR Master Mix, as described in section 2.6.3.

2.10 Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins (RIME)

Rapid Immunoprecipitation Mass Spectrometry (RIME) was performed by Active Motif from cell lysis onwards. Cells were crosslinked as per the company's protocol, as follows. $200x10^6$ cells per experiment were crosslinked by adding 1/10 of formaldehyde solution (16% Methanol Free Formaldehyde; Thermofisher Scientific, 0.1M NaCl, 1 mM EDTA, 50mM HEPES pH 7.9) to the media and incubating the cells for 8 minutes whilst agitating. The reaction was terminated by adding glycine to 125mM and incubating at RT for 5 minutes. Cells were then collected, centrifuged at 800g for 10minutes at 4°C and resuspended in 0.5% cold Igepal-PBS solution. Cells were centrifuged again and resuspended in 0.5% cold Igepal-PBS solution, this time also adding 1mM PMSF protease inhibitor. Cells were then centrifuged again, and cell pellet was snap frozen on dry ice. Following cell fixation, RIME experiments and mass spectrometry results analysis were performed by Active Motif. Briefly, cells were lysed with lysis buffer and homogenised using a Dounce homogenizer. Chromatin was sonicated to 300-500bp and pre-cleared using Protein G agarose beads (Invitrogen). Subsequently, 15µg of antibody was used to immunoprecipitate the protein, which was then captured using Protein G magnetic beads. Beads were washed and on-bead trypsinisation was performed to digest the proteins. The peptides were then purified using a C18 spin column (Harvard Apparatus) and were vacuum dried, before being analysed using a Thermo Scientific Q Exactive Orbitrap mass spectrometry instrument. Quantification of the chromatin yield was performed by decrosslinking a small aliquot of the cell lysate prior to pre-clearing and subsequently quantifying using NanoDrop. Each RIME reaction was performed using the rabbit monoclonal MAF antibody (Abcam; see **Table 2.8** above) and an IgG control, supplied by Active Motif. Two technical replicate reactions in the same cell line were performed. Analysis was performed by Active Motif, using PEAKS X software. Peptides spectrum match was performed at a cut off of -10logP > 20, giving an FDR of 1.5% using the "decoy-fused method" (**Appendix B Fig. S1**). Proteins that were considered significantly enriched were those present in the intersection of the results of the two MAF immunoprecipitation reactions, after subtracting the proteins present above the background cut-off (defined as ≤1 unique peptide and <5 peptide spectrum matches for a given protein) in the IgG reactions.

2.11 Bio-informatic analysis

Bio-informatic analysis of RNA-seq, ATAC-seq and ChIP-seq data was performed by Nikolaos Trasanidis, Imperial College London as described below. Primary sample analysis was performed by Dr Jaime Alvarez-Benayas, Department of Molecular Biology and Biotechnology, University of Sheffield, as part of a larger collaborative project[100]. Computational workflows used are outlined below.

2.11.1 RNA-seq

A schematic of the analysis pipeline is shown in **Appendix B Fig. S2**. In short, 75bp paired-end reads obtained per RNA sequencing experiment were demultiplexed (by the Imperial BRC genomics facility) and aligned against the GRCh38 human reference genome using "STAR 2.5.3a"[101]. Raw counts per gene were obtained with the "Rsubread" Bioconductor package[102]. In order to compare gene expression between the experimental conditions, the "DESeq2" Bioconductor package[103] was used to perform data normalisation and differential gene expression analysis across conditions. For knockdown experiments, the likelihood ratio test within the "DESeq2" package was used to determine differential gene expression across conditions (padj<0.01). For overexpression experiments, the wald test was used and a combined padj<0.05 and log2FC 0.2 cut off was applied. For primary sample analysis, performed as part of a larger collaborative project by Dr Jaime Alvarez-Benayas, genes with padj<0.05 and log2FC>1.5 were considered significantly differentially expressed[100]. For heatmap visualisation, log2 fold change (FC) values or TPM were z-score standardised and visualized using tools within the R environment. TPM values (z-score standardised) of differentially expressed genes between myeloma and normal PC samples were used to perform hierarchical clustering using the Spearman correlation method. Gene set enrichment and pathway analysis of differentially expressed genes was performed using the GSEA software 4.0.3[104]and the EnrichR online tool[105].

2.11.2 ChIP-seq

A schematic of the analysis pipeline is shown in **Appendix B Fig. S3.** Following demultiplexing (BRC Genomics Facility) and quality control using the FASTQC tool, ChIP-seq reads (50bp single-end) were aligned against the GRCh38 human reference genome using the Bowtie2 package[106]. After removing duplicate reads using Picard MarkDuplicates[107], peak calling was performed using MACS2[108], with signals analysed against INPUT DNA (parameters used: *-B -q 0.01 --SPMR --call-summits*). ChIP-seq replicates were combined following peak calling using the MSPC package (default parameters) and ChIPpeakAnno package to obtain overlapping peaks[109]. Data tracks and heatmaps depicting signal values normalised to fold change over input were generated using *deeptools* commands set[110]. The Homer package v4.9 was used for motif analysis and genomic annotation[111]. The package was also used for super-enhancer calling, using ChIP-seq against H3K27ac in MM.1S obtained from the ENCODE data portal (*ENCBS241TLI*). Data tracks were visualised with the Integrative Genomes Browser (IGV). The BETA-plus package enabled integration of ChIP-seq and RNA-seq data (settings: $c=1, df=0.05$), by assigning TF binding sites to the nearest differentially expressed gene (within 100kb). Two separate analyses were performed for each shRNA-cistrome combination; only genes with a rank score <0.05 in both shRNAs were considered regulatory targets of the transcription factor. Enrichment analysis was performed using the EnrichR online tool[105], Gene Set Enrichment Analysis (GSEA)[104] software and GREAT online tool[112], as indicated.

IRF4 ChIP-seq was obtained from the Young lab (GSM1195560). Due to the lack of INPUT DNA, data were analysed as for ATAC-seq described below. Co-binding between IRF4 and MAF was determined using *bedtools intersect*, with co-bound peaks considered those with at least one overlapping base pair. The *bedtools closest* package was used to obtain the distribution of distance between MAF and IRF4 summits.

2.11.3 Epigenomic annotation

Chromatin state annotation was performed by Nikolaos Trasanidis, as outlined below. Histone mark ChIP-seq peaks in MM.1S and primary plasma cells and B cells were accessed through ENCODE and Blueprint portals respectively (https://www.encodeproject.org/; http://blueprint-data.bsc.es/). The ChromHMM software package was used[113] to define the chromatin states for genomic loci in both MM.1S (25 chromatin states) and primary cells (12 chromatin states, as defined by the Blueprint epigenome project[114, 115]). Chromatin states generated were subsequently annotated based on ENCODE[116, 117], Blueprint epigenome project[115] and RoadMap Epigenome Consortium[118] guidelines and combined to form 6 final chromatin states (super-states) used to annotate the "MAF-associated" regions (**Appendix B Fig. S4**).

2.11.4 ATAC-seq

ATAC-seq data for primary samples were analysed by Dr Jaime Alvarez-Benayas as part of a collaborative project[100]. ATAC-seq reads (75bp paired-end) were aligned against the GRCh38 human reference genome using the Bowtie2 programme[106]. After removing duplicate reads using Picard MarkDuplicates[107] and mitochondrial chromosomes (chrM), peak calling was performed using the MACS2 package[108], with signals analysed against a bioinformatically computed background (parameters: *-g hs -q 0.01 --nomodel --shift -100 --extsize 200 -B --SPMR --keep-dup all --call-summits*). For primary samples, a comprehensive analysis strategy was followed to identify subgroup-specific, differentially accessible areas: (i) identify consensus regions set, (ii) calculation of depth-normalised enrichment of each sample on consensus regions set, (iii) likelihood ratio test with "DESeq2" package[103] to allow comparisons across patient subgroups versus normal donor (ND), by adjusting for batch effects, tumour status (diagnosis/relapse) and patient's sex. In this study, peaks with a padj<0.05 and log2FC>1 were considered differentially accessible between MAF and ND samples. Accessibility regions in primary samples were assigned to differentially expressed genes within 1Mb of the peak. MAF binding on the differentially accessible peaks was determined using *bedtools intersect*. For overexpression experiment analysis, performed by

Nikolaos Trasanidis, the ODIN package[119] was used to identify differentially accessible areas between MAF and EV samples (q value<0.01). Genomic annotation of peaks was performed using the Homer package. The *deeptools* package was used for signal files construction, signal tracks, heatmap and metagene plots visualisation. Differential ATAC signal (z-score standardised) was used to perform hiearchical clustering using the Spearman correlation method. Enrichment analysis was performed using the GSEA[104] software and GREAT online tool[112], as indicated. A schematic of the pipeline is shown in **Appendix B Fig. S5.**

2.11.5 Data from other public datasets

Gene expression across the haematopoietic cells, measured in FPKM, was made available from the Blueprint Consortium and obtained via the data analysis portal[120]. Gene expression profiles (FPKM) across primary myeloma cells was obtained from The Multiple Myeloma Research Foundation (MMRF) Compass study data portal[121].

2.12 Statistical analysis

Statistical analyses were performed using GraphPad Prism v6.0.1. One-way or two-way ANOVA, parametric t test or the non-parametric Mann-Whitney U test were used as appropriate, with a cut off for significance set at p value <0.05. All graphs depict mean and standard deviation (SD), unless otherwise specified.

2.13 Other software

Flow cytometry analysis was performed using FlowJo X.07 software. The SnapGene 3.2.1 software was used to plan cloning strategies and confirm cloned sequences. Schematics were created using the BioRender software (www.biorender.com). STRING networks were created on Cytoscape 3.7.2[122].

Chapter 3

MAF regulome in primary myeloma cells and myeloma cell lines

3.1 Introduction

MAF over-expression due to translocation t(14;16) occurs in 3-5% of multiple myeloma (MM) patients while in another 30% of patients MAF is overexpressed by alternative mechanisms [18, 45, 61]. The t(14;16) juxtaposes *MAF* to the chromosome 14 *IgH* enhancer and is considered to be a primary myeloma–initiating event; it occurs in the germinal centre B cells, from which MM is thought to arise, during somatic hypermutation in the immunoglobulin genes and IgH class switch recombination.

In MM, MAF is known to upregulate expression of four genes: cyclin D2 (*CCND2*), involved in cell cycle regulation and cell proliferation[28, 45], integrin β7 (*ITGB7*), an adhesion molecule mediating adhesion to bone marrow stromal cells and plasma cell migration[28, 45, 79], c-c motif chemokine receptor 1 (*CCR1*), mediating adhesion, proliferation, chemotaxis and migration out of the bone marrow niche[45, 80, 81, 84], and finally Nuak family kinase 1 (*NUAK1*), a kinase, that promotes plasma cell migration[28, 85]. These genes were identified as targets of MAF by knockdown and overexpression studies, and in the case of *NUAK1* and *CCND2* the presence of the MARE sequence at the gene promoters and the ability of MAF to transactivate them[45, 85].

In addition, gene expression profiling studies in cohorts of newly diagnosed MM patients compared gene expression between the different myeloma subgroups and identified a number of genes ("top 50") selectively deregulated in MAF-translocated myeloma[67, 68, 88]. However, whether these signature genes are directly regulated by ectopically over-expressed MAF has not been studied.

As these studies suggest and given the t(14;16) is a primary genetic event leading to aberrant expression of a transcription factor, I hypothesised that MAF would have a significant impact on the transcriptome and epigenome in plasma cells, driving critical aspects of myelomagenesis. Since there has been no study to date comprehensively delineating the MAF regulome in MM, the experiments described in this chapter aimed to address this question.

3.2 Pattern of MAF expression in myeloma and normal PC

3.2.1 MAF expression in myeloma cell lines

In the first set of experiments I set out to determine MAF expression in myeloma cell lines (MMCL) and primary myeloma samples and compare them to normal bone marrow plasma cells (PC). As expected and previously described[45], MAF expression was highest in MAF-translocated MMCLs, whereas some MAF expression was observed in MMCLs with the MMSET translocation [t(4;14; **Fig. 3.1A**)]. On average, MAF expression was 10-fold higher in t(14;16) MMCLs compared to MMSET-translocated H929 cells. No MAF expression was detected in t(11;14), CCND1-translocated cell lines. Expression of known MAF targets, *CCND2*, *ITGB7* and *CCR1*, was higher in MAF-expressing MMCLs (**Fig. 3.1A**), with highest MAF expression associated with higher target expression, though this was not absolute (**Fig. 3.1B**). Although these target genes are MAF-associated, they are not unique to MAF myeloma (explored further in section 3.3), and other transcription factors and regulatory proteins could be involved in their regulation, explaining these findings.

Next, I sought to confirm MAF expression across MMCLs at the protein level, using intracellular flow cytometry, a technique which has been previously used to assess MAF expression in T-regulatory and T-helper17 cells[52, 123] (**Fig. 3.2**). Reflecting MAF expression at mRNA level, the highest MAF protein was found in the MAF-translocated cell lines, followed by the MMSET cell lines, whilst no protein was detected in t(11;14) MMCLs.

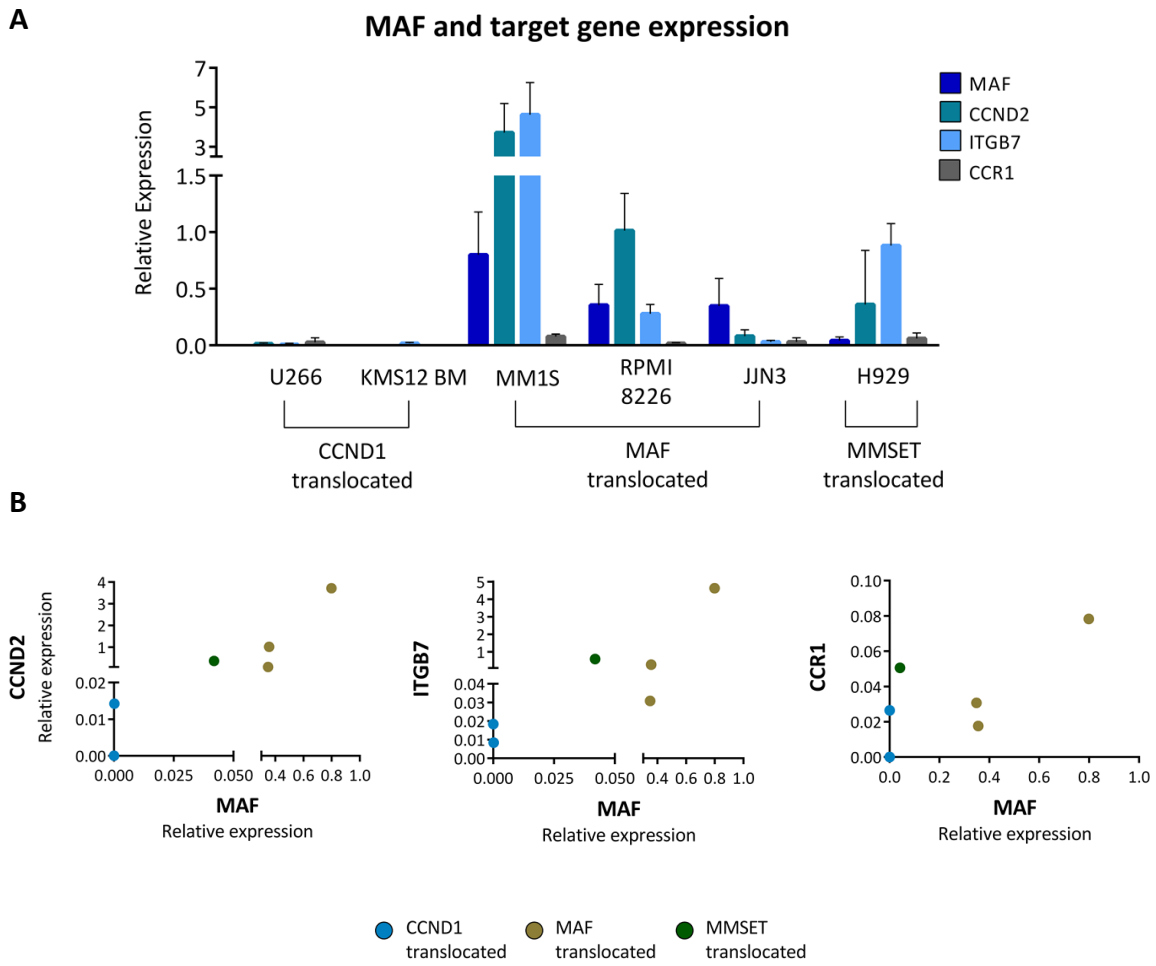


Figure 3.1: MAF and target gene mRNA expression in MMCLs

(A) qPCR assay of *MAF* and target gene expression in multiple myeloma cell lines. Expression normalised to *GAPDH*. Bars show Mean + SD, n=3. **(B)** Relative expression (normalised to *GAPDH*) of MAF against each of the target genes, *CCND2*, *ITGB7* and *CCR1*, in myeloma cell lines of different primary translocations.

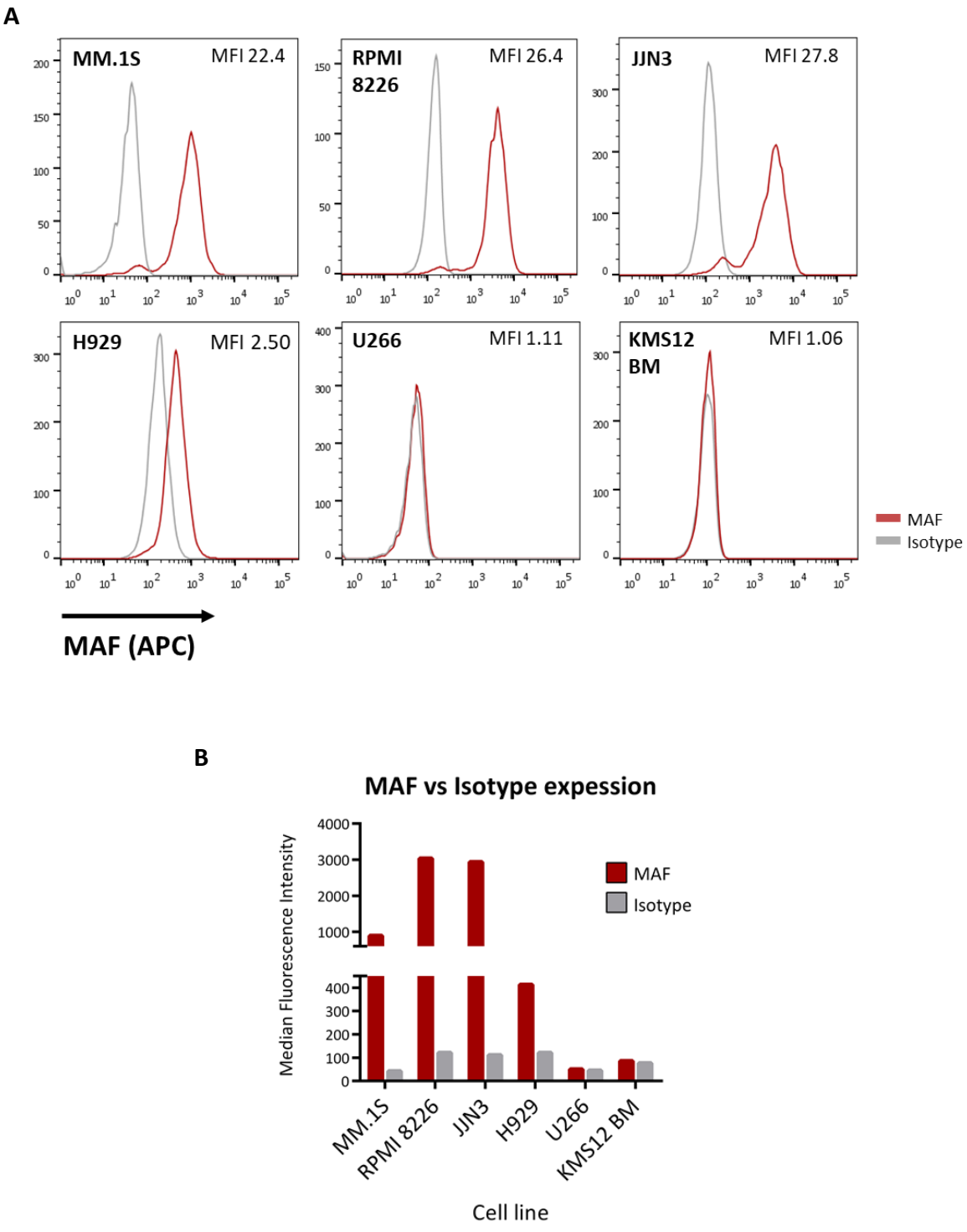


Figure 3.2: MAF protein expression in MMCLs

(A) Representative plots of MAF protein expression in myeloma cell lines, as assessed by flow cytometry. MFI; Median Fluorescence Intensity, expressed as ratio of MFI of MAF stain (red) over MFI of isotype control (grey). **(B)** Median Fluorescence Intensity of MAF (red) vs isotype control (grey) per cell line. Bars show mean of n=2 independent experiments.

3.2.2 MAF expression in primary myeloma cells

Next, I compared MAF expression in primary myeloma PC with that of MM cell lines and normal bone marrow PC.

To achieve this, I obtained purified bone marrow PC from MM patients, either at diagnosis or relapse, as part of a larger collaborative work[100]. PC purification was performed using standard CD138 positive immunomagnetic bead selection[124, 125]. Pre- and post-selection flow cytometry was performed to assess selection purity, using CD138 and CD38 markers (**Fig. 3.3**). Only samples achieving a post-selection PC purity >80% were used for downstream applications, with the MAF t(14;16) samples attaining over 95% purity (**Appendix B Table S1**).

As part of the same collaborative work, I obtained bone marrow PC from healthy haematopoietic stem cell transplantation donors, as controls. For this purpose, we established an antibody panel, as described in section 2.3.2, and following mononuclear cell isolation and CD3 and CD14 cell depletion, sorted the normal PC using fluorescence activated cell sorting (FACS). Gating strategy is shown in **Fig. 3.4**. T cells, monocytes and erythroblasts were excluded by negative gating on CD2, CD3 (T cells), CD14 (monocytes) and glycophorin-A (GPA; erythroid marker). PC were identified using CD38 and CD138 high fluorescent signal[126]. These cells were positive for CD27 (bright) and CD319 as previously described[126-128], contained both CD19+ and CD19- PC[129, 130] and had a variable expression of CD45, as is characteristic of normal PC[126].

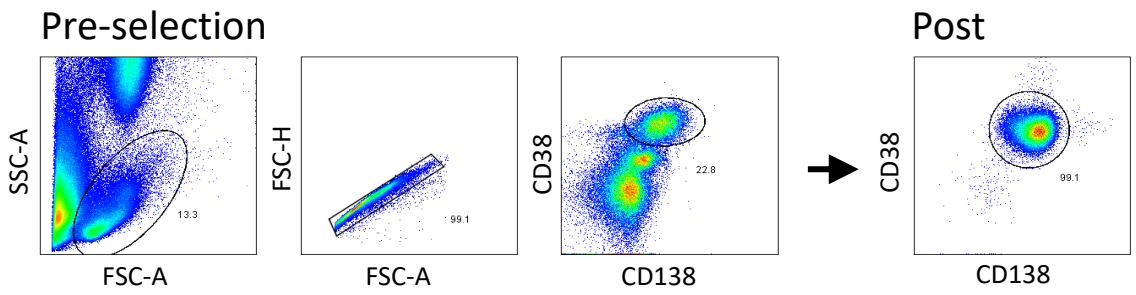


Figure 3.3: Pre- and post-selection flow cytometry of t(14;16) myeloma sample

Representative flow cytometry panels obtained pre and post CD138+ plasma cell selection from a primary myeloma sample, depicting gating strategy and post-selection purity based on CD138 and CD38 markers.

I then determined MAF expression in purified myeloma PC, and compared to that of normal PC and MMCLs, according to their primary genetic abnormality, using qPCR and RNA-seq assays (**Fig. 3.5**). As expected, t(14;16) primary samples had the highest expression of MAF, comparable but higher than the t(14;16) MMCLs. Some t(4;14) samples also expressed a small amount of MAF, as well as some hyperdiploid samples; levels in these samples were comparable to that of the t(4;14) H929 cell line. By contrast, MAF expression was not detected in normal plasma cells.

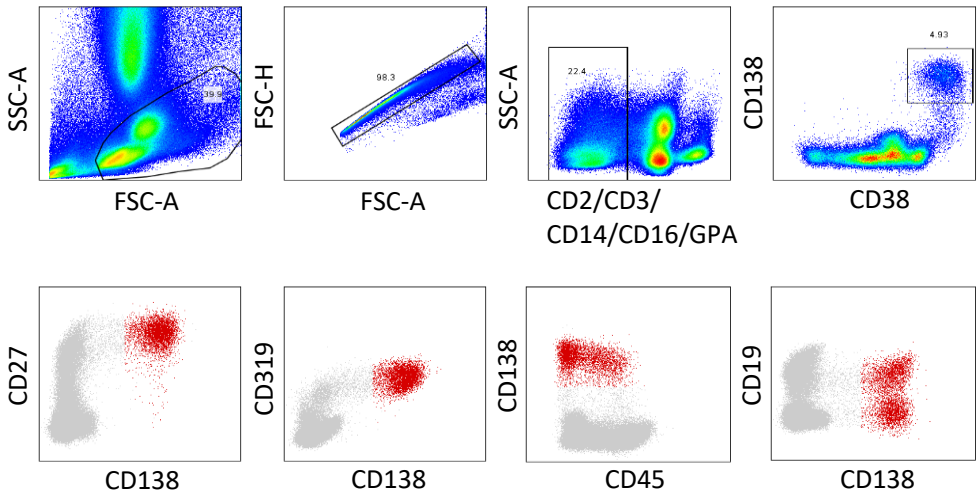


Figure 3.4: Healthy donor plasma cell flow cytometry

Representative flow cytometry plots obtained from FACS-sorting normal plasma cells from a healthy bone marrow donor, depicting gating strategy and signal of different flow cytometry markers in the normal plasma cells.

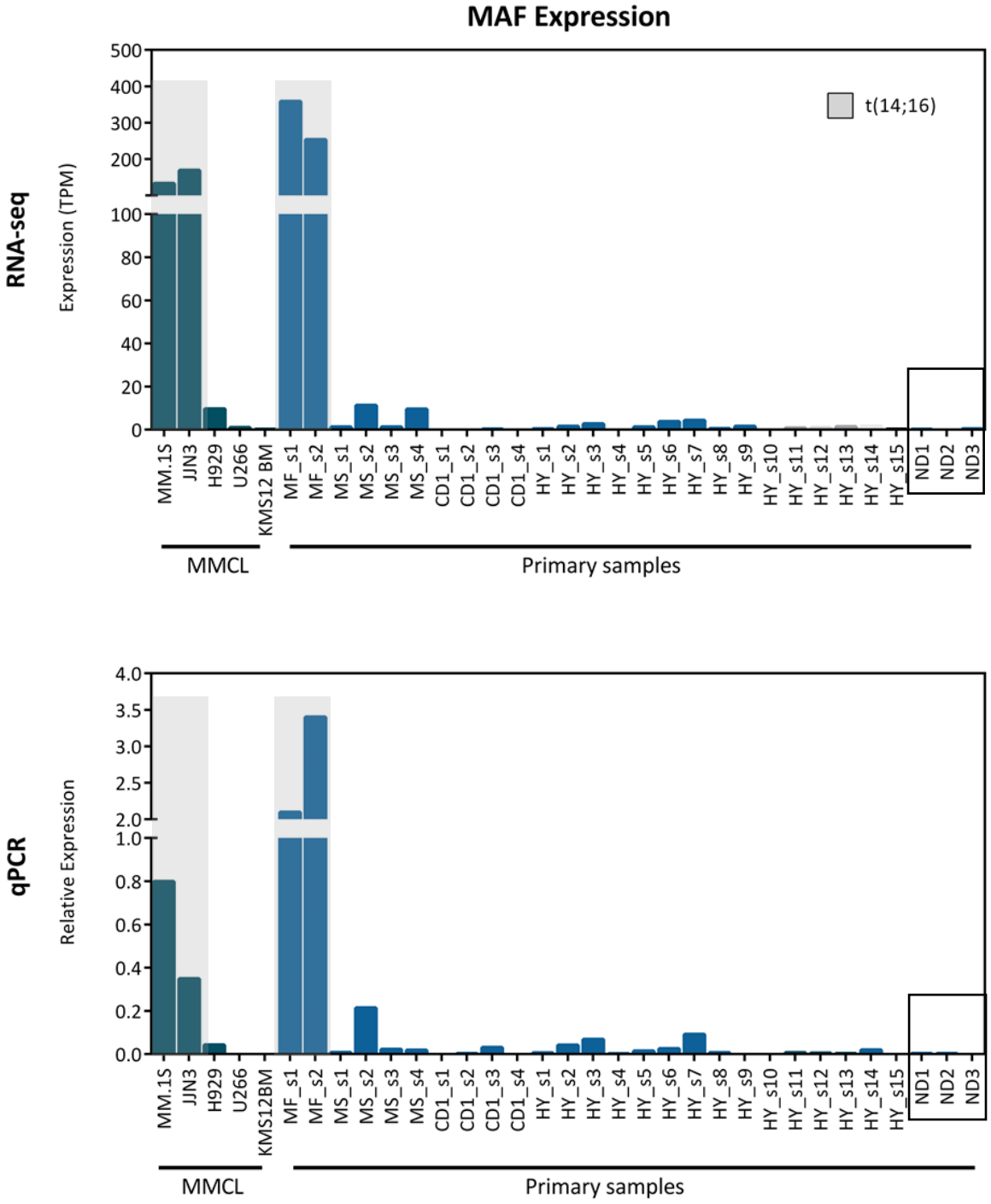


Figure 3.5: MAF expression in primary myeloma plasma cells

MAF expression by RNA-seq (top) and qPCR (bottom) in primary myeloma plasma cell samples (blue) compared to myeloma cell lines (green) and normal donor plasma cells (boxed). Shaded in grey are the MAF translocated cells. Expression by qPCR normalised to GAPDH. MF; MAF, MS; MMSET, CD1; CCND1, HY; hyperdiploid, ND; healthy donor normal plasma cells, MMCL; multiple myeloma cell lines.

3.3 Transcriptome of MAF-translocated primary myeloma PC

In order to explore genes associated with MAF in primary MM, I performed RNA-seq on purified PC from the two t(14;16) patients (MAF s1, MAF s2) and compared their gene expression to that of normal bone marrow PC (ND1, ND2, ND3). At a padj<0.05 and absolute log_2FC≥1.5 cut off, 904 and 1045 genes were up- and downregulated respectively in the MAF samples as compared to normal PC (**Fig. 3.6**).

Published gene expression profiling studies (GEP), have identified a set of genes, whose expression is significantly upregulated in MAF samples as compared to other MM genetic groups, identifying a MF molecular signature[28, 67, 68]. Notably however, these studies did not compare transcriptome differences between myeloma and normal PC. Several of the known MAF targets and MF molecular group associated genes, as described in GEP studies, were identified in the upregulated gene set of my samples (**Fig. 3.6**).

In order to confirm that my gene set was representative of larger MAF patient datasets, I performed gene set enrichment analysis (GSEA)[104] against the MSigDB C2 (Molecular Signature Database C2) containing curated gene sets from some of these GEP studies. Genes differentially expressed in the two MAF samples were highly enriched in genes of the MF molecular group[67] (**Fig. 3.7**). Indeed, of the genes found upregulated in the MF molecular groups described by Zhan et al.[67] and Broyl et al.[68], over 2/3 and 1/3 respectively were also differentially upregulated in my two samples. However, the results of these GEP studies are not directly comparable to the genes identified in my analysis, as the former compared between patient myeloma PC transcriptomes obtained by microarray technology whereas my study compared t(14;16) myeloma PC against normal PC using RNA-seq. Despite this, these results indicate the molecular profile seen in my samples is comparable to that found in larger collections of MAF MM patients, and therefore results are representative of the MAF MM group as a whole. By contrast, genes I found downregulated in the MAF samples did not enrich in the MF molecular group downregulated genes, indicating those results are less specific. Interestingly, the second most enriched gene set was that of angioimmunoblastic T cell lymphoma as compared to normal CD4+T cells[131], a malignancy that also overexpresses MAF.

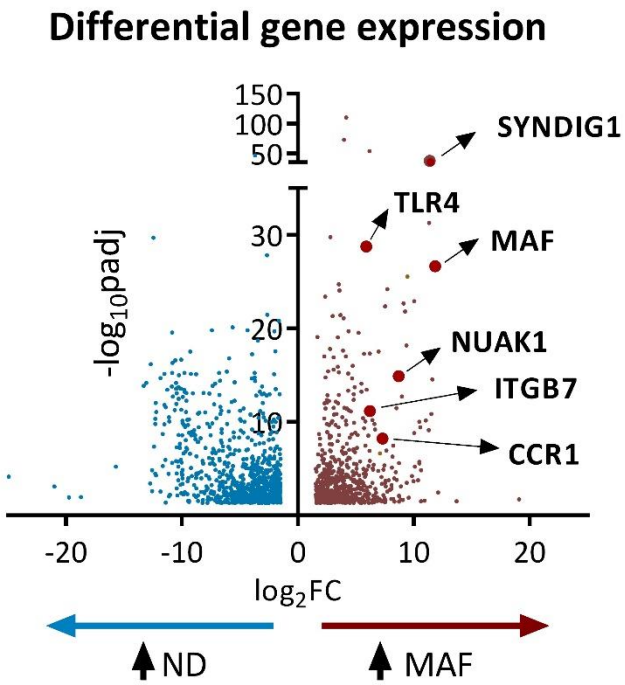


Figure 3.6: Differential gene expression in MAF primary myeloma

Genes found to be differentially expressed in MAF primary myeloma cells when compared to normal bone marrow plasma cells (padj<0.05, $log_2FC \geq 1.5$). In this Volcano plot, highlighted genes represent previously validated MAF target genes in myeloma (*NUAK1*, *CCR1*, *ITGB7*) or genes previously associated with the MF molecular group in gene expression studies.

In order to determine the extent of similarity or difference between the molecular profile of the MAF-translocated samples and other MM PC, I compared the differential gene expression of the MAF samples with that of myeloma samples belonging to different genetic subgroups obtained as part of a collaborative work in the lab[100]. First, to ensure these samples were representative of the MM genetic subgroups, as part of the same collaborative project the transcriptomes obtained were compared to that of 892 MM samples found in the Multiple Myeloma Research Foundation (MMRF) CoMMpass trial[121] (**Appendix C Fig. S1**)[100]. Reassuringly, our samples clustered well with the respective genetic subgroups of the larger cohort and indeed so did the t(14;16) samples, further suggesting that the two MAF-translocated samples used in my studies can be considered representative of t(14;16) MM. I then compared the differential transcriptome profiles obtained in the MAF samples to that of different MM genetic subgroups in our study. Hierarchical clustering analysis of the

differentially-expressed genes in the myeloma samples, revealed that MAF samples indeed clustered distinctly together (**Fig. 3.8**). Genes found to be differentially expressed in the MAF samples, could be shared between one or more other subgroups or be uniquely upregulated (and sometimes downregulated) in the MAF samples (**Fig. 3.8**). Thirty eight percent of MAF upregulated genes were uniquely upregulated in the MAF subgroup (followed by MMSET group with 33% differentially expressed genes, **Fig. 3.9A**). A much lower proportion of downregulated MAF genes were unique to the MAF group (16%), explaining why the genes didn't enrich for the top 50 downregulated genes of the MF molecular group identified in GEP studies. In addition, focusing on some of the genes previously identified as MAF targets or MAF-associated genes in the literature, I observed that many of them are upregulated both in the MAF samples as well as other samples, most frequently the MMSET group (**Fig. 3.9B**). This was not surprising as MAF is frequently found overexpressed in the context of a t(4;14)[62]. For example, *CCND2*, a previously validated MAF target was upregulated in all subgroups apart from the CCND1-translocated samples. *CCR1* and *NUAK1* on the other hand, were upregulated only in the MAF- and MMSET-translocated myelomas. By contrast *CX3CR1*, a gene previously identified as MAF-associated in GEP studies, is uniquely upregulated in the MAF subgroup.

Having identified the transcriptome profile associated with the MAF translocation in primary myeloma samples, and noting from the above observations that genes differentially upregulated in MAF samples when compared to normal PC may be either unique to t(14;16) myeloma or upregulated in other subgroups as well, I proceeded to identify the set of genes regulated by MAF on a genome-wide scale.

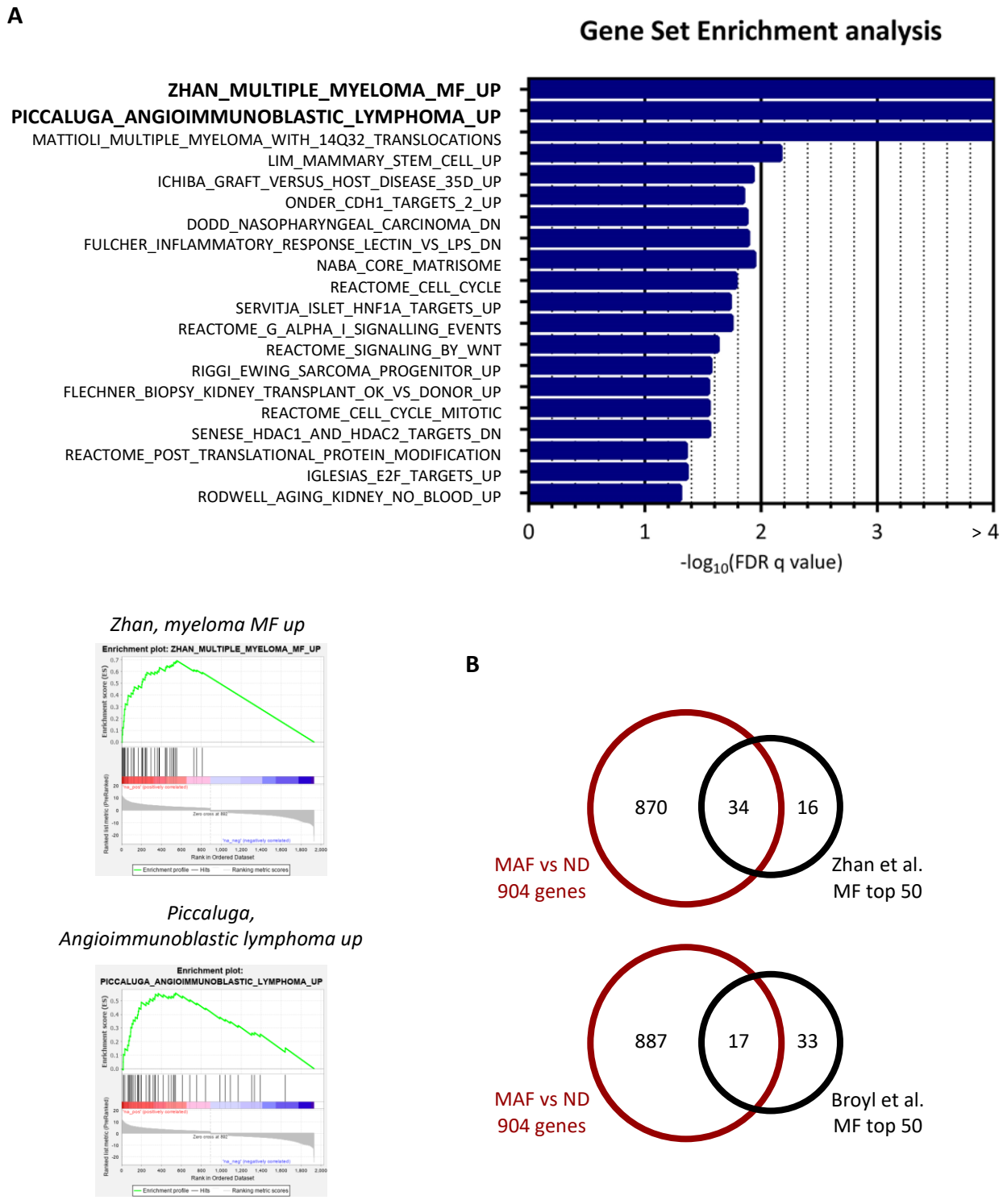


Figure 3.7: Gene set enrichment analysis of MAF upregulated genes in primary samples

(A) Gene set enrichment analysis performed using the GSEA software on the genes differentially upregulated in MAF primary samples (MAF s1 and MAF s2) compared to normal plasma cells (ND1, ND2, ND3). Only pathways reaching FDR q value<0.05 shown, with enrichment plots of the two most enriched pathways shown at the bottom. **(B)** Number of genes shared between the MAF upregulated genes and top 50 genes upregulated in the MF molecular group as identified by previous GEP expression studies in myeloma patients[67, 68].

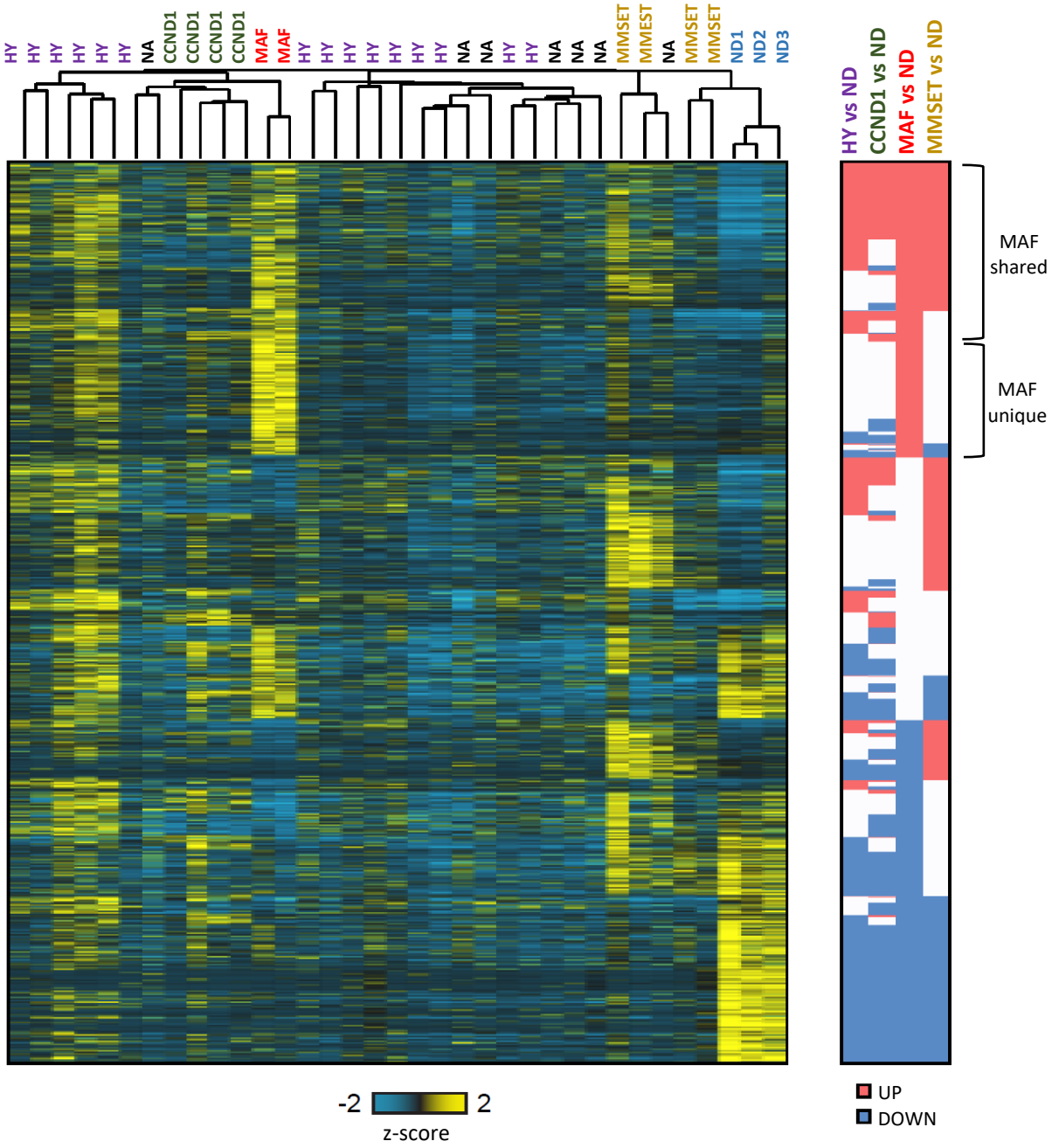


Figure 3.8: Hierarchical clustering on differentially expressed genes in myeloma PC

Heatmap of differential gene expression in primary myeloma PC as compared to normal plasma cells. Dendrogram showing hierarchical clustering, performed using differentially expressed genes in myeloma samples compared to ND (Spearman correlation method). On the right, genes whose expression (in TPM) is > 1.5 $\log_2$FC (UP, red) or <-1.5 $\log_2$FC (DOWN, blue) as compared to normal plasma cells. Brackets indicate genes differentially upregulated in the MAF samples that are either shared with other myeloma subgroups, or uniquely upregulated in the MAF samples. HY; hyperdiploid, ND; normal donor plasma cells, NA; others, includes samples with either no primary translocation detected or unknown cytogenetics.

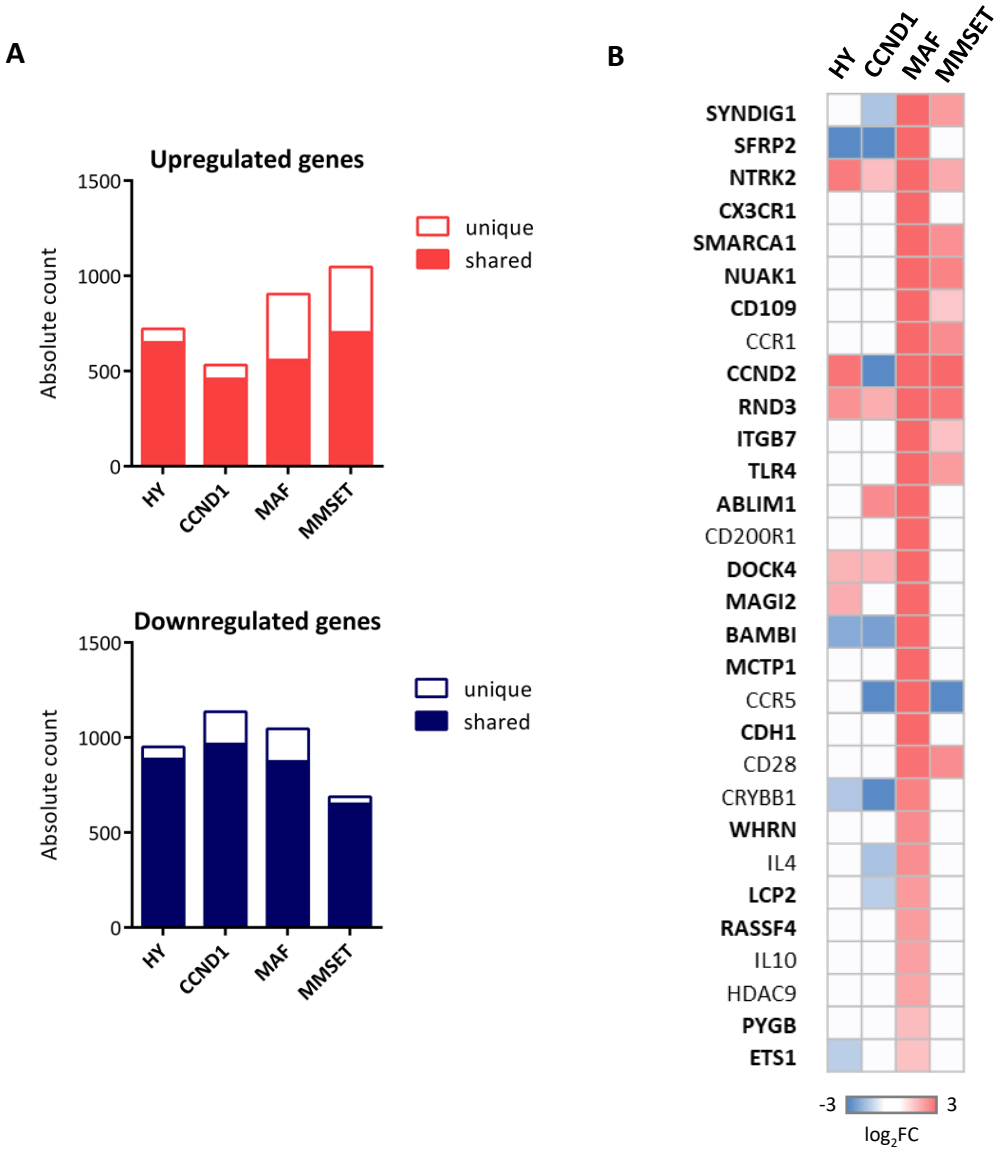


Figure 3.9: Comparison of MAF differentially expressed genes across myeloma genetic subgroups

(A) Number of genes differentially upregulated (top, red bars) or downregulated (bottom, blue bars) in the different MM cytogenetic subgroups as compared to normal plasma cells, identifying those that are uniquely differentially expressed in each subgroup. **(B)** Selected genes, (in bold genes previously identified as MAF-associated in GEP studies), with heatmap depicting expression log_2FC over normal plasma cells across genetic subgroups.

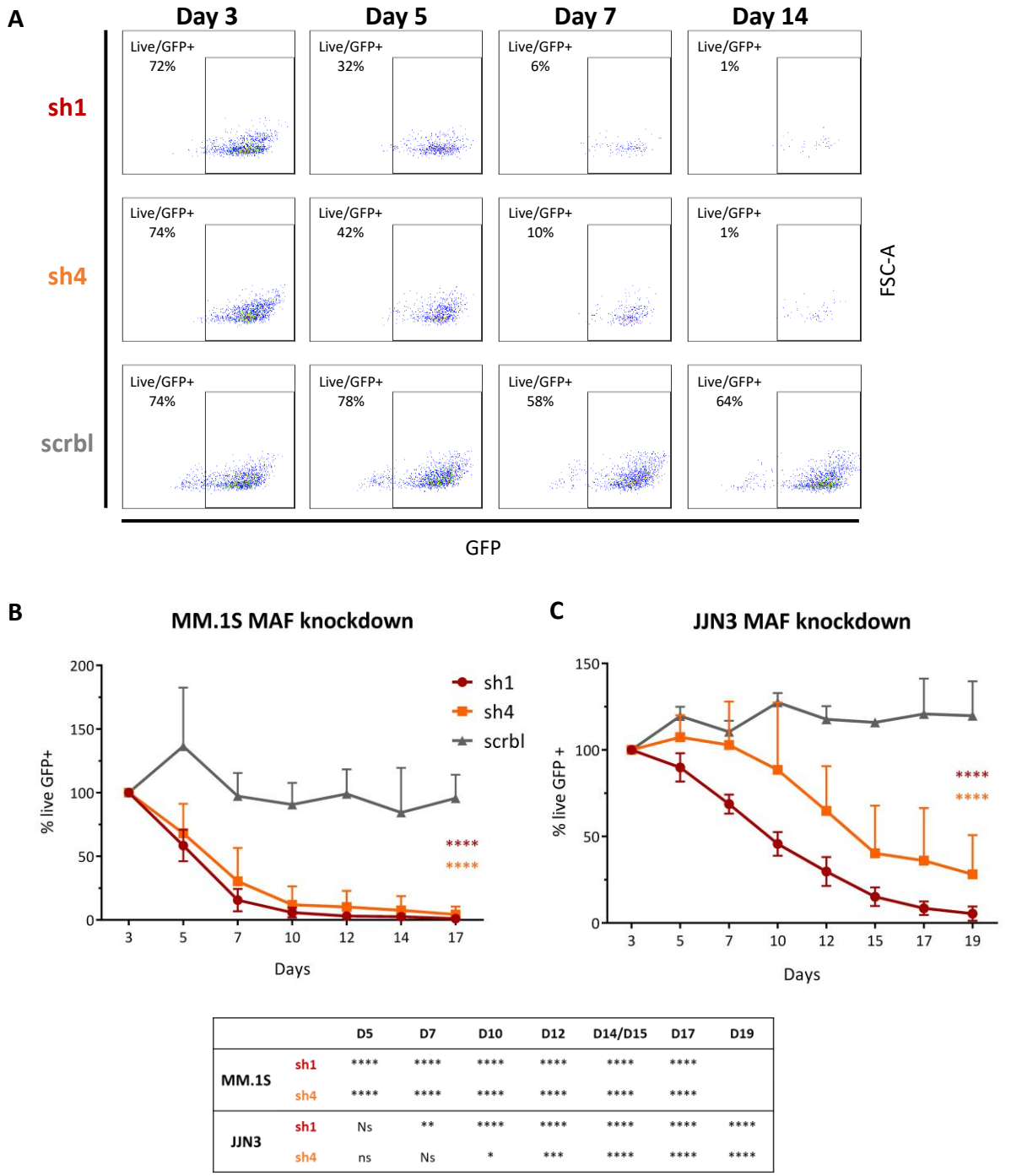
3.4 Functional validation of MAF targets

3.4.1 shRNA MAF knockdown

In order to identify the full spectrum of genes that are regulated by MAF in MM, I performed shRNA-mediated MAF knockdown in two t(14;16) MMCLs, using two shRNA constructs, sh1 and sh4, and a non-targeting scramble control (scrbl). Cloning was performed into the lentiviral GFP+ pLKO.1 vector, as described in section 2.4.1. Successful cloning was confirmed by sanger sequencing (**Appendix C Fig. S2)**, following which lentivirus was produced and used for cell transduction.

Hypothesising that MAF is essential for t(14;16) myeloma cell survival, cells were transduced and followed up over a period of several days using flow cytometry. Transduction efficiency was assessed on day 3 based on percentage of live GFP+ cells (**Appendix C Table S1**), using untransduced cells as controls to establish the gating strategy (**Appendix C Fig. S3**). Indeed, MAF depletion with either of the two shRNAs was detrimental to MM.1S cell viability (**Fig. 3.10A & B**). The same effect was observed in the JJN3 myeloma cell line (**Fig. 310C**), although JJN3 cells required a longer period compared to MM.1S for the same effect. This could be explained by the higher levels of MAF protein seen in JJN3. However, it could also be the result of the lower transduction efficiencies achieved in this cell line. In addition, sh1 appeared in both cell lines to be more potent in its effect on cell viability; this difference was more pronounced in JJN3 than MM.1S cells. By contrast, MAF depletion in a t(11;14) MMCL, the KMS12 BM, which lack MAF, did not affect cell viability (**Appendix C Fig. S4**).

To confirm effective MAF knockdown and assess the effect of this to the expression of known targets, cells were FACS-sorted on day 2 post transduction and qPCR was performed on the extracted RNA. A reduction in the expression of *MAF* and its known targets was confirmed in both MAF translocated cell lines (**Fig. 3.11**). By contrast, the expression of other genes such as the housekeeping gene *G6PD* or ubiquitin ligase *MDM2* was not reduced. MAF protein expression, assessed on day 2 and day 3, was also progressively reduced following shRNA-mediated knockdown (**Fig. 3.12**). The effect was more pronounced for sh1 than sh4, despite



		D5	D7	D10	D12	D14/D15	D17	D19
MM.1S	sh1	****	****	****	****	****	****	
	sh4	****	****	****	****	****	****	
JJN3	sh1	Ns	**	****	****	****	****	****
	sh4	ns	Ns	*	***	****	****	****

Figure 3.10: Viability of cells transduced with MAF shRNA

(A) Representative flow cytometry plots of GFP+ cells in the sh1, sh4 and scrbl control transduced MM.1S over time. Live/singlets/GFP+ gated cells shown, with gating strategy illustrated in Appendix C Fig. S3. Percentages indicate live GFP+ cells. **(B, C)** Viability of transduced MM.1S **(B)** or JJN3 cells **(C)** over time, as assessed by percentage of live GFP+ cells, normalised to day 3 of transduction. Legends for (B) and (C) are the same. * p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001, ns; non-significant, two-way ANOVA (Tukey's post hoc multiple comparisons correction) for n=3, two technical replicates per experiment.

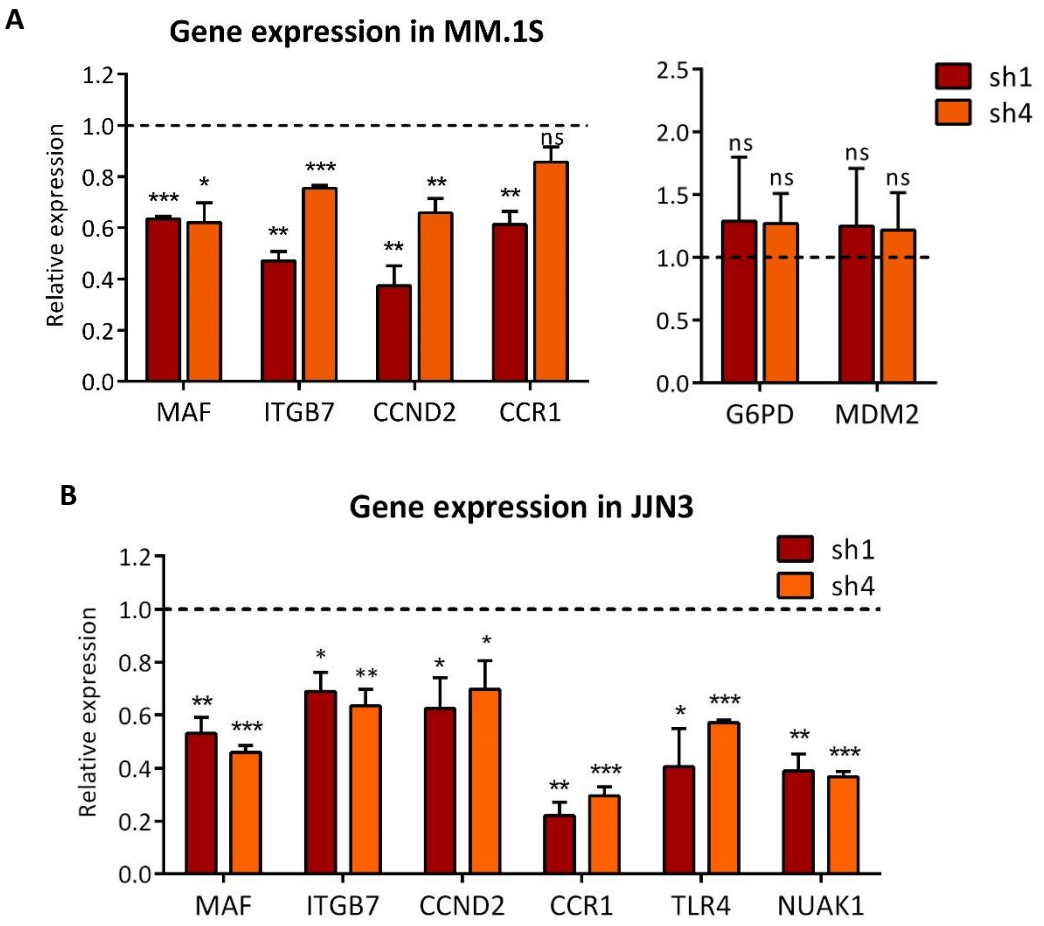


Figure 3.11: *MAF* and target gene expression upon MAF knockdown

Relative gene expression of *MAF* and target genes, normalised to *GAPDH* and scrbl control, as assessed by qPCR. Expression assessed on FACS-sorted GFP+ MM.1S **(A)** and JJN3 **(B)** on day 2 of transduction with MAF shRNA constructs, sh1 and sh4. By contrast, two indicative genes not known to be MAF targets, *G6PD* and *MDM2*, are shown on the right. Bars represent mean + SD for n=3 independent experiments. Statistical analysis comparing relative gene expression to scrbl control was performed using parametric t tests; * p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001, ns; non-significant.

MAF mRNA levels being similar between the two constructs. This result was in agreement with the difference seen between the two constructs in their effect on cell viability.

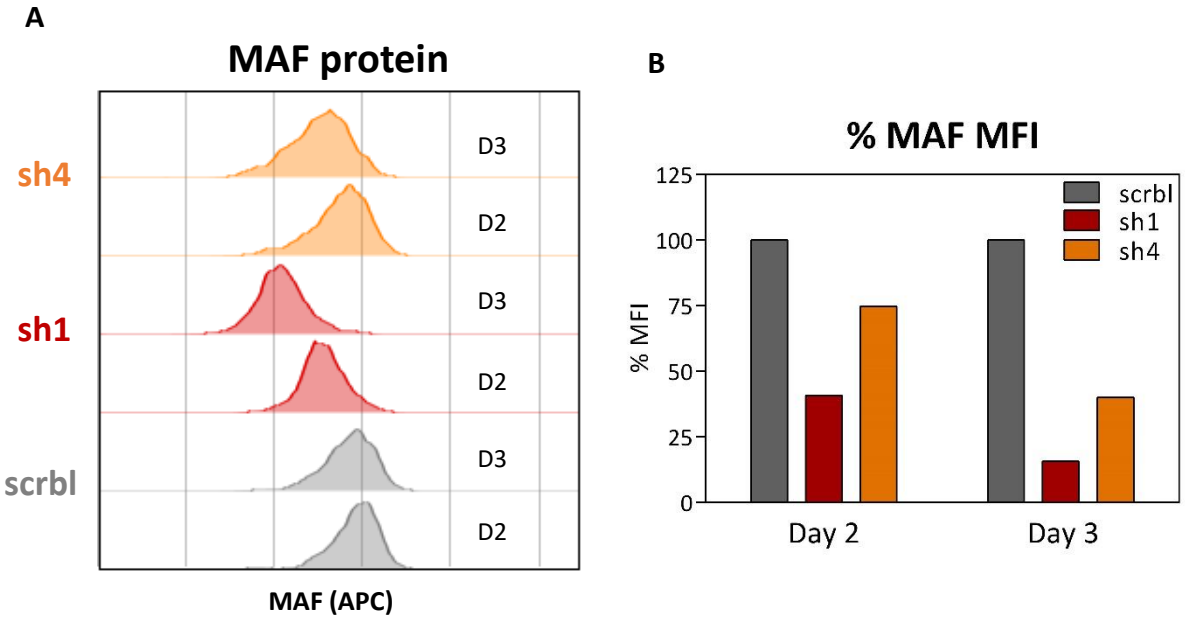
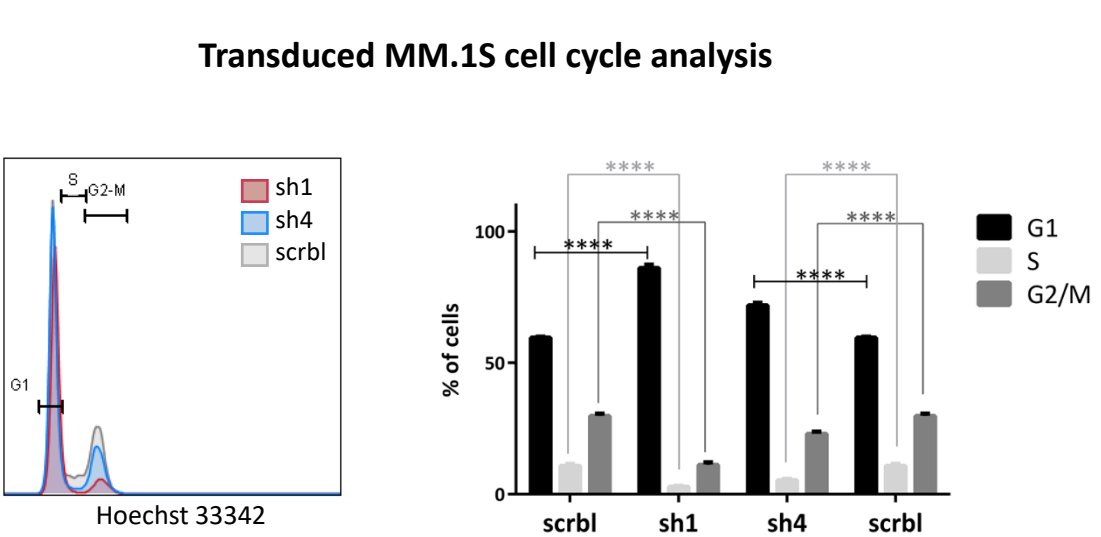


Figure 3.12: MAF protein expression upon MAF knockdown

MAF protein expression as assessed by flow cytometry on FACS-sorted GFP+ MM.1S cells on day 2 and day 3 of transduction with MAF shRNA constructs, sh1 and sh4. **(A)** Representative histograms of MAF expression for each construct on day 2 and day 3, compared to that of scrbl control. **(B)** Percent median fluorescence intensity, expressed as a ratio of MAF vs isotype control median fluorescence for each of sh1, sh4 and scrbl and then normalised to scrbl control. MFI; Median Fluorescence Intensity

Finally, it has previously been reported that MAF knockdown results in reduced cell proliferation[45]. To investigate this in my shRNA system, I performed cell cycle analysis on sh1 and sh4 transduced cells compared to scrbl control (**Fig. 3.13**). Transduced cells that were negative for both Annexin V (indicating apoptosis) and Propidium iodide (a marker for necrosis) were assessed for DNA content using the Hoechst 33342 dye. Consistent with previous findings, shRNA MAF knockdown cells showed an increase in the G1 and reduction in S and G2/M phase populations, consistent with cell cycle arrest.

A



B

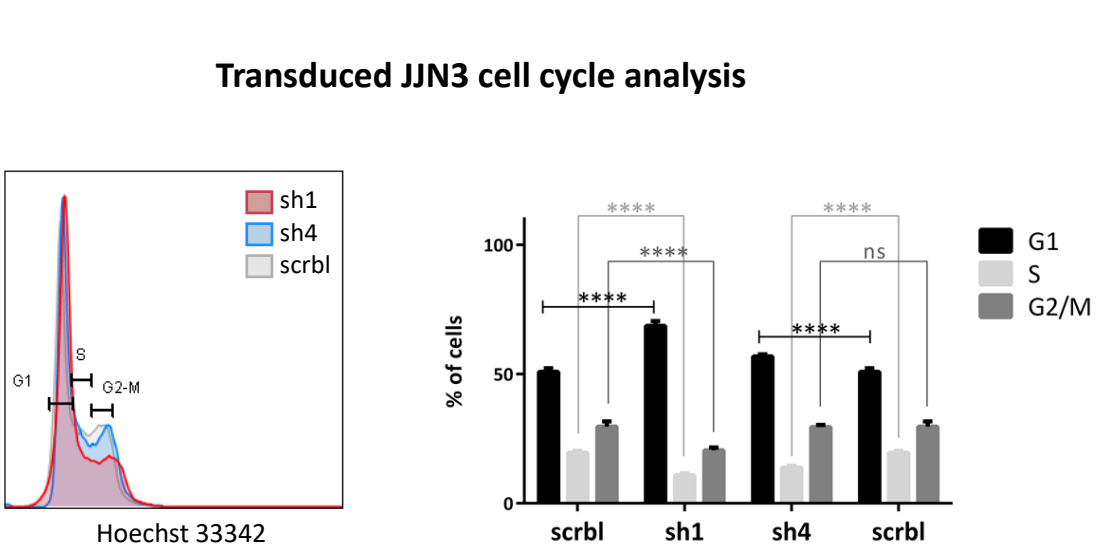


Figure 3.13: Cell cycle analysis upon MAF knockdown

Cell cycle analysis performed on transduced MM.1S **(A)** and JJN3 cells **(B)**, on day 6 of the experiment. Cells were gated on GFP+ and double negative for Annexin V and Propidium Iodide cells to examine transduced live cells. Hoechst stain was used to assess cell cycle. In each panel, an overlay of representative flow cytometry histograms is shown and bar graphs showing mean percentage (and SD) of cells found in each cell cycle stage. **** $p<0.0001$, ns; non-significant; two-way ANOVA (Tukey's post hoc multiple comparisons correction) for n=3 independent experiments.

3.4.2 MAF-regulated transcriptional programme

Having validated the downregulation of known MAF target genes using the shRNA knockdown system, I set out to establish the genome-wide transcriptional programme regulated by MAF in multiple myeloma cell lines. To that end, RNA sequencing was performed on FACS-sorted GFP+ sh1 and sh4 cells, and was compared to that of the scrbl control.

In order to establish the optimal time for cell harvest, a time course experiment was conducted, during which RNA was collected from FACS-sorted GFP+ cells on day 2, day 3 and day 5 post transduction (**Appendix C Fig. S5**). sh1 was observed to be more potent than sh4 in downregulating the MAF targets at every timepoint. This was in agreement with the observation on MAF protein reduction, where sh4 had a less potent and more delayed effect. In addition, as a general trend, targets were most highly downregulated on day 2, with sometimes an increase in their expression observed the following days. Based on these findings, RNA sequencing was performed on cells harvested on day 2.

RNA sequencing was performed on MAF shRNA knockdown MM.1S cells, transduced with sh1, sh4, and scrbl control, following FACS-sorting of GFP+ cells on day 2 of transduction. Two independent replicates were performed, with good correlation seen between the replicates in each construct, with greater variance observed in gene expression compared to scrbl control following sh1-mediated knockdown than sh4 (**Appendix C Fig. S6**).

Comparative analysis between genes significantly up- or down-regulated by MAF knockdown compared to scrbl (padj<0.01), revealed a total of 1,779 genes downregulated by both shRNAs (**Fig. 3.14A**). Conversely, less than half that number was seen upregulated, suggesting that in these cells, MAF has primarily an activatory role in gene expression. Amongst the most significantly downregulated genes were those previously validated as MAF targets: *NUAK1*, *CCND2*, *CCR1* and *ITGB7* were in the top 10% of downregulated genes, when ranked by differential z score (**Fig. 3.14B**). Furthermore, they were significantly enriched for genes identified by Zhan et al. as specifically upregulated in the MF molecular group[67] (**Fig. 3.14C**). By contrast, only sh1-upregulated genes were enriched for MF molecular group genes,

whereas those upregulated by sh4 or both shRNAs were not, suggesting the upregulated genes are less specific or form part of downstream pathways not differentially expressed in MF samples.

To validate these results further and investigate if downregulated genes following MAF depletion indeed reflect the MAF "signature" found in patients, I compared the genes downregulated by both shRNAs to the MF molecular profile identified by 3 GEP studies[28, 67, 68] (**Fig. 3.15**). Twenty-nine - 50% of the MF group upregulated genes, were also downregulated following MAF knockdown, indicating that a significant proportion of the genes differentiating the MF transcriptional profile from other myeloma subgroups are regulated by MAF. Nine out of 12 MAF signature genes identifying by combining GEP data[28] were also seen downregulated by MAF knockdown while 22-40% of genes downregulated in the MF group were conversely upregulated following MAF depletion.

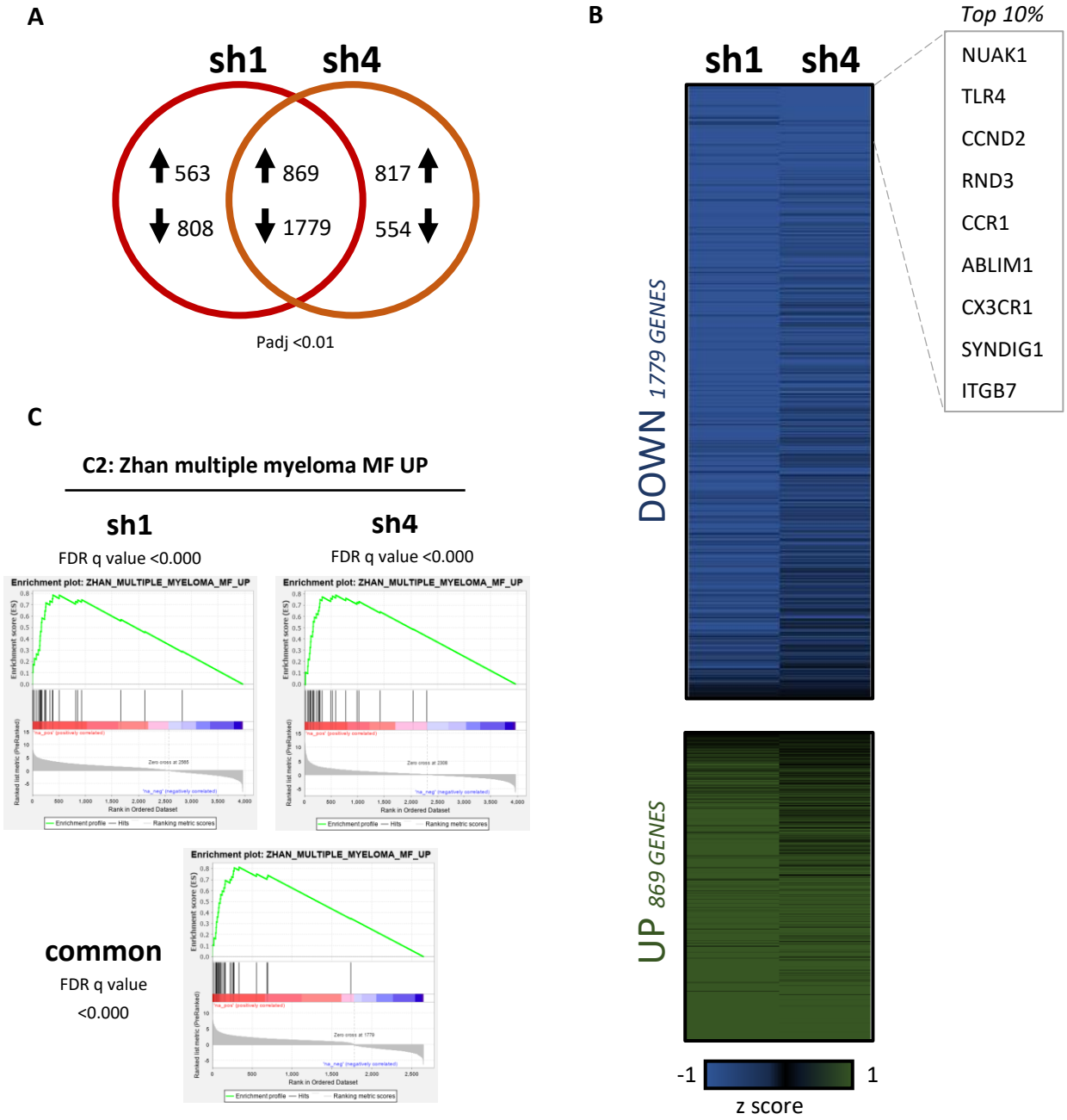


Figure 3.14: Differential gene expression upon MAF knockdown

Differential gene expression of anti-MAF sh1- and sh4-RNA, as compared to scrbl control, in transduced MM.1S cells. **(A)** Venn diagram of the differential genes up- and downregulated in sh1 and sh4, padj<0.01. **(B)** Heatmap of genes differentially regulated by both sh1 and sh4, arranged in order of their average z score. Box indicates genes within the top 10% known to have a role in MAF myeloma. **(C)** Gene set enrichment analysis of downregulated genes by sh1, sh4 and both, against a list of top 50 differential upregulated genes in the MF molecular group identified by Zhan et al[67], performed using the GSEA software[104].

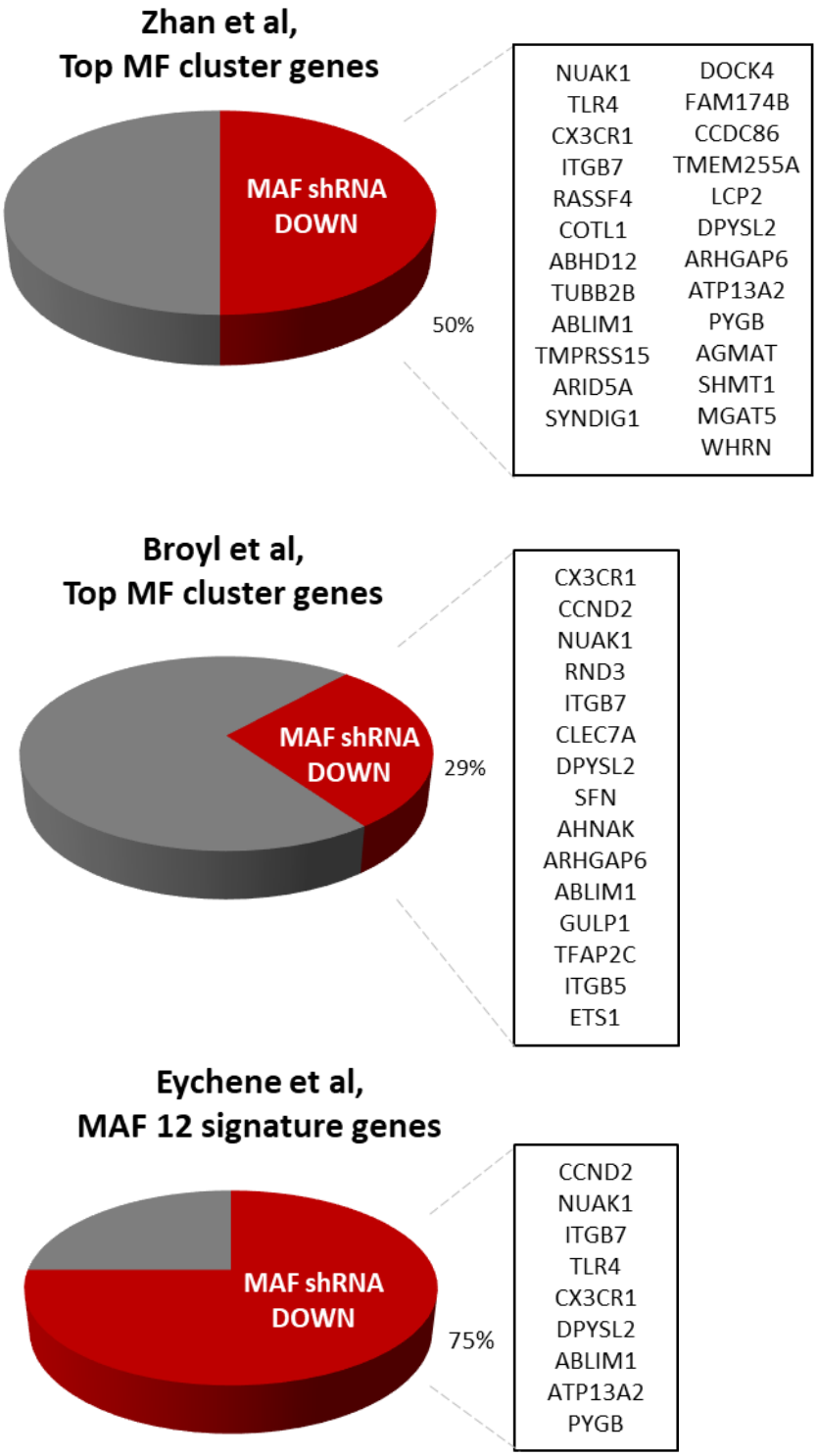


Figure 3.15: Comparison of anti-MAF shRNA downregulated genes with MF subgroup public datasets

Pie charts illustrating the proportion of genes upregulated in the MF molecular profile in GEP studies (conducted by Zhan et al.[67], and Broyl et al.[68] and a distilled MAF signature gene list from Eychene et al.[28]), also found downregulated by anti-MAF shRNA (shared between sh1 and sh4).

Gene pathway enrichment analysis was performed using the GSEA[104] and the EnrichR online tool[105]. Genes downregulated by MAF knockdown were significantly enriched in genes involved in cytokine interactions, the integrin signalling pathway, DNA replication and cell cycle (**Fig. 3.16**). These pathways had been predicted to be regulated by the MAF in multiple myeloma, when main MAF targets, such as *CCND2*, *ITGB7*, and *CCR1* were identified, and following identification of genes differentially expressed in MAF myeloma patients compared to other genetic groups, including *CX3CR1*[28, 45, 67, 68, 88]. However, these data comprehensively link the molecular pathways to the function of the MAF transcription factor in the context of multiple myeloma. On the contrary, genes upregulated following MAF knockdown were not found statistically enriched in any of the pathways interrogated, suggesting again non-specific upregulation.

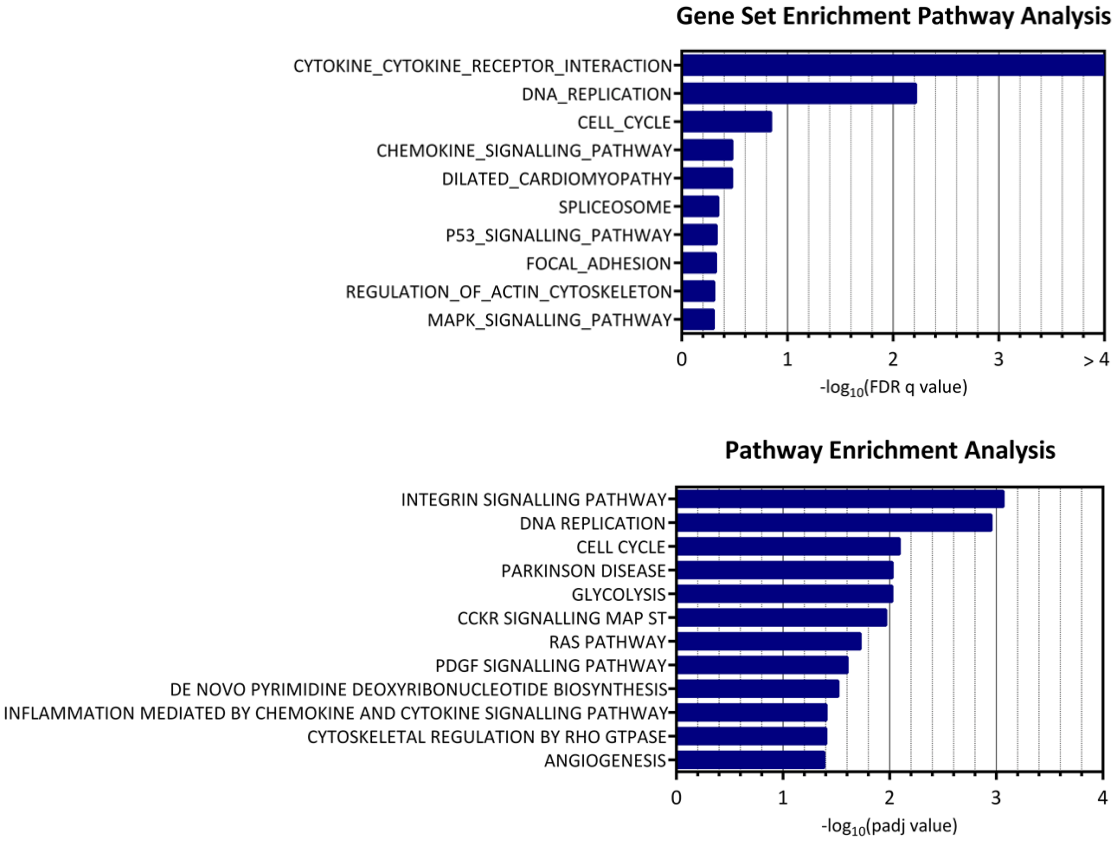


Figure 3.16: Pathway enrichment analysis of downregulated genes upon MAF knockdown

Pathway enrichment analysis of genes downregulated by both sh1 and sh4 in shRNA-mediated MAF knockdown, using Gene Set Enrichment analysis tool[104] (top) or EnrichR[105] (Panther; bottom). Only pathways with a padj<0.05 in the Enrich-R database are shown.

To further validate these results, shRNA mediated knockdown was performed in another t(14;16) MMCL, JJN3, shown earlier to be dependent on MAF for proliferation and survival (**Fig. 3.10C, 3.13B**). Following a time course experiment of target gene expression to identify the optimal time for cell harvesting (**Appendix C Fig. S7**), and in order to maintain consistency with RNA sequencing experiments in MM.1S, GFP+ cells were FACS-sorted on day 2 of transduction and RNA-seq was performed. In these cells, a much lower number of differentially up- or down-regulated genes was found (**Fig. 3.17A**), though the majority were shared between sh1 and sh4. This could be explained by the early time point of RNA collection during the transduction experiment; phenotypic changes in JJN3 were observed much later than MM.1S (**Fig. 3.10 and Fig. 3.13**). As a result, day 2 was likely too early a time point for RNA collection in transduced JJN3. Despite this, genes differentially affected by both sh1 and sh4 were enriched for MF molecular group genes, as identified by Zhan et al.[67], such as *NUAK1*, *TLR4* and *ITGB7* (**Fig. 3.17B**). This suggests that despite cells having been harvested early, the transcriptome profile seen upon *MAF* knockdown is MAF-specific. Due to the small number of genes involved, GSEA was unable to generate statistically significant enriched pathways. However, pathway enrichment analysis using the EnrichR online tool[105] yielded similar results to the knockdown experiments on MM.1S cells, and identified pathways such as integrin signalling and cytokine signalling as significantly enriched in the downregulated genes (**Fig. 317C**). Interestingly, the T cell activation pathway was found to be most significantly enriched, driven by NF-kB and MAPK proteins, as well as CD28, a surface glycoprotein known to provide co-stimulatory signals for T cell activation [132] and Lymphocyte Cytosolic Protein 2 (LCP2), involved in T cell receptor mediated signalling. Indeed, MAF is known to physiologically regulate Th2 cell development and activation[47, 48]. Although many of these genes were also found downregulated by MAF knockdown in MM.1S, this pathway did not reach statistical significance.

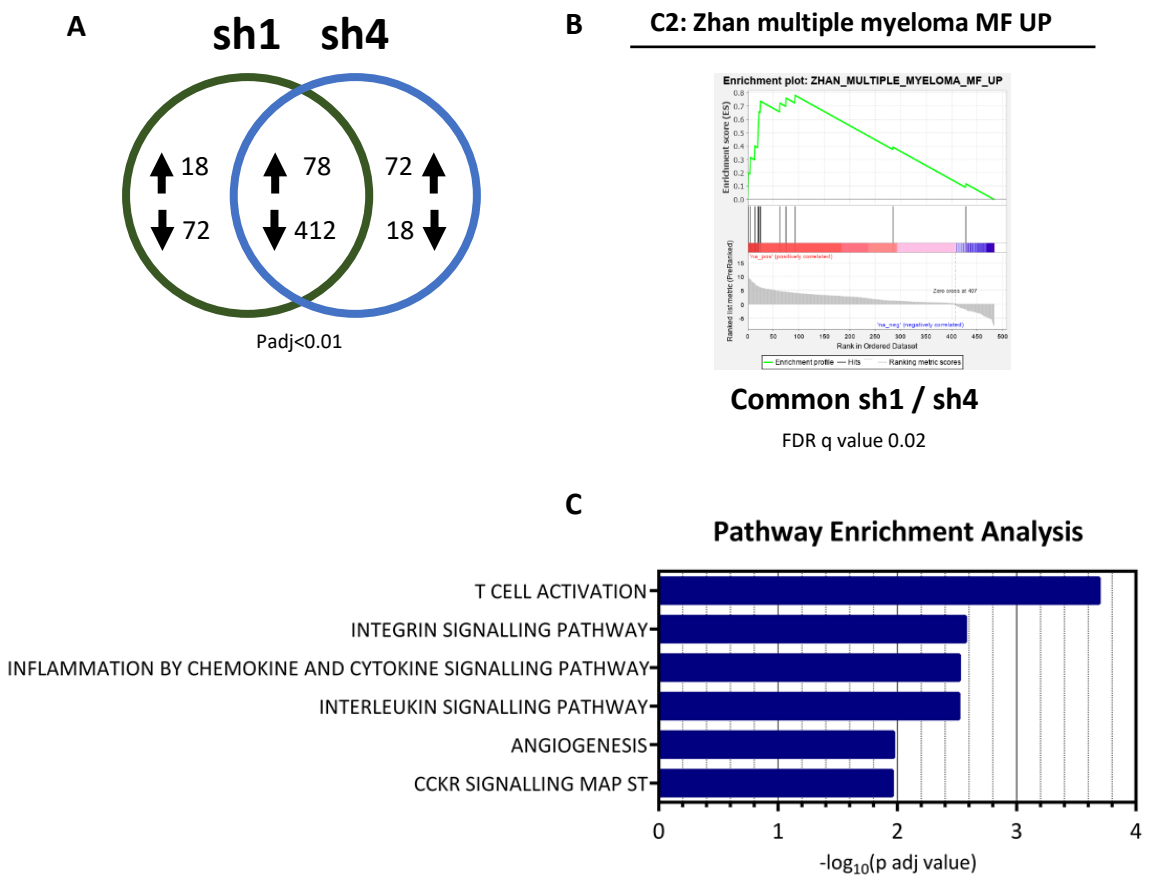


Figure 3.17: Gene expression analysis following MAF knockdown in JJN3 cells

Differential gene expression analysis performed on JJN3 cells, transduced with anti-MAF sh1 and sh4-RNA, or scrbl control. **(A)** Venn diagram of genes differentially up- or downregulated. **(B)** Gene set enrichment analysis of genes differentially regulated by the anti-MAF shRNAs, against the MF subgroup top upregulated genes as identified by Zhan et al.[67], performed using the GSEA software[104]. **(C)** Pathway enrichment analysis of the same genes against Panther pathways, obtained using the EnrichR online tool[105].

Combining the genes differentially up- or downregulated following MAF knockdown in the two cell lines, MM.1S and JJN3, revealed a list of common genes regulated by MAF; the majority of which were upregulated by the transcription factor (**Fig. 3.18**). Amongst these genes were genes previously identified to be associated with the MF molecular group of patients, some not previously validated as MAF targets, such as toll-like receptor 4 (*TLR4*) and Rho Family GTPase 3 (*RND3*) implicated in tumour cell proliferation and migration[133]. In addition, new functional MAF targets in multiple myeloma were also identified, such as Friend Leukaemia Integration 1 (*FLI1*) TF, *CD28* and *IRF4*.

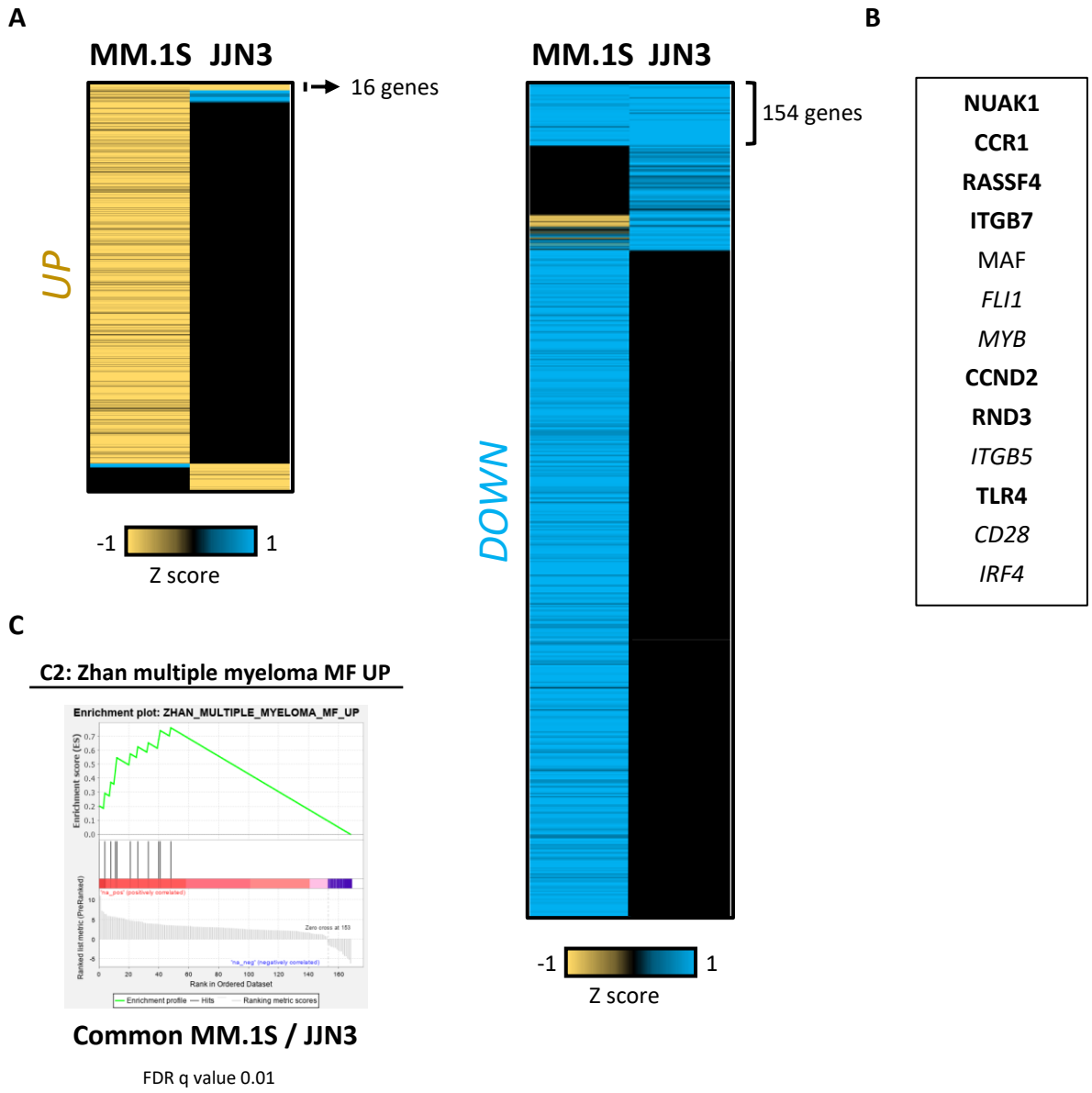


Figure 3.18: MAF transcriptional signature identified following MAF knockdown in MM.1S and JJN3 cells

Genes differentially up- and downregulated by anti-MAF sh1- and sh4-RNA, compared to scrbl control, in the MM.1S and JJN3 MAF-translocated cell lines. **(A)** Heatmaps depicting genes up- and downregulated in each cell line, and shared between the two. **(B)** Selective gene list of differentially downregulated genes common to both cell lines, indicating previously known MAF targets (*CCND2*, *NUAK1*, *CCR1*, *ITGB7*), previously associated with the MF molecular group (bold), and not previously linked to MAF in myeloma (italics). **(C)** Gene set enrichment analysis of genes differentially downregulated by MAF knockdown in both cell lines, against the MF subgroup top upregulated genes as identified by Zhan et al.[67], performed using the GSEA software[104]

3.5 MAF cistrome in MAF-expressing myeloma cell lines

Having identified the genes that are directly or indirectly regulated by MAF, I aimed to refine this further by identifying the set of genes directly regulated by MAF. As a first step, I performed ChIP-seq against MAF in three MMCLs, (as described in section 2.9). Of note, the genome wide binding profile of MAF in MM has not been reported to-date.

Using two anti-MAF antibodies, MAF G (Santa Cruz, sc10017) and MAF R (Santa Cruz, sc7866)[42], I first performed ChIP-qPCR assays in the MAF-translocated MM.1S and in the MAF-negative U266 t(11;14) MMCL as a negative control (**Appendix C Fig S8**). Both antibodies demonstrated enrichment in the *ITGB7* and *CCND2* loci, although there was some evidence of non-specific binding by MAF G, as indicated by modest enrichment in the *ITGB7* locus in U266 cells. Three independent replicates of each ChIP-seq reaction in MM.1S were performed. Comparing the motif analysis of each reaction, enrichment in the MAF family motif for the MAF G antibody was much lower than that observed in the MAF R reactions (**Appendix C Table S2**), suggesting less specific binding of this antibody. In addition, MAF G produced broader, less sharp peaks (**Appendix C Fig. S9**), again suggesting non-specific DNA binding. MAF G reactions were therefore excluded from downstream analyses. Focusing on the three MAF R reactions, the latter two replicates correlated well with each other, whereas the first replicate didn't correlate with the other two (**Appendix C Fig. S10**). Enrichment score for the MAF family motif was also low in this reaction (**Appendix C Table S2**). Therefore, the first MAF R ChIP-seq assay was also excluded, with the latter two replicates being used for downstream analyses. A further two immunoprecipitation reactions were then performed using this antibody on JJN3 cells, which are also MAF-translocated, as well as one replicate in H929 cells, a t(4;14) MMCL that overexpresses MAF.

The number of significantly enriched MAF binding sites was similar for the two t(14;16) MMCLs (**Fig. 3.19**). However, the total number of peaks obtained in H929 cells was much lower. Given that MAF mRNA and protein expression levels are approximately 10-fold lower in H929 than in t(14;16) cells (**Fig. 3.1, 3.2**), this finding suggests that the amount of transcription factor present affects its genome-wide binding.

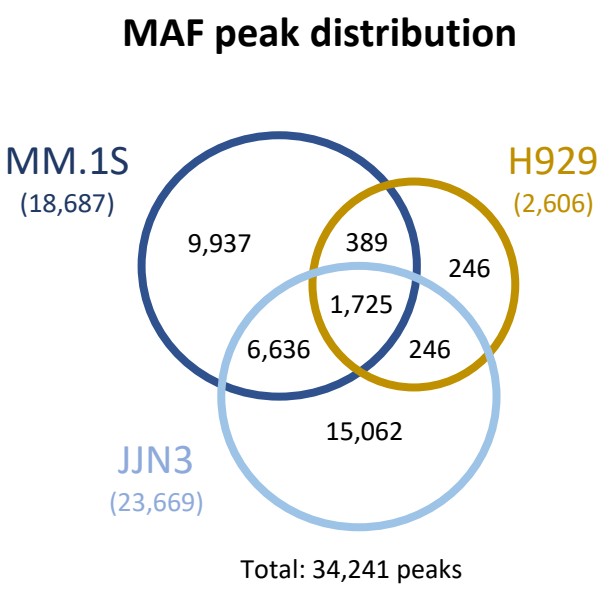


Figure 3.19: Peak distribution of MAF binding across three myeloma cell lines

Number of MAF-bound peaks obtained from ChIP-seq against MAF in three MMCL. n=2 ChIP-seq experiments were performed in the MAF-translocated cell lines, MM.1S and JJN3, and n=1 in the MMSET-translocated myeloma cell line, H929.

Focusing on MM.1S, a total of 18,687 MAF binding sites were identified (**Fig. 3.20**). Interestingly, genomic annotation of these peaks revealed that the majority of MAF binding sites are situated in intergenic and intronic regions (45% and 47% respectively), as opposed to promoters, suggesting that MAF preferentially binds enhancers. *De novo* motif analysis showed the MAF motif as the most significantly enriched. Integrating the MAF cistrome in MM.1S with histone mark maps in the same cell line made available through the ENCODE consortium project[117], revealed that MAF binding sites are enriched in activatory marks associated with promoters and enhancers (H3K27ac, H3K4me1, H3K4me2, H3K4me3, H3K9ac), whereas conversely did not display enrichment in marks signifying transcriptional repression such as H3K9me2, H3K9me3 and H3K27me3 (**Fig. 3.21**). Superenhancer calling using the H3K27ac ChIP-seq in MM.1S and juxtaposing this against the MAF bindings sites interestingly revealed that 92% of the MM.1S identified superenhancers, are also MAF-bound.

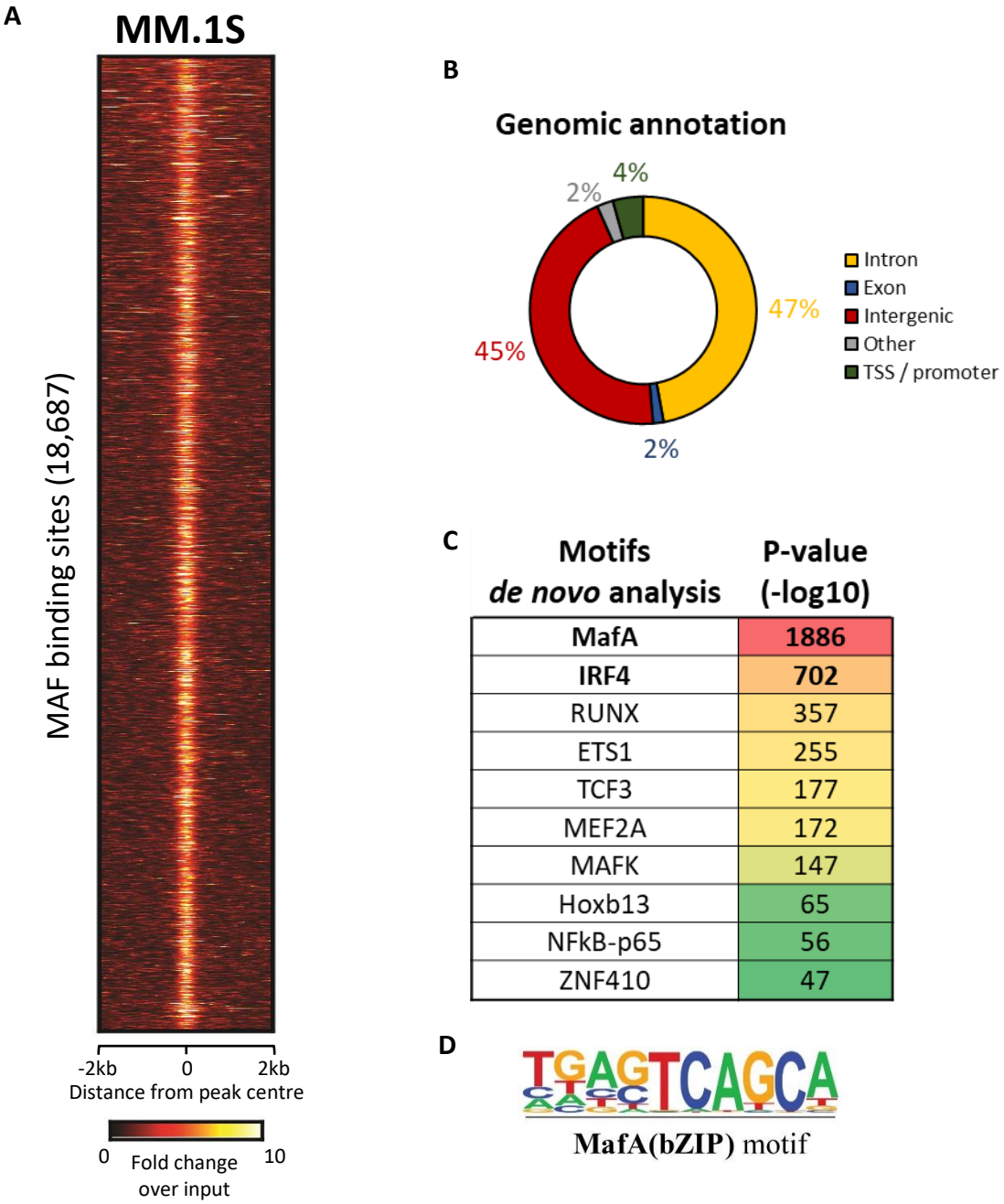


Figure 3.20: MAF binding in MM.1S

(A) Heatmap of MAF binding sites in MM.1S, depicted in fold change over input. **(B)** Genomic annotation of MAF ChIP-seq peaks **(C)** *De novo* motif analysis of ChIP-seq peaks **(D)** MAFA weblogo, the most enriched motif obtained. TSS; transcription start sites

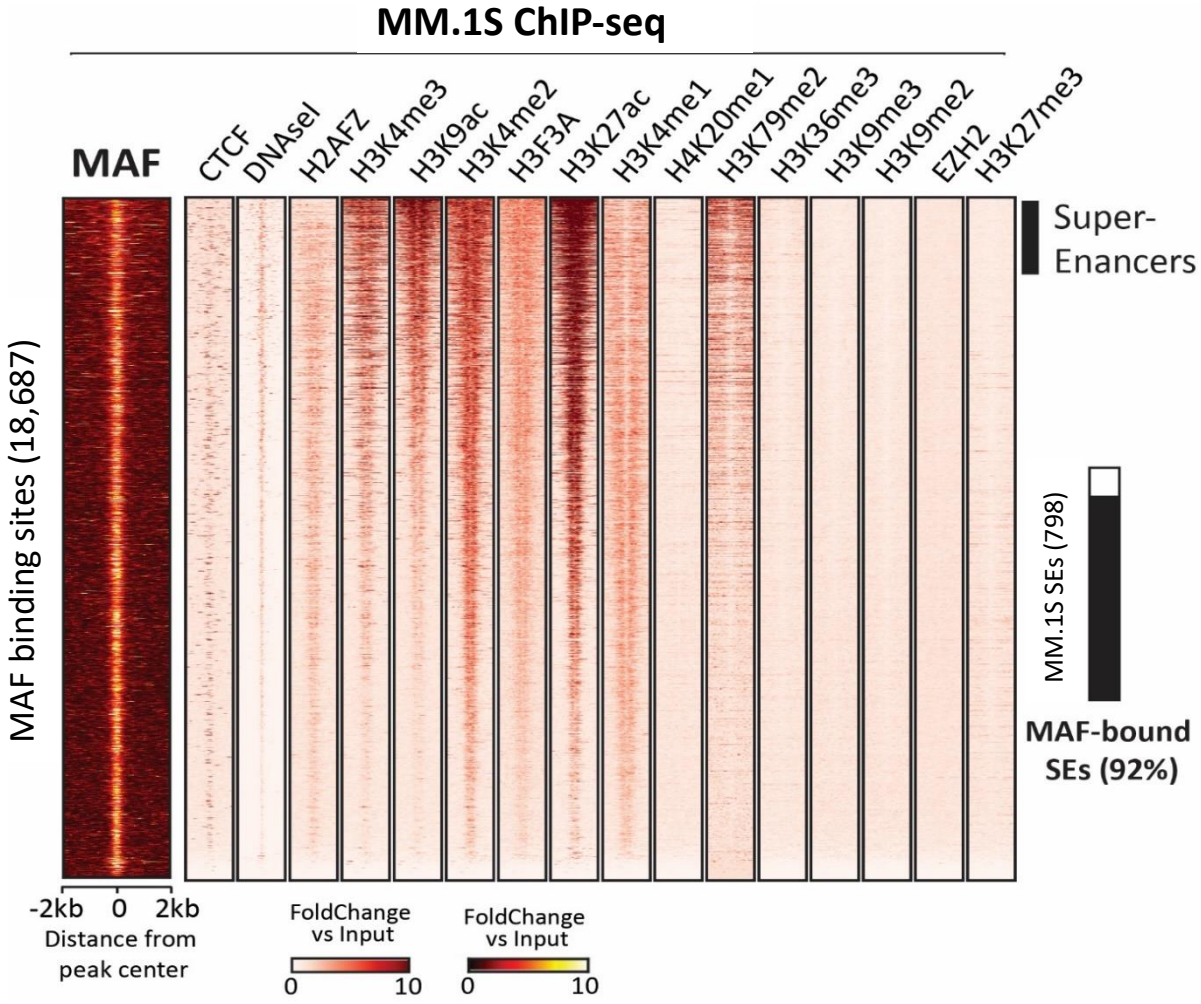


Figure 3.21: Histone mark maps corresponding to MAF binding sites

Heatmap depicting the MAF binding sites (left), as obtained from ChIP-seq assays against MAF in MM.1S, represented as fold change over input. On the right, respective maps of histone marks and other epigenetic markers in MM.1S on the same binding sites (fold change over input), obtained from ENCODE consortium project repository[117]. Superenhancers in the MAF binding sites depicted by the black bars on the right.

In JJN3, a total of 23,669 MAF binding sites were seen, displaying very similar genomic distribution as the one observed in MM.1S (**Fig. 3.22**). Peaks were mostly located in intronic and intergenic regions, with a smaller proportion present on promoters. Similarly, although much fewer peaks were obtained in H929, the genomic location of these strongly favoured intergenic and intronic regions (**Fig. 3.23**).

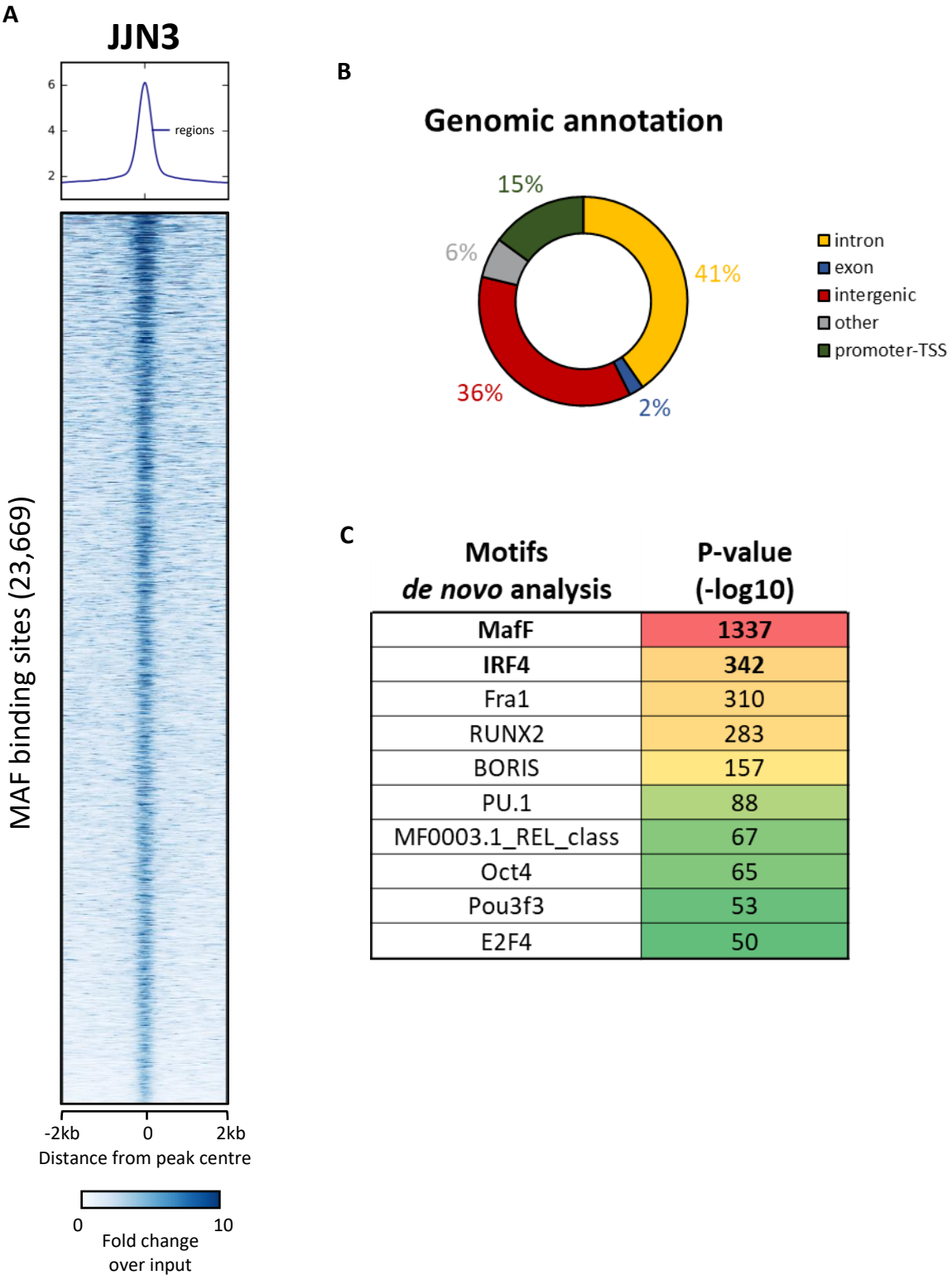


Figure 3.22: MAF ChIP-seq in JJN3

(A) Heatmap of MAF binding sites in JJN3, depicted in fold change over input. **(B)** Genomic annotation of these peaks. **(C)** *De novo* motif analysis of the ChIP-seq peaks.

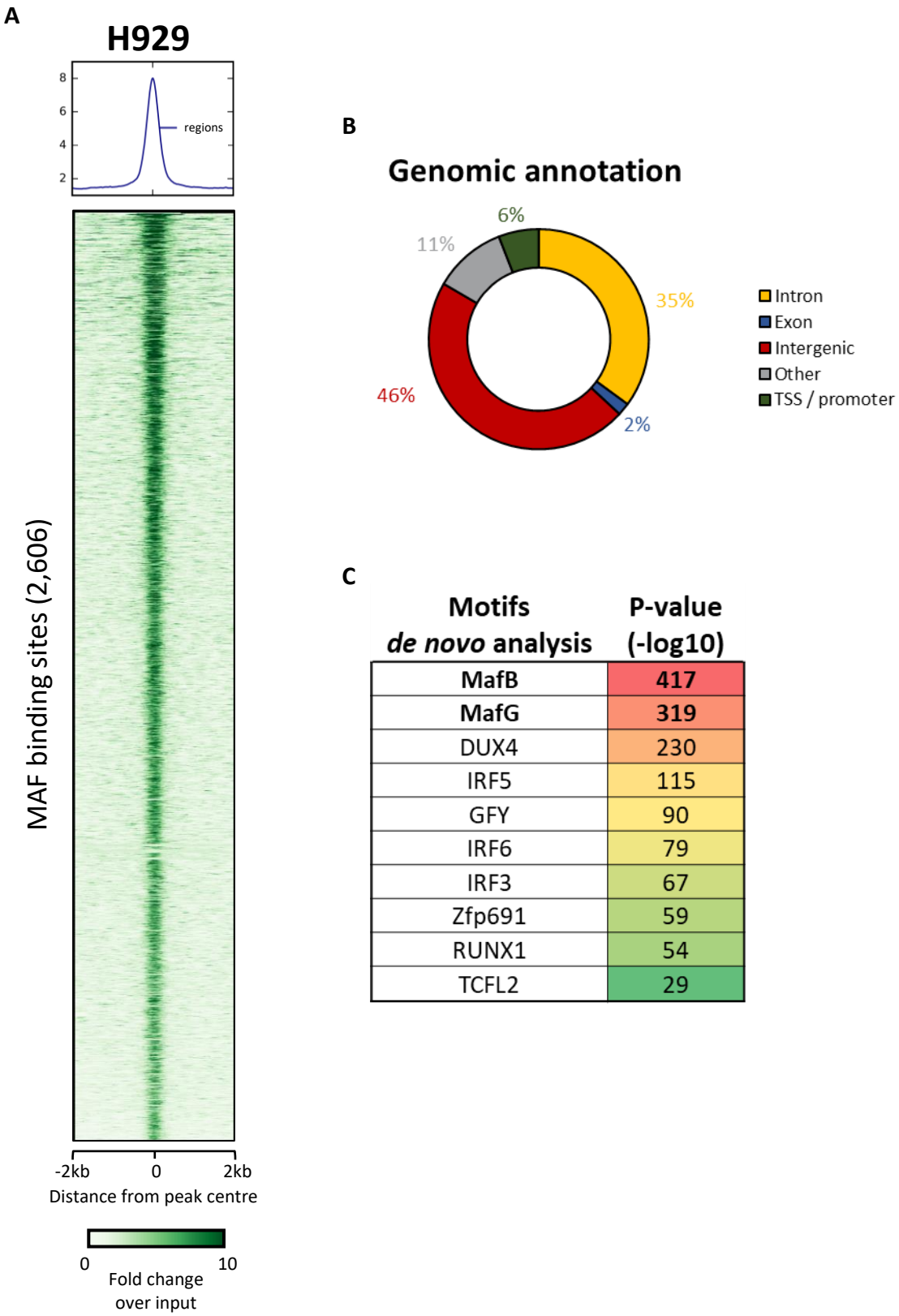


Figure 3.23: MAF ChIP-seq in H929

(A) Heatmap of MAF binding sites in H929 (fold change over input). **(B)** Genomic annotation of these bindings sites and **(C)** *de novo* motif analysis of ChIP-seq peaks.

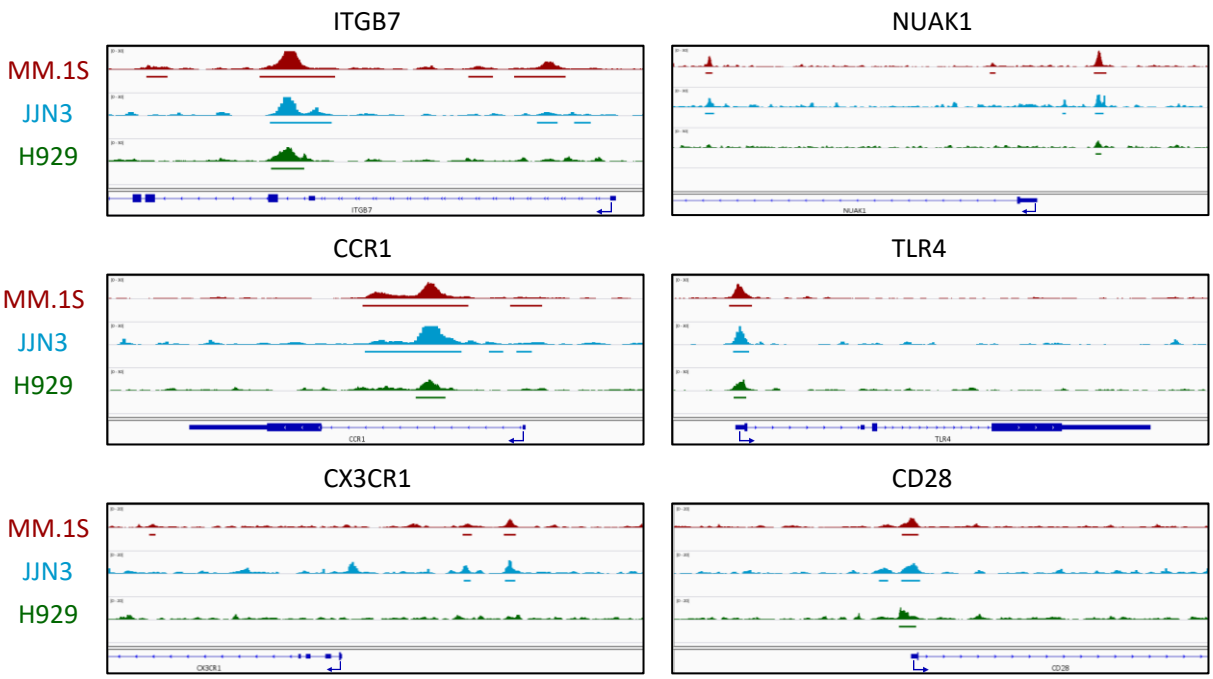


Figure 3.24: MAF binding on target genes

IGV plots depicting examples of MAF binding targets in three myeloma cell lines. Graphs illustrate key known MAF target genes (*ITGB7*, *NUAK1*, *CCR1*), some known MF molecular group associated (*TLR4*, *CX3CR1*) and other genes previously not known to be MAF targets (*CD28*). Peaks represent ChIP-seq fold change over input, bars underlying peaks indicate significant peaks with q value<0.0.1.

Exploring the MAF binding sites indeed revealed direct binding of the transcription factor in key known MAF targets, such as *ITGB7*, *CCR1*, and NUAK1 as well as MAF associated genes, as identified in GEP of myeloma molecular groups, such as for example *TLR4* and *CX3CR1* (**Fig. 3.24**). In addition, new MAF-bound targets were identified, such as in the case of *CD28*. In order to identify the genome-wide direct targets regulated by MAF, the MAF cistrome was then integrated with the differential gene expression as defined by MAF knockdown.

3.6 Identification of the MAF directly regulated targets

3.6.1 Integration of MAF cistrome with differential gene expression upon MAF knockdown

In order to explore which genes are directly regulated by MAF-binding, the ChIP-seq profile in MM.1S cells was integrated with the differential gene expression upon MAF knockdown in the same cell line, using the Beta-plus package[134]. This tool assigns target genes as a function of distance of the binding site from the transcription start site (within 100kb), number of binding sites and gene expression changes. The MM.1S cell line was chosen, instead of JJN3, as shRNA-mediated MAF knockdown in JJN3 produced a very small number of differentially expressed genes. Analysis yielded 2,112 genes predicted to be directly regulated by MAF (**Fig. 3.25A**). The majority of them were upregulated by MAF, with an overall activating function predicted (**Fig. 3.25B & C**). Closer inspection of the MAF-directly regulated genes revealed that known MAF targets, such as *ITGB7*, *CCND2*, *CCR1* and *NUAK1* were amongst the top ranking MAF-activated genes (**Fig 3.26A**). In addition, GEP MF subgroup signature genes were also amongst the top 20% of MAF-activated genes (e.g., *ARID5A*, *RASSF4*, *LCP2*, *CCDC86*, *ITGB5* and *CX3CR1*). Indeed, the predicted direct MAF targets enriched for genes in the top 50 upregulated genes defining the MF molecular group in the GEP study conducted by Zhan et al.[67] (**Fig. 3.26B**). Finally, interesting new MAF target genes were also revealed, such as *CD28*, *FLI1* and plasma cell identity-defining transcription factors *XBP1*, *PRDM1* and *IRF4*.

Pathway enrichment analysis on the MAF-activated genes revealed they are mainly involved in cytokine pathways, cell cycle regulation and the integrin signalling pathway (**Fig. 3.27**), pathways previously hypothesised to be regulated by MAF based on a small number of genes associated with MAF[28]. By contrast and as previously seen with genes upregulated upon MAF knockdown, MAF-repressed genes did not show any significantly enriched pathway using GSEA. Pathways found enriched using the EnrichR online tool[105] are shown in **Appendix C Fig. S11**.

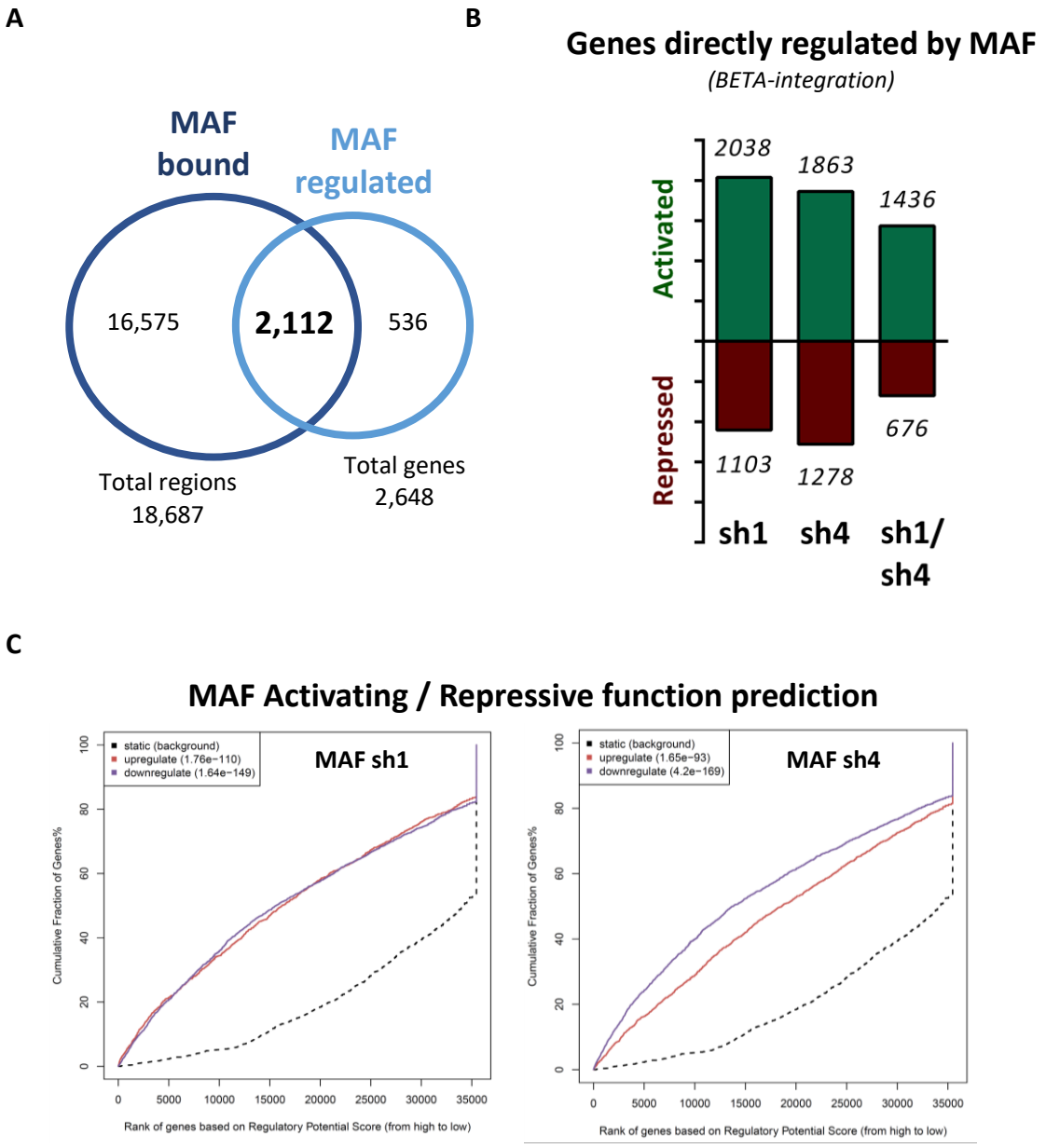


Figure 3.25: Integration of MAF cistrome with transcriptome following MAF depletion

Beta-plus integration of MAF cistrome with differential gene expression upon MAF knockdown in MM.1S cells **(A)** Venn diagram depicting intersection of MAF-bound regions, as identified by ChIP-seq in MM.1S, with MAF-regulated genes, found up- or downregulated upon shRNA-mediated MAF knockdown in the same cell line. **(B)** Number of activated or repressed genes identified per shRNA construct. **(C)** Activating (purple) and repressive (red) function prediction of MAF as identified by BETA-plus analysis.

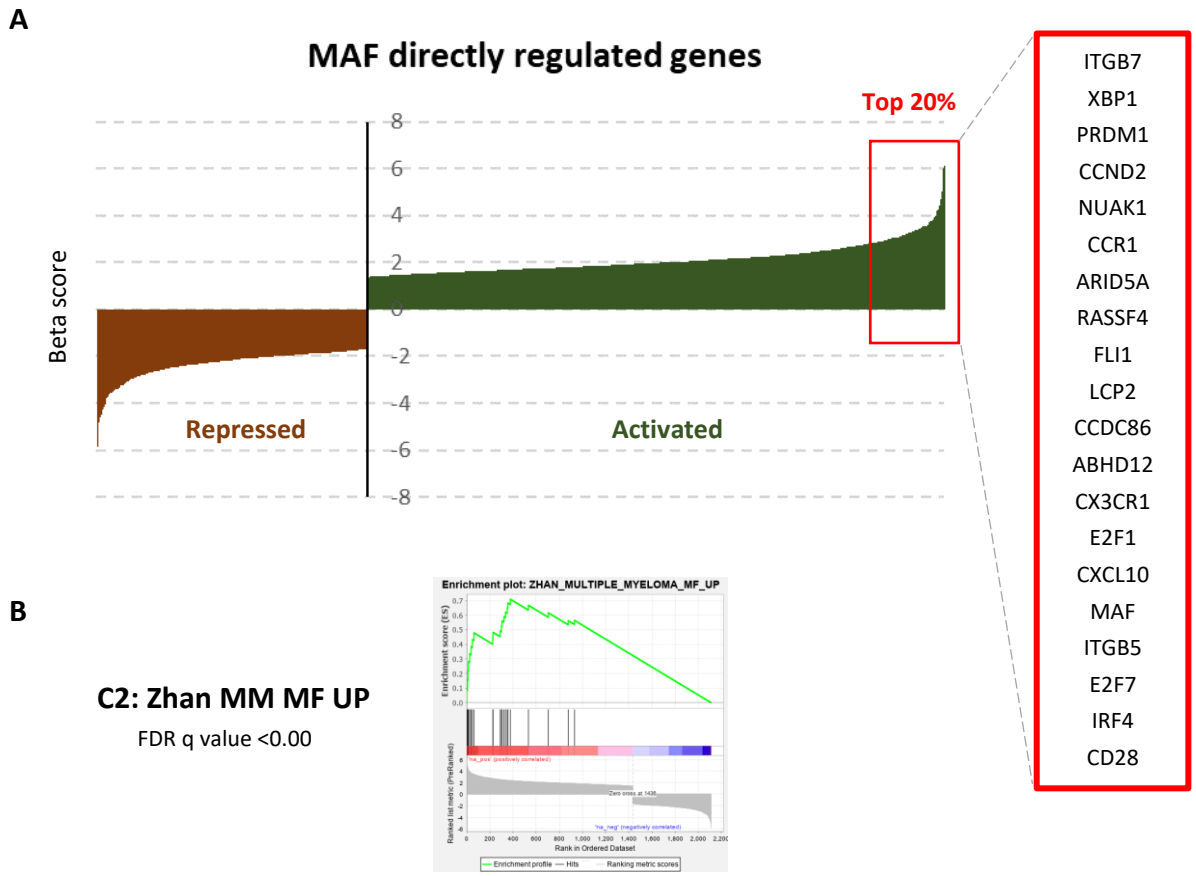


Figure 3.26: MAF directly regulated targets

(A) MAF directly regulated targets as identified by Beta-plus analysis of shRNA up- and downregulated genes, ranked by Beta score. Highlighted are some of the genes in the top 20% of MAF regulated targets, some of which are previously identified MAF targets or MF molecular group associated genes in GEP studies. **(B)** GSEA[104] analysis shows significant enrichment of MAF targets against the top 50 upregulated genes in the MF molecular group identified by Zhan et al[67].

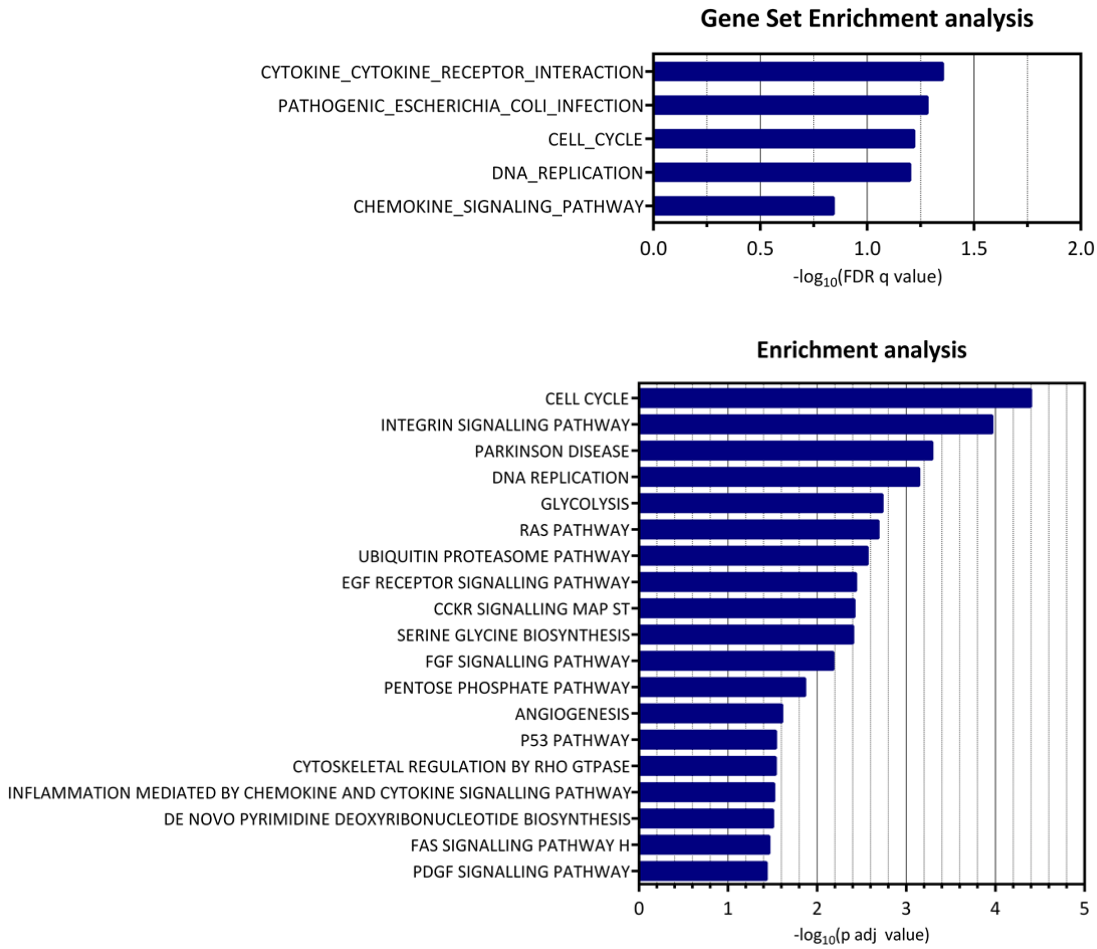


Figure 3.27: Pathway enrichment analysis on MAF activated genes

Pathway enrichment analysis performed on MAF directly activated genes, using the GSEA software[104] (top). EnrichR online tool enriched pathways (Panther pathways)[105] depicted in the bottom panel for pathways achieving a padj value<0.05.

3.6.2 Integration of MAF directly regulated targets with genes differentially expressed in MAF primary samples

Finally, I set out to identify which of the MAF targets validated *in vitro* in myeloma cell lines in the aforementioned experiments are also differentially expressed in primary MAF myeloma PC. For this purpose, I integrated the predicted directly MAF-activated target genes (obtained from the Beta-plus integration of MAF ChIP-seq with the genes downregulated upon MAF knockdown) with the genes differentially upregulated in the MAF primary samples as compared to normal PC. This process identified 102 genes, excluding MAF itself, forming a group of MAF core target genes (**Appendix C Table S3**).

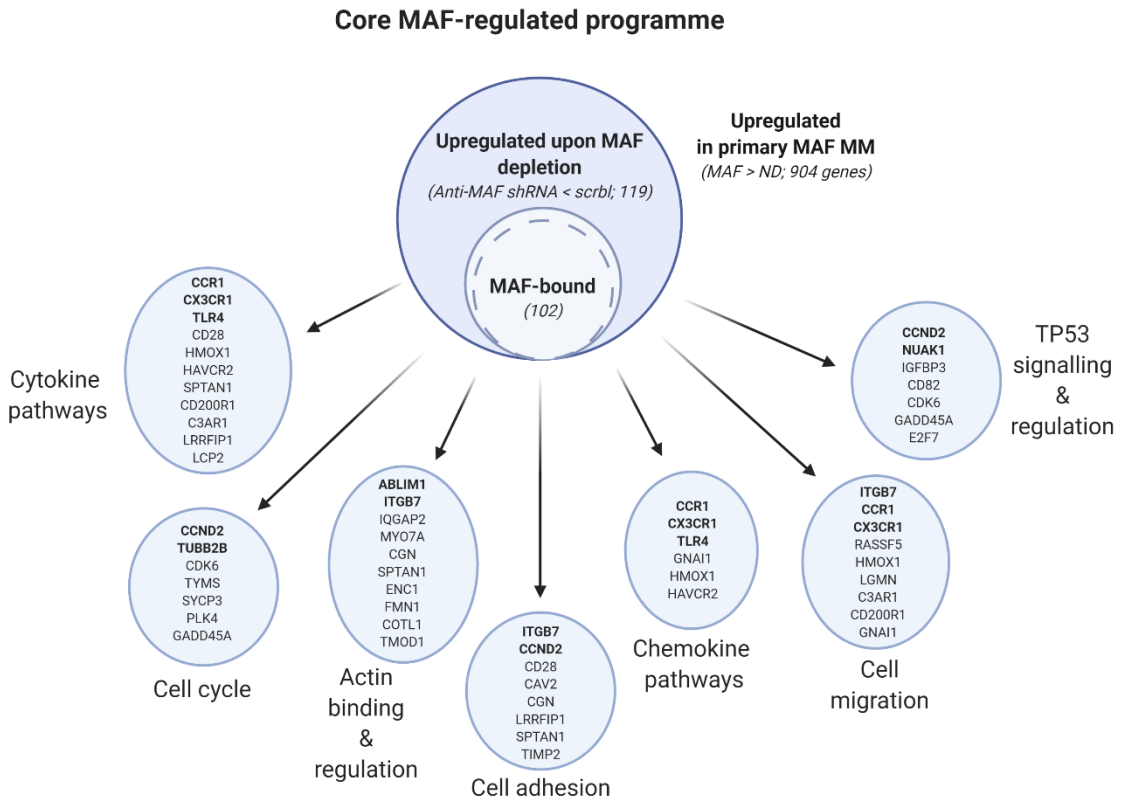


Figure 3.28: MAF core gene targets

Integration of predicted MAF direct targets obtained from MMCL *in vitro* assays (MAF bound and downregulated upon MAF depletion) with genes found differentially expressed in MAF primary samples (genes upregulated in MAF samples compared to normal PC) yielded a set of 102 core MAF-regulated genes. Functional annotation of the genes was performed manually using multiple GSEA curated datasets[104]. Only pathways involving 4 or more MAF core target genes are shown. In bold, genes previously associated with the MF molecular group in GEP studies[67, 68]. ND; normal donor plasma cells.

Due to the small number of genes, GSEA and pathway enrichment analysis using the EnrichR online tool did not yield any results at q value <0.05. In order to therefore ascertain the pathways regulated by the MAF core target genes and determine a core programme of cellular functions influenced by the transcription factor, I manually annotated the genes, using multiple curated GSEA datasets[104], focusing only on pathways involving 4 or more core MAF target genes (**Fig.3.28**). This revealed key cellular functions that are regulated by ectopic MAF expression in myeloma PC, including in addition to cell cycle, pathways involved in cell to cell interactions, such as cell adhesion, cell migration and cytokine and chemokine signalling pathways.

3.7 Discussion

In the set of experiments described in this section I aimed to define the transcriptome changes imposed by the aberrant expression of MAF in plasma cells, in the context of t(14;16) myeloma and used this to predict the impact of the translocation on cellular function.

A small number of MAF gene targets have been described before. *CCND2, CCR1*, *NUAK1* and *ITGB7* have previously been identified as MAF-regulated genes using MAF depletion and overexpression assays[45, 85]. In addition, *NUAK1* and *CCND2* were hypothesised to be direct MAF targets following the detection of the MARE sequence at the gene promoters and the ability of MAF to transactivate them[45, 85]. In addition, GEP studies conducted on MM patient PC have identified a MF molecular group with a distinct transcriptome profile when compared to other MM genetic groups[67, 68, 88]. Although these genes have been found to be MAF-associated, they have not been validated as MAF-regulated targets.

Despite the above available data, the full MAF transcriptome profile in MM, as this relates functionally to the MAF transcription factor and the genome-wide direct MAF targets in MM have not been identified before. In addition, the MAF-associated transcriptome characteristic of t(14;16) myeloma PC has never been compared to that of normal bone marrow PC enabling, thus, identification of the changes occurring upon acquisition of the t(14;16) translocation setting the cell *en route* to myelomagenesis. Therefore, in this section I set out to address some of these outstanding questions. An illustration of the workflow and analysis followed can be seen in **Fig. 3.29**.

3.7.1 MAF transcriptome profile in primary t(14;16) MM

I first compared the levels of MAF expression in primary cells and MMCLs, showing that MAF-translocated primary cells have similar if not higher expression than the translocated MMCLs. In addition, some of the other primary samples, expressed MAF at a much lower level, comparable to that of MMSET-translocated MMCLs. Although the number of samples tested

here is relatively small to ascribe precise percentages (23 non-t(14;16) samples), this has been described before[45, 61], at a rate varying between 30-50%. MAF has previously been seen aberrantly expressed in a proportion MMSET-translocated MM and a very small number of other myelomas[62]; this is confirmed in the samples assayed here where a low level MAF expression was detected in at least half of the MMSET samples but also some of the hyperdiploid MM cases. Furthermore, I have confirmed that normal bone marrow PC indeed do not express MAF, as has been reported in studies of normal marrow and tonsillar PC[45, 46]. This rendered them ideal to serve as a negative control molecular background, as opposed to PC residing in other tissues, usually tonsillar, or PC differentiated *in vitro* from B cells, commonly serving as negative control cells in MM studies.

I then compared the transcriptome profile of two t(14;16) primary MM samples to that of normal bone marrow PC. Although I was only able to obtain two samples, given the rarity of t(14;16) MM, I attempted to confirm that their transcriptome was representative of the MAF MM group as a whole by comparing their differential gene expression to the MF molecular profile identified in GEP studies involving large groups of patients. Indeed, my dataset obtained from the two patients was significantly enriched in genes identified as characteristic of the MF molecular group. Importantly however, the GEP-identified MF molecular profile was obtained from comparing different MM samples as opposed to the transcriptomes of MM vs normal PC. Furthermore, differential gene expression of 30 MM samples of different genetic backgrounds (including my two MAF samples) obtained as part of a larger collaborative project[100], was compared to the transcriptomes of 892 MM samples available from the MMRF CoMMpass study[121]. Indeed, the two t(14;16) samples used in my dataset clustered well with the MMRF MAF samples, further suggesting they are representative of the MAF MM. However, the MMRF dataset is derived from MM transcriptomes as opposed to differential expression analysis between MM and normal PC.

The transcriptome profile of my t(14;16) samples, in addition to significantly enriching in genes identified as part of the MF molecular signature, also enriched in genes identified as upregulated in angioimmunoblastic T cell lymphoma (AITCL), as compared to normal CD4+ T cells[131]. In this study, authors compared the gene expression profiling of primary AITCL with that of normal CD4+ T cells, and found the gene expression of the lymphoma cells most

resembles that of T follicular helper (Tfh) cells, a subset of CD4+ T cells[131], from which AITCL is thought to arise [135]. Indeed, MAF is expressed in Tfh cells, regulating cytokine production and their differentiation from early CD4+ cells[49]. Forty-five genes were found to be shared between the transcriptome of MAF myeloma and the upregulated genes in AITCL (which identified just over 200 differentially expressed genes), including *TLR4* and *RND3*, also identified as part of the MF group molecular signature. However, in addition to MAF-driven upregulation, this dataset may also include genes reflecting the malignant transformation of AITCL cells (as MAF is expressed in Tfh cells and indeed other CD4+ subsets) as well as the genes differentially expressed between Tfh cells and other CD4+ cells.

I then compared the transcriptome obtained after comparison with normal PC in the MAF samples, with that obtained from MM of other genetic backgrounds. Indeed, the differential transcriptome clustered the MAF samples away from other genetic MM subgroups, suggesting they have a distinct molecular profile. Closer inspection revealed that 38% of MAF upregulated genes were uniquely upregulated in the MAF samples (such as *CX3CR1*, and *LCP2*) whereas the rest were shared with other myeloma subgroups, most commonly the MMSET group. Indeed, many of the MF molecular group genes identified in GEP studies[67, 68, 88], or indeed previously identified MAF targets, fell into that category, including *NUAK1*, *TLR4*, *CCND2* and *ITGB7*. It is therefore reasonable to conclude that genes uniquely upregulated in MAF cells as well as genes shared with other MM genetic subgroups form part of MAF MM biology, as the same genes could be upregulated via different mechanisms in different cytogenetic subgroups. In addition, in the case of genes shared with MMSET-translocated cells, MAF could be responsible for their upregulation in both MM genetic subgroups.

3.7.2 Effects of MAF depletion

Having confirmed that t(14;16) MMCLs and primary cells express similar levels of MAF, I used the t(14;16) MMCLs MM.1S and JJN3 for further experimentation. shRNA-mediated MAF knockdown led to cell death, confirming my hypothesis that MAF is critical for myeloma cell survival in these cell lines. At the time of these experiments this had not been shown before, although a recent abstract published in the American Society of Haematology conference at

the end of 2019, reported MAF to be a myeloma cell dependency in CRISPR knock-out assays[136]. These data however have not yet been published in a peer-reviewed journal. In addition, MAF knockdown induced cell cycle arrest, consistent with previous reports of a reduction in cell proliferation upon MAF inhibition[45].

Comparing gene expression of MAF primary samples to that of normal PC yielded interesting results and linked genes found to be differentially expressed with the presence of MAF-translocation, but not in a causal manner. I therefore used the MAF shRNA knockdown system to compare gene expression between two anti-MAF shRNAs and a scrbl control and thus established the genome-wide functional targets of MAF. Comparative analysis revealed many more genes were downregulated by MAF knockdown than upregulated, suggesting MAF acts primarily as a gene activator in this context. Downregulated genes were found to be enriched in genes identified as characteristic of the MF molecular profile, increasing my confidence in the specificity of these results. By contrast, genes found to be upregulated by MAF knockdown appeared less specific and didn't reflect genes in these datasets. This could mean that upregulation of genes was non-specific, or they form part of secondary, downstream effects. I therefore focused only on the genes downregulated upon MAF knockdown. In addition, there was a significant difference in the number of genes found differentially expressed between the two MAF-translocated MMCLs upon MAF depletion, likely explained by the early timepoint of RNA collection in JJN3 cells, much before effects on cell viability were evident. In future experiments, harvest of JJN3 cells following knockdown should be performed at a later timepoint. This would enable better integration of the RNA-seq results in the two cell lines, increasing the confidence in the results. Despite these limitations, these experiments confirmed some of the MF molecular group associated genes identified in GEP studies as being functional targets of MAF, such as *CX3CR1*, *TLR4* and *RND3*. In addition, new putative MAF targets were also identified such as *CD28*, *IRF4* and *FLI1*.

3.7.3 Defining the MAF cistrome

In order to further refine these results, I then identified the cistrome of MAF by performing ChIP-seq experiments on MAF-translocated and expressing cell lines. Interestingly, based on genomic annotation, the majority of MAF binding sites were found in intronic and intergenic regions in both MM.1S and JJN3 cells, suggesting MAF preferentially binds enhancers. This has previously been observed in ChIP-seq against MAF performed in activated Th17 cells: following activation of Th17 cells with anti-CD3/CD28 antibodies that resulted in MAF upregulation, ChIP-seq against MAF revealed that over 80% of MAF binding sites were found outside a gene promoter[123]. The authors then used ChIP-qPCR to show that the majority of subsequently tested regions were enriched in H3K27ac and H3K4me1 instead of H3K4me3, thus identifying these regions as enhancers. Motif analysis in the same ChIP-seq experiments revealed in addition to the MAF motif, the motifs of AP-1, ETS, RUNX and TCF also highly enriched, leading the authors to hypothesise that some of these transcription factors may cooperate with MAF in Th17 cells[123]. These motifs were also identified in the analysis of ChIP-seq in MM.1S and JJN3, though mostly these were seen in one or the other cell line. Interestingly though, the two motifs found consistently between the two t(14;16) cell lines were that of RUNX and IRF4, suggesting that these transcription factors could be acting co-operatively with MAF in myeloma cells.

ChIP-seq against MAF was also performed in H929, an MMSET-translocated MMCL, yielding similar results with respect to genomic annotation of MAF binding sites, although the number of binding sites was much lower. This led me to the hypothesis that the level of MAF protein expression may influence the function and genomic binding of MAF. However, only one replicate of ChIP-seq against MAF in H929 was performed, and further experiments would be required to confirm these findings. In addition, in order to further investigate how the function of MAF may differ in this genetic context, knockdown experiments and a comparative gene expression analysis would need to be performed.

3.7.4 Predicting the direct gene targets of MAF

In order to predict targets directly regulated by MAF, i.e., genes that are both bound by MAF and differentially expressed upon knockdown (mostly downregulated), I integrated the ChIP-seq data with the differential gene expression following shRNA-mediated MAF knockdown in the MM.1S cell line. Unfortunately, only one of the two t(14;16) cell lines could be used, as in JJN3 very few downregulated genes were identified.

Some of the genes found to be associated with the MF molecular group in GEP studies[67, 68, 88] but not known to be regulated by MAF were indeed predicted by this integration to be directly upregulated by MAF, such as *CX3CR1*, *TLR4* and *LCP2*. CX3CL1, the ligand of the chemokine receptor CX3CR1, has been shown to increase myeloma cell adhesion to fibronectin and VCAM-1 expressed on bone marrow stromal cells, a process thought to upregulate pathways promoting myeloma cell survival[137]. In addition, CX3CL1 has been reported to promote bone marrow microvessel density in MM and angiogenesis[138]. TLR4 is a member of the Toll-Like Receptor family, the ligands of which have previously been shown to stimulate myeloma cell growth and survival through autocrine action by activation of the NFκB pathway and induction of IL-6 secretion[139-141].

In addition, new direct predicted targets of MAF, not previously described, were identified through my analysis, such as *FLI1* and *CD28*, as well as plasma cell identity defining transcription factors *XBP1*, *PRDM1* and *IRF4*. CD28 is a surface glycoprotein whose expression has previously been associated with poor prognosis in MM and t(14;16)[142]. CD28 has been shown to mediate myeloma cell interaction with bone marrow dendritic cells via its ligands CD80/CD86, which in turn induces IL-6 production and promotes myeloma cell resistance to starvation and chemotherapeutic agents[143]. FLI1, an ETS transcription factor, has previously been shown to be critical for myeloma cell survival and is a key transcription factor in the myeloma regulatory network, by binding to MM superenhancers[144]. On the other hand, XBP1, PRDM1 and IRF4 are transcription factors involved in PC differentiation[15], and it is tempting to hypothesise that dysregulation of their expression induced by aberrant expression of MAF in the developing PC may have a role in MAF-induced myelomagenesis. Indeed, PRDM1 and XBP1 were both found upregulated in the tumour cells of transgenic mice

that expressed MAF in the B-cell compartment and subsequently developed a lymphoma with myeloma-like features, compared to wild type B cells[57]. IRF4 was also found elevated in these cells, but did not reach statistical significance. Further experiments would be required to investigate this, perhaps by molecular studies following overexpression of MAF in a normal B cell or PC.

3.7.5 The MAF core transcriptional programme

In order to reflect these results obtained from experiments in MMCLs back to the transcriptome profile of primary t(14;16) samples and ensure the *in vitro* results are relevant to the molecular profile of real-life patients, I finally integrated the MAF directly upregulated target genes with those found upregulated in the primary samples as compared to normal PC. Eleven per cent, i.e., 102, of the differentially expressed genes in primary samples were predicted to be direct MAF targets, thus constituting a core MAF programme that is active in primary MAF-translocated MM. Functional annotation of these 102 genes identified pathways previously hypothesised to be regulated by MAF based on a handful of identified target genes, such as regulation of cell cycle, cell adhesion and cytokine pathways[28, 45]. Overall, pathway enrichment analysis of gene sets obtained either following MAF knockdown, MAF ChIP-seq or after their integration with data from primary samples, consistently confirm that in addition to cell cycle regulation, MAF overexpression mostly affects pathways involved in cell to cell interaction. Genes activated by MAF are involved in cell adhesion, cell migration and the regulation of cytokine and chemokine production and signalling pathways. This suggests therefore that overall aberrant expression of MAF in PC affects primarily their interaction with each other and their surrounding bone marrow microenvironment.

A relatively small number of genes have been identified as being part of this MAF core transcriptional program, in comparison to the number of genes found upregulated in MAF primary samples. This is not surprising; comparative analysis between the MAF samples and normal PC would have picked up many genes upregulated in MM in general, that are not necessarily functionally related to MAF. Other genes may be driven by MAF overexpression, but may not be directly bound by the transcription factor, and instead represent downstream

events. Finally, although I confirmed that the molecular profile of the two MAF samples is representative of larger MAF MM cohorts, the small sample size in this analysis may have compounded the results. As a result, the core MAF programme identified here may have been unable to capture the molecular profile of primary MAF MM in its entirety. In order therefore to confirm these results and further improve on this analysis, a larger number of primary MAF samples would be needed. In addition, future experiments would include ChIP-seq against MAF in primary t(14;16) cells, provided a large enough cell number could be obtained.

In summary, genes found to be regulated by MAF using shRNA-mediated knockdown, were integrated with the MAF cistrome, in order to predict directly MAF-regulated targets. Genes found to be differentially expressed upon MAF knockdown, but not MAF bound, are likely to form part of secondary, downstream pathways. Genes directly regulated by MAF were then analysed against the differentially expressed genes found in primary t(14;16) cells. This identified a MAF core programme, that is activated in primary MM. Genes found to be directly regulated by MAF but not differentially expressed in primary cells could either be MMCL specific or alternatively could include genes that may be MAF targets but are already significantly expressed in the normal plasma cell, such as *IRF4*, *XBP1* and *PRDM1*. Finally, the MAF core programme includes genes that are both uniquely upregulated in the MAF subgroup, but also genes found upregulated in other subgroups, with *CCND2* being a notable example.

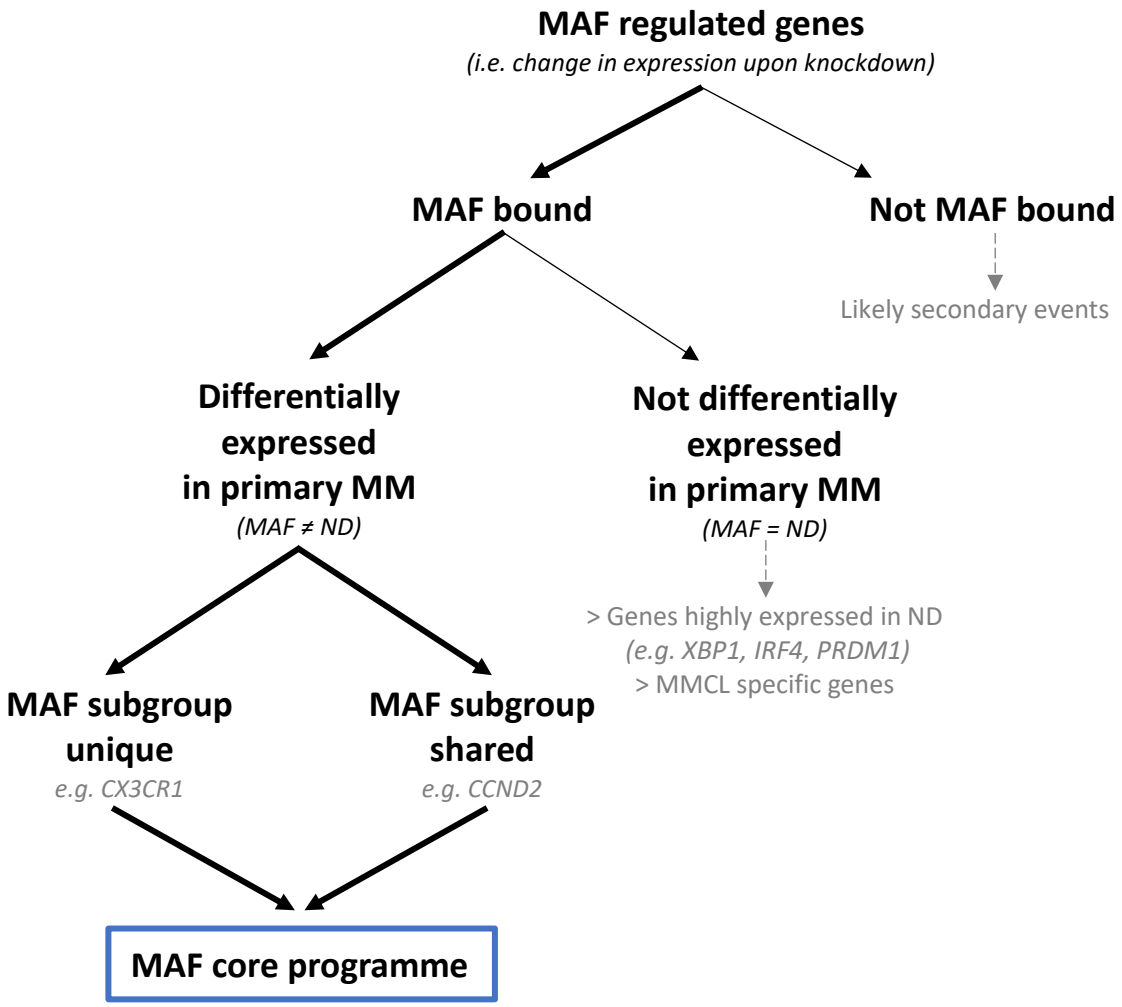


Figure 3.29: Diagram of analysis workflow followed to identify MAF core regulated genes

Diagram outlining the analysis followed in this chapter. ND; normal donor plasma cells, MMCL; multiple myeloma cell line

Chapter 4

MAF associated regulatory genome

4.1 Introduction

Transcription factors act to regulate gene expression by accessing genomic non-coding regulatory regions, such as promoters and enhancers. Promoters, located at the 5' UTR region of a gene, mark the location where transcription starts. Enhancers contain binding sites for multiple transcription factors, are often located up to tens or hundreds of kilobases away from their target genes and activate transcription via long distance interactions with promoters[145].

Cancers are characterised by epigenetic aberrations, especially at enhancers[146]. Several studies have charted the cancer enhanceome, i.e. the full compendium of distal regulatory elements, including in solid organ cancers[147] and haematological malignancies such as chronic lymphocytic leukaemia (CLL)[148, 149] and multiple myeloma[144]. In addition, large scale international projects such as the Encyclopaedia of DNA Elements (ENCODE) Consortium Project[117] and Blueprint Project[115], set up to characterise the human epigenome as a whole or specifically to the haematopoietic system respectively, are an invaluable resource for studies of regulatory genomics.

Mapping of regulatory elements has shown specificity in clustering cancer subtypes[147, 149] while cancer-specific regulatory regions, such as the superenhancers of *BCL2* and *CTLA4* in CLL, can be defined with reference to the normal cellular counterpart[148]. Furthermore, correlating the enhanceome with transcriptome analysis can help predict functional interactions of oncogenic enhancers with their target genes. Finally, identifying enhancers and predicting interactions between regulatory elements and gene expression has enabled modelling of transcription factor regulatory networks and homing in to genes functioning as oncogenic drivers , such as for example the transcription factor MYC in MM[144].

Having observed that MAF binds predominantly outside transcription start sites, in this section I set out to describe the changes in chromatin accessibility, a measure of regulatory potential, associated with oncogenic MAF in the context of multiple myeloma. For this purpose, I used the assay of transposase-accessible chromatin coupled with next generation

sequencing (ATAC-seq). ATAC employs a modified, hyperactive transposase enzyme which cleaves DNA and inserts sequencing adapters in chromatin accessible genomic areas. This enables the mapping of genome-wide chromatin accessibility and identification of regulatory elements. In order to delineate the epigenome associated with the MAF transcription factor, I integrated the chromatin accessibility profile of MAF-translocated myeloma cells with the MAF cistrome and histone mark maps.

4.2 Chromatin accessibility in primary MAF-myeloma plasma cells

To investigate the chromatin changes associated with the MAF translocation, I used ATAC-seq to define the chromatin accessibility of CD138+ selected bone marrow myeloma plasma cells (PC) from two patients with t(14;16) and compared it to chromatin accessibility of normal bone marrow PC (**Fig. 4.1A**). The computational pipeline is outlined in section 2.11.

In total, 6,641 differentially accessible regions were identified. Compared to normal PC, MAF-translocated myeloma displayed enhanced accessibility, with 88% (5,828/6,641) of differentially accessible regions being more accessible in myeloma than normal PC. In addition, similarly to the MAF cistrome, genomic annotation of these regions revealed the majority are intronic or intergenic, suggesting most of the t(14;16)-associated enhanced chromatin accessibility is derived from more accessible enhancer areas as opposed to promoters (**Fig. 4.1A**).

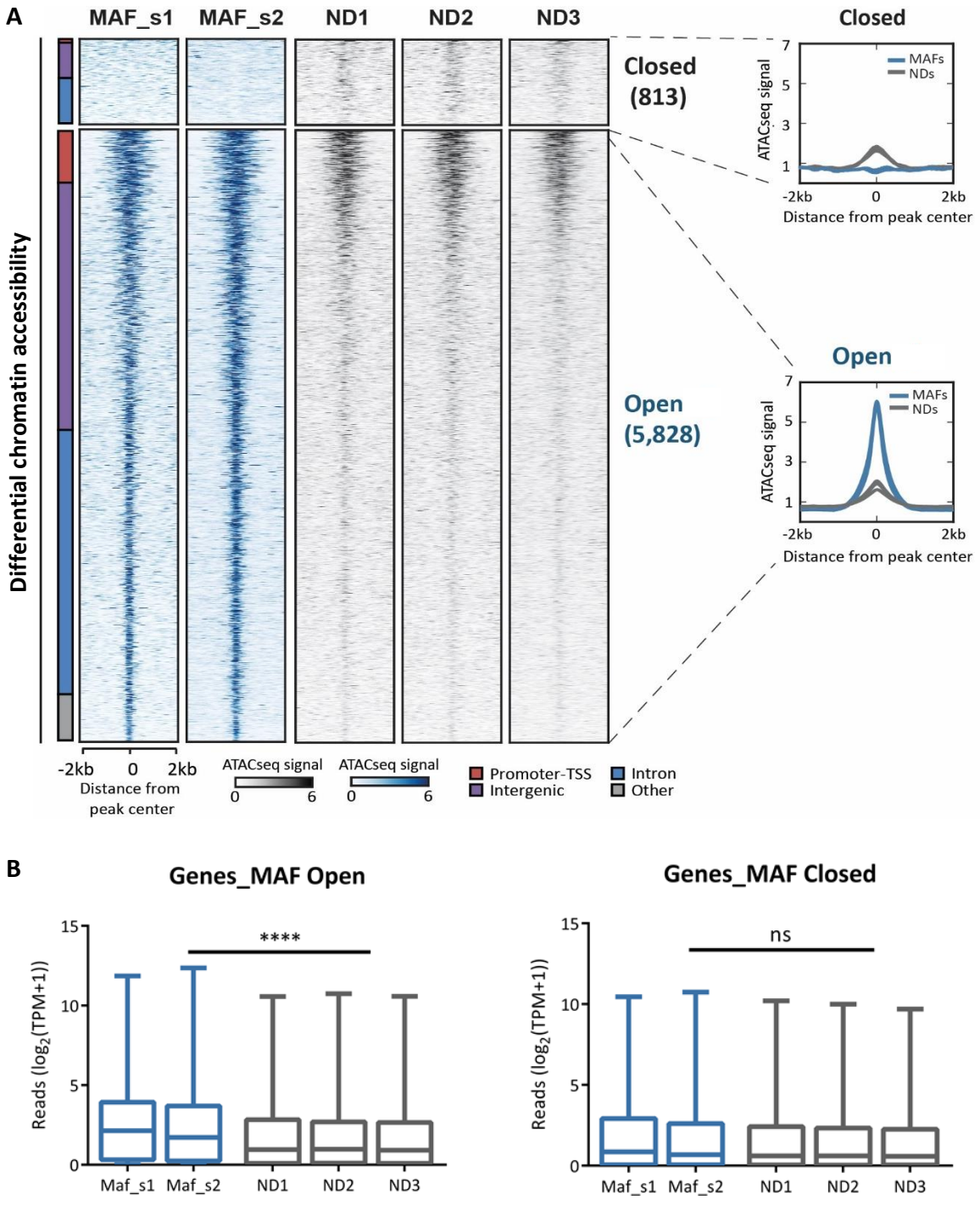


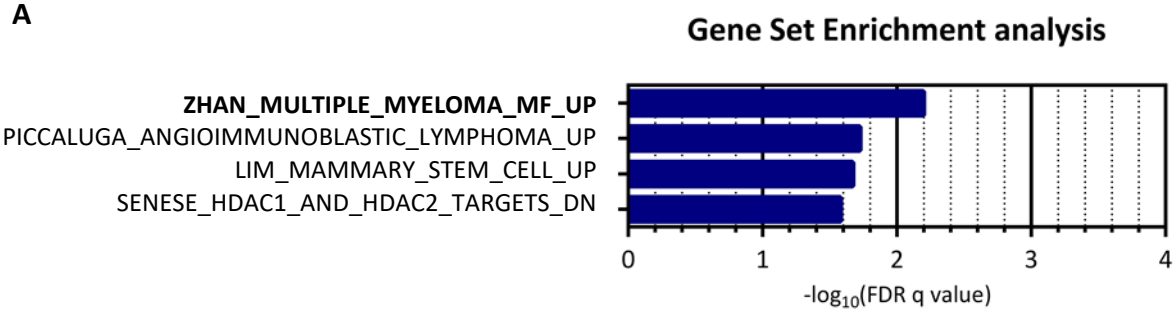
Figure 4.1: Chromatin accessibility in MAF primary samples

Differential chromatin accessibility (ATAC-seq) in MAF primary samples (MAF_s1, MAF_s2) as compared to normal PC (ND1, ND2, ND3). **(A)** Heatmaps depicting regions with differential chromatin accessibility between MAF samples and normal PC (padj<0.05, log_2FC=1); either displaying enhanced accessibility in MAF samples (Open) or reduced accessibility in MAF samples (Closed). Bar on the left illustrating genomic annotation of these regions. **(B)** Box plots depicting expression (in reads) of genes assigned to MAF open (left) or MAF closed (right) regions per sample, **** p<0.0001, ns; non-significant, non-parametric Mann-Whitney U test.

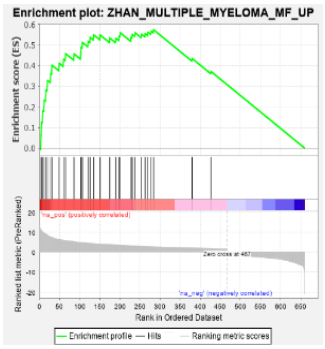
Annotation of the differentially accessible peaks to the nearest gene revealed that, overall, peaks displaying enhanced accessibility in MAF samples were associated with increased gene expression ($p<0.0001$), whereas peaks with reduced accessibility in MAF samples were not associated with a significant change in gene expression ($p=0.0531$; **Fig. 4.1B**), compared to normal PC. This suggests that peaks with enhanced accessibility drive the oncogenic transcriptome in MAF samples.

Gene Set Enrichment analysis (GSEA[104]) of differentially expressed genes annotated to differentially accessible peaks (within 1Mb), revealed enrichment in genes characterising the MF molecular group as identified by Zhan et al.[67], with hits including *TLR4*, *CX3CR1*, *NUAK1* and *ITGB7*, i.e., genes characteristically upregulated in MAF MM (**Fig. 4.2A**). The second most enriched gene set was that of genes upregulated in angioimmunoblastic T cell lymphoma, a malignancy also associated with *MAF* overexpression[56], reflecting the lymphoma's CD4+ follicular T helper cell origin[150]. Furthermore, some of the top differentially expressed genes (log_2FC expression MAF vs ND) annotated to peaks with enhanced accessibility in MAF samples included other known MAF targets such as *CCND2* and *CCR1*, as well as *MAF* itself. Interestingly, the downregulated genes characterising the MF group were not enriched for in my dataset, perhaps because GEP studies defined MF group genes based on comparison to other MM subgroups rather than normal PC. Moreover, downregulated genes annotated to differentially accessible MAF regions did not yield any significant enrichment upon GSEA. This further supports the notion that over-accessible regions mainly drive transcriptional activation in MAF myeloma PC. To confirm GSEA findings, pathway enrichment analysis was also performed using the online tool GREAT[112]. Indeed, the top most significantly enriched pathway associated with peaks displaying enhanced accessibility in MAF samples was that of genes overexpressed in the MF group[67] (**Fig. 4.2B**). Here, genes associated with MAF peaks with reduced accessibility were also significantly enriched for genes in the top 50 downregulated MF group genes, though this was not the top most enriched gene set. This discrepancy with the GSEA enrichment analysis might be explained by the method of gene annotation to each peak: GSEA was performed on a gene set based on differential gene expression between the MAF MM and normal PC as well as distance from the peak, whereas the online tool GREAT, assigns genes to peaks based solely on proximity.

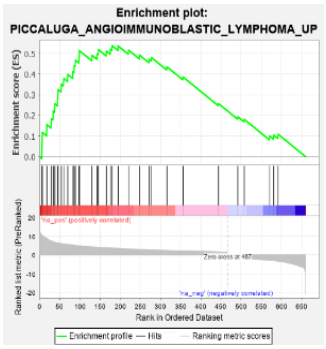
A



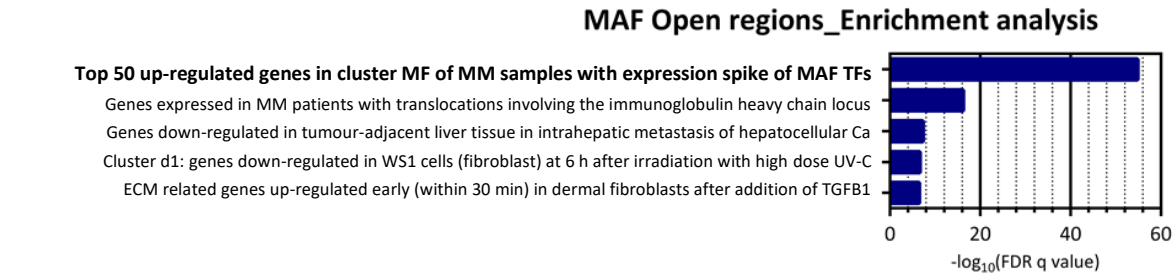
Zhan, myeloma MF up



Piccaluga, Angioimmunoblastic lymphoma up



B



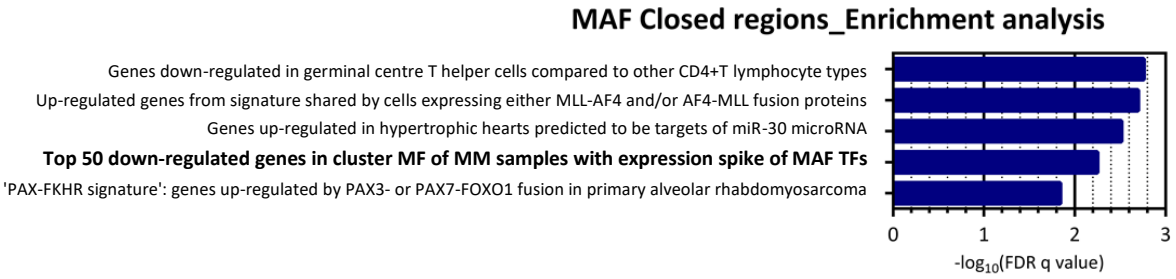


Figure 4.2: Enrichment analysis of MAF differentially accessible regions

(A) Enrichment analysis of genes assigned to MAF differentially accessible regions compared to normal PC, against gene sets using the GSEA software[104] (FDR q<0.05). Enrichment analysis plots of the two top enriched gene sets depicted below. **(B)** Enrichment analysis of differentially accessible regions (accessibility enhanced in MAF, top, accessibility reduced in MAF, bottom) using the GREAT online tool[112]. Top 5 significantly enriched genes sets shown, based on FDR q value. Genes were assigned based on GREAT default settings.

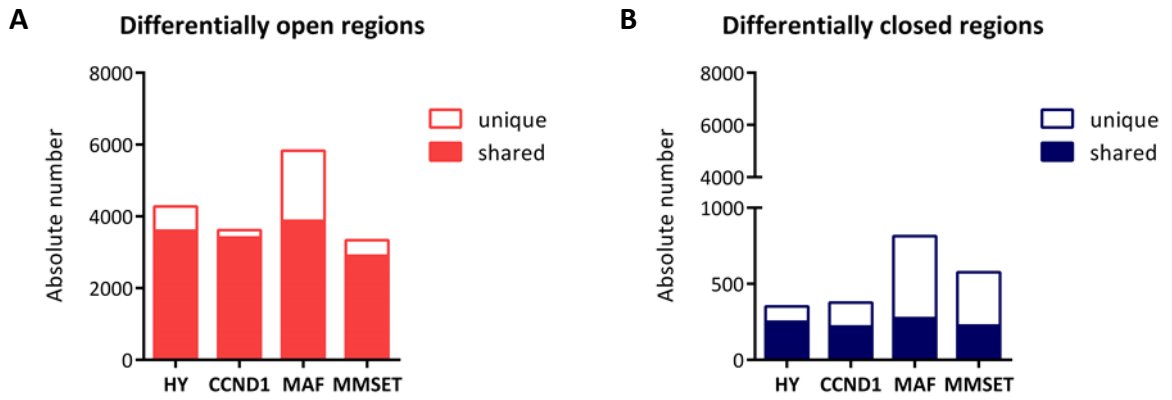


Figure 4.3: Comparison of differentially accessible regions across myeloma genetic subgroups

(A) Number of regions with differentially enhanced accessibility and **(B)** number of regions with differentially reduced accessibility in the different MM genetic subgroups as compared to normal PC, identifying those that are uniquely differentially accessible in each subgroup.

Using ATAC-seq data from primary MM samples myself and my colleagues obtained as part of a larger collaborative project[100], I was able to compare the accessibility profile of the MAF samples with that seen in CCND1, MMSET and hyperdiploid MM genetic subgroups. All myeloma PC, irrespective of genetic background, displayed overall enhanced accessibility compared to normal PC, however, the MAF subgroup had the highest number of differentially open regions (**Fig. 4.3A**). These peaks could be shared between two or more subgroups or be unique to each genetic group; however, the MAF samples displayed the highest number of uniquely differentially accessible regions (33% in MAF, 16% in hyperdiploid, 13% in MMSET and 6% in CCND1, **Fig. 4.3A**). The same was true for regions with reduced accessibility as compared to normal PC (66% unique in MAF, 41% unique in CCND1, 36% in MMSET and 29% in hyperdiploid myeloma samples), although the total number of regions with reduced accessibility was much smaller than the number of peaks with enhanced accessibility in any myeloma subgroup compared to normal PC (**Fig. 4.3B**). Hierarchical clustering analysis of the differentially accessible peaks in the myeloma samples, revealed that MAF samples indeed clustered distinctly together (**Fig. 4.4**), indicating a significantly different accessibility profile compared to other myeloma subgroups.

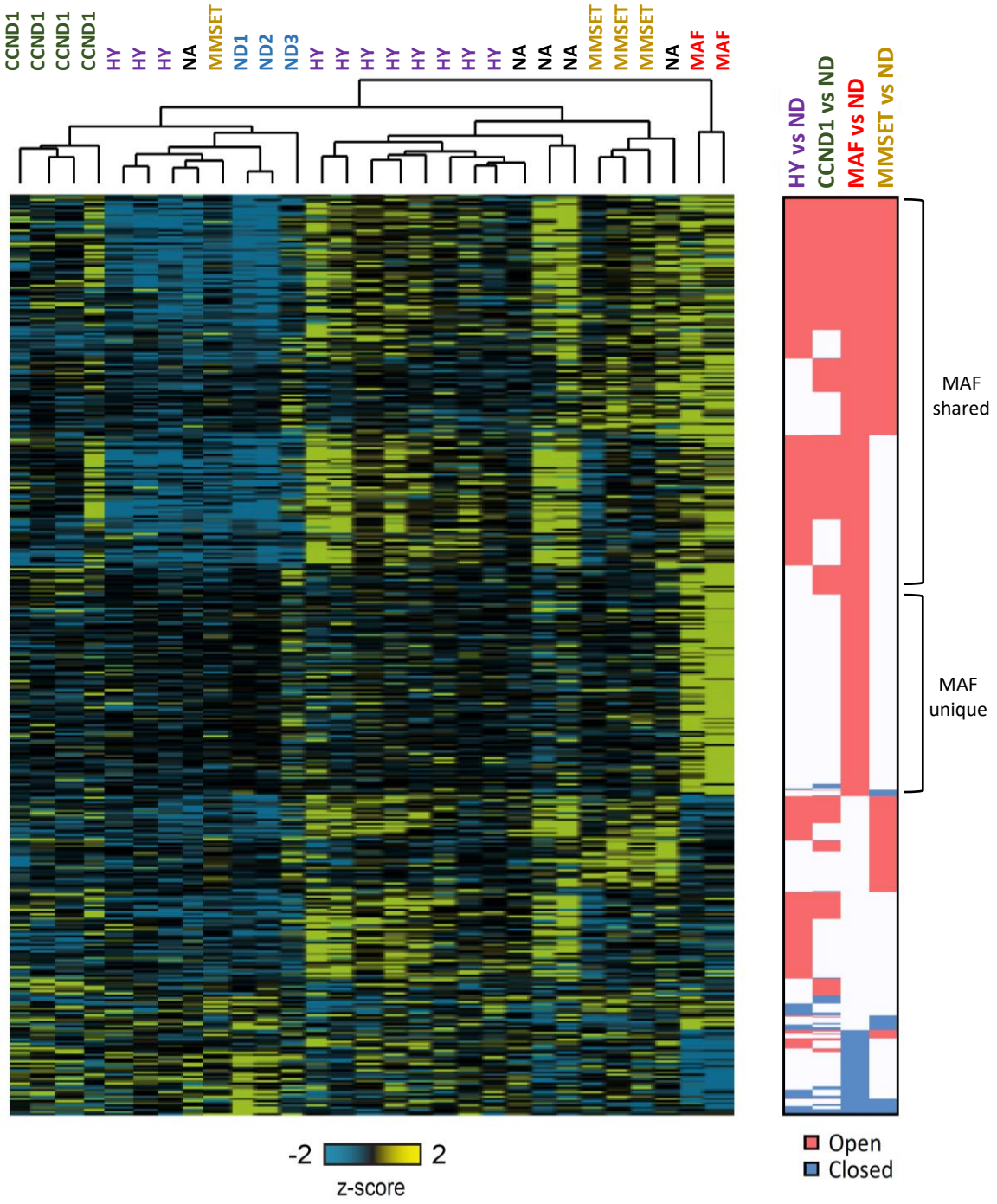


Figure 4.4: Hierarchical clustering on differential chromatin accessibility in myeloma PC

Heatmap of differentially accessible regions in primary MM PC as compared to normal PC. Dendrogram showing hierarchical clustering, performed using differentially accessible regions between myeloma and ND samples (clustering using Spearman correlation method). On the right, regions whose accessibility (ATAC-seq signal) is > 1.0 log_2FC (OPEN, red) or <-1.0 log_2FC (CLOSED, blue) compared to normal PC. Brackets indicate regions differentially upregulated in the MAF MM that are either shared with other MM subgroups, or uniquely upregulated in the MAF samples. HY; hyperdiploid, ND; normal donor PC, NA; others, includes samples with no primary translocation detected or unknown cytogenetics.

4.3 Defining the MAF regulatory genome

4.3.1 Defining the "MAF-associated" regions

In order to further delineate the MAF regulatory genome and determine which of the accessible regions found in MAF-translocated myeloma may be regulated by the transcription factor, I integrated the ATAC-seq data from the MAF primary samples with the MAF cistrome obtained by ChIP-seq against MAF in the MAF-translocated MM.1S myeloma cell line. This entailed combining the regions found to have enhanced accessibility in MAF primary myeloma cells when compared to normal bone marrow PC, with MAF binding sites. Interestingly, over one fourth (1,661/5,828; 29%) of the regions with enhanced accessibility in MAF samples were also found to be MAF-bound (**Fig. 4.5**). I termed these the "MAF-associated" regions. Only 17 out of the 813 regions with reduced accessibility in MAF samples (as compared to normal PC) were also MAF-bound (2%). Given that the differentially "closed" regions were not associated with changes in transcription (**Fig. 4.1B**) and that MAF had been shown to act as a transcriptional activator in my earlier work with myeloma cell lines (see section 3.6.1), in addition to the low MAF occupancy of these regions, I discounted the regions with decreased chromatin accessibility from further downstream analysis.

Focusing only on the regions with differentially enhanced accessibility, the "MAF-associated" regions, I found that these could be differentially accessible either uniquely in the MAF subgroup (37%), or shared with other myeloma genetic subgroups (**Fig.4.6A**). Genomic annotation revealed the majority to be intronic and intergenic (**Fig.4.6B**), suggesting a predominance of enhancers over promoters. Epigenomic annotation of the same regions was based on the MM.1S chromatin marks available from ENCODE[117], and performed as described in section 2.11.3 and illustrated in **Appendix B Fig. S4**. Although many more regions were identified as promoters based on their epigenetic marks, 35% of "MAF-associated" regions displayed epigenetic marks consistent with enhancers (**Fig. 4.6C**). The discrepancy between the genomic and epigenomic annotation may be explained by the presence of secondary promoters and the criteria used by the Homer package to annotate TSS. Motif analysis of the 1,661 "MAF-associated" confirmed the large MAF motif as the most enriched, followed by members of the IRF transcription factor family (**Fig. 4.6D**).

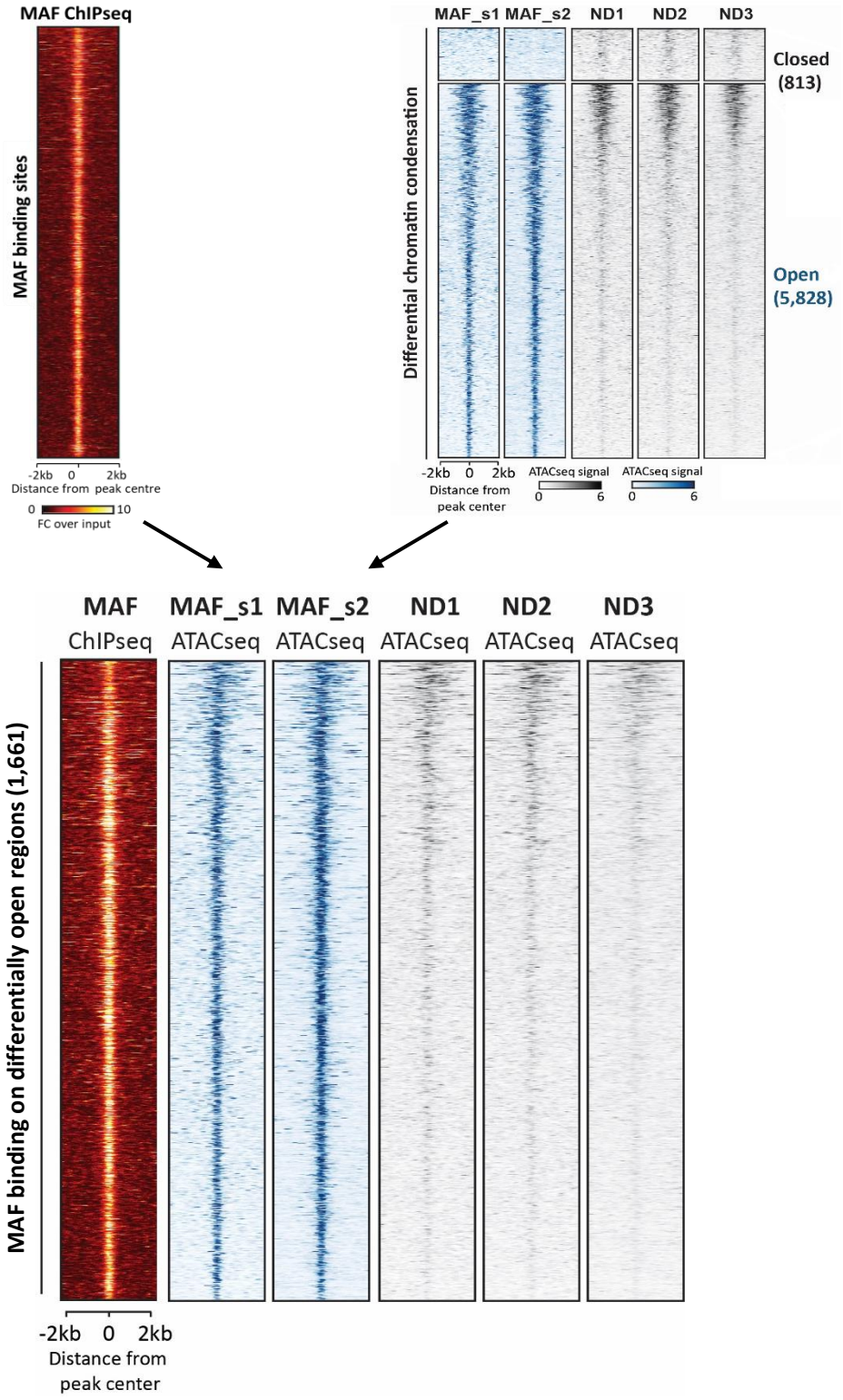


Figure 4.5: MAF bound and differentially open regions, the "MAF-associated regions"

Schematic and heatmap depicting the 1,661 "MAF-associated regions". The MAF cistrome (top left) was obtained by ChIP-seq against MAF in the MAF-translocated MM.1S myeloma cell line. This was integrated with regions in primary MAF samples (MAF_s1, MAF_s2) with enhanced accessibility as compared to normal plasma cells (ND1, ND2, ND3, top right) to obtain 1,661 "MAF-associated regions" that are both MAF-bound and differentially open in MAF myeloma cells.

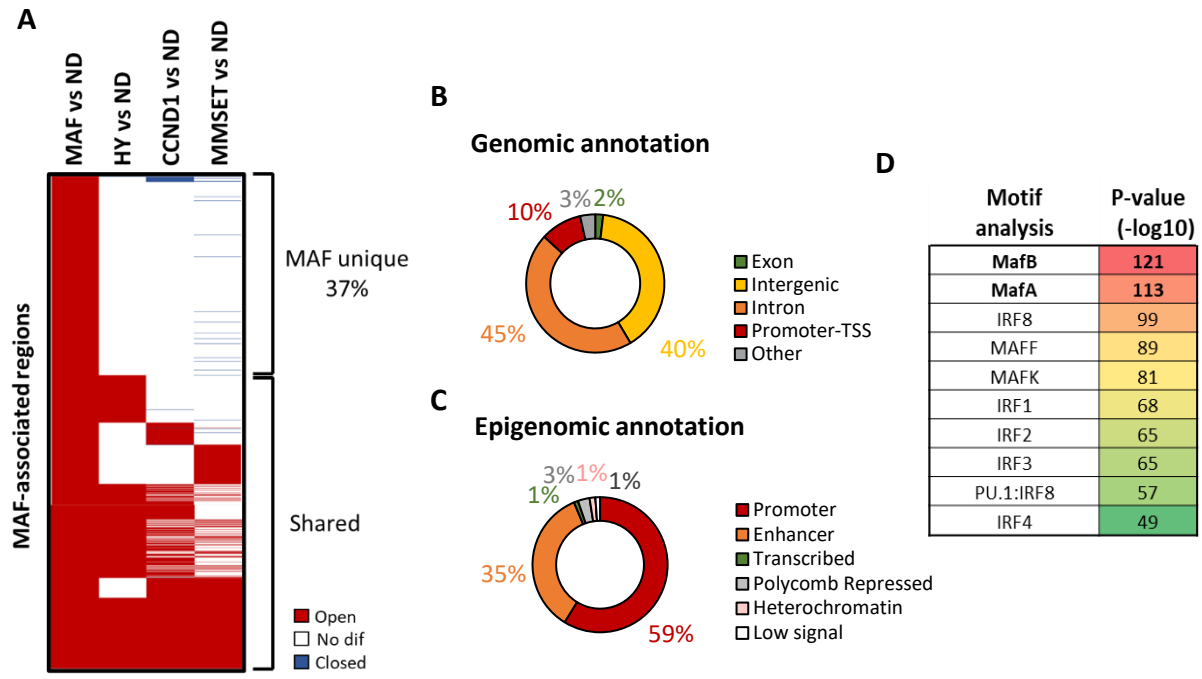


Figure 4.6: Characteristics of the "MAF-associated" regions

(A) Distribution of the 1,661 "MAF-associated" regions across the myeloma genetic groups, with 37% differentially open uniquely in MAF myeloma and the rest shared with at least one more myeloma subgroup **(B)** Genomic annotation of the 1,661 "MAF-associated" regions. **(C)** Epigenomic annotation of the same regions, based on their chromatin mark status of MM.1S available from ENCODE[117]. **(D)** Motif analysis of the "MAF-associated" regions.

4.3.2 "MAF-associated" regions across development

In order to characterise the "MAF-associated" regions across B cell development, I first explored the Blueprint Consortium RNA-seq data in normal haematopoietic cells[120], establishing that MAF is not expressed during the stages of B cell maturation (**Fig. 4.7**). Although MAF was very lowly expressed in some of tonsillar PC, the database had no samples of bone marrow (BM) PC to confirm our in-house data illustrating lack of MAF expression in BM PC (**Fig. 3.5**).

Using the epigenomic annotation of MM.1S cells available from ENCODE[117] and the chromatin marks in normal B cells and tonsillar PC obtained by the Blueprint Consortium[115], I was able to compare the chromatin mark maps of the 1,661 "MAF-associated" regions with

that of normal tonsillar PC and across the stages of B cell maturation. Interestingly, 30-40% of the "MAF-associated" regions lacked activatory chromatin marks in normal B cells and PC (**Fig. 4.8**). This was also the case for the chromatin states characterising the same regions in U266, a non-MAF MMCL; just over 40% of the "MAF-associated" regions had no epigenetic marks of activation in U266, suggesting therefore that this observation was not the effect of comparing a cell line to primary cells.

In order to specifically identify the regions that were not active in normal PC and B cells, but become activated in MAF myeloma, I juxtaposed the chromatin marks of the "MAF-associated" regions in MM.1S with those of normal tonsillar PC and B cells. This enabled me to trace each region's activity through the developmental trajectory of the mature B lineage including PC. Forty-four percent (726/1,661) of the "MAF-associated" regions did not display any activatory epigenetic marks in normal PC; I termed these the "MAF-activated" regions (**Fig. 4.9**). Indeed, the majority of these regions (582/726, 80%) also lacked activatory marks throughout the B cell maturation stages (**Fig. 4.9**). These regions were roughly equally split between epigenetically defined promoters ("MAF-activated" promoters, 369/726) and enhancers ("MAF-activated" enhancers, 357/726). The activation seen in these regions in MAF-myeloma, having previously been epigenetically inactive during the normal cell's development, suggests activation at some point during the malignant transformation of the cell. Just over half of these regions (371/726, 51%) were uniquely over-accessible in MAF-myeloma compared to other subgroups, suggesting the activation of these regions is not generic to MM as a whole. However, although epigenetic activation of the putative "MAF-activated" regions is therefore linked to MAF-myeloma, a significant proportion of them are not unique to t(14;16) myeloma PC, and instead display differential accessibility in at least one more MM subgroup. This observation does not on its own contradict the possibility of MAF-associated activation, as these regulatory regions may have been activated in other genetic MM subgroups via different mechanisms. For the same reason, when identifying the putative "MAF-activated" regions, I did not include the U266, a CCND1-translocated MMCL, chromatin mark maps. On the other hand, although activation of these regions was observed in MAF MM compared to normal PC, this does not necessitate a causative link between the transcription factor and the regions' activation, a question which I set out to address in the following section.

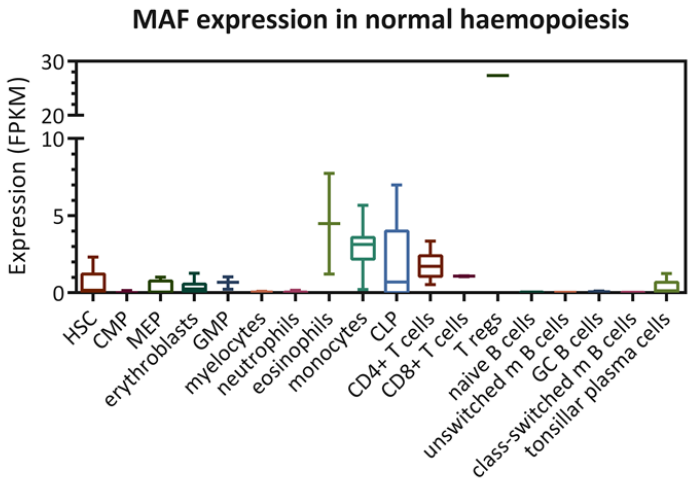


Figure 4.7: MAF expression across normal haemopoietic cells

MAF expression (in FPKM) across normal haemopoiesis, including the various stages of normal B cell development. Data available from Blueprint Consortium, using the data analysis portal[120]. HSC; haematopoietic stem cell, CMP; common myeloid progenitor, GMP; granulocyte-monocyte progenitor, CLP; common lymphoid progenitor, m B cells; memory B cells, GC; germinal centre

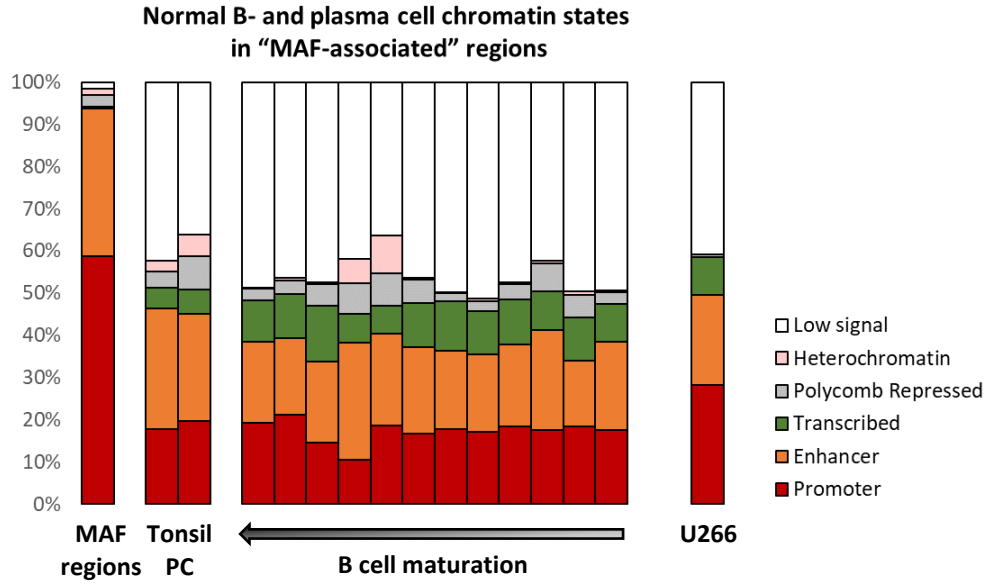


Figure 4.8: Comparative chromatin states of "MAF-associated" regions across B cell development

Chromatin states of the 1,661 "MAF-associated" regions in normal tonsillar plasma cells and across the stages of B cell maturation, from naïve B cell through to class-switched memory B cell (Blueprint Consortium data[115]). From left to right, bars represent "MAF-associated" regions, tonsil plasma cells (2 samples), class-switched memory B cells (3 samples), germinal centre B cells (2 samples), unswitched memory B cell (1 sample) and naïve B cells (6 samples). U266 ChromHMM status of the same regions shown on the right for comparison.

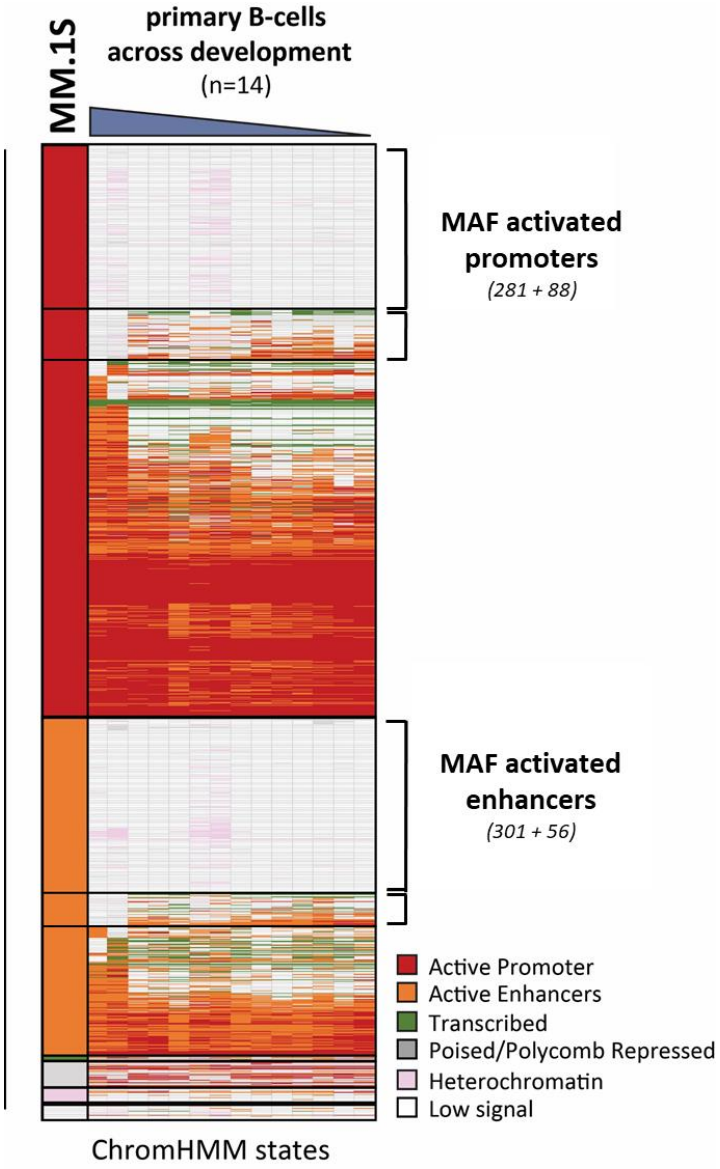


Figure 4.9: Putative MAF-activated regions

Heatmap of the "MAF-associated" regions against the histone map marks of normal tonsillar plasma cells (first two samples on the right of MM.1S) and B cells across B cell maturation stages (Blueprint Consortium data[115]), identifying a group of epigenetically defined promoters and enhancers which lack activatory marks in tonsillar plasma cells or in both plasma cells and throughout B cell development, the "MAF-activated" regions. B cell maturation stages in order as per figure 4.8.

4.4 MAF as an activator of specific regulatory regions

Thus far, my exploration of the MAF regulatory genome revealed the presence of 1,661 "MAF-associated" regions, that are both bound by the transcription factor and display enhanced accessibility in MAF-myeloma in comparison to normal PC. Using chromatin marks of the regions in MM.1S cells, these were identified as both promoters and enhancers, although on genomic annotation 85% lie outside TSS. Juxtaposing the chromatin mark maps of these regions against those of normal PC and developing B cells revealed 43% of those to be inactive in normal cellular counterpart; the majority lacked activatory marks throughout B cell maturation as well as PC, whereas a smaller minority displayed activatory marks at some point during B cell maturation but not in normal PC. These regions are therefore associated with MAF-myeloma in that they are MAF-bound and epigenetically active in t(14;16) MM, having previously been inactive during normal B cell development.

In order to investigate the role of the MAF transcription factor in the activation of these putative "MAF-activated" regions, I ectopically overexpressed MAF in a MAF-negative myeloma cell line, U266, which carries the t(11;16) translocation, juxtaposing *CCND1* to the *IgH* enhancer. For this purpose, the *MAF* cDNA (see Appendix B Table S2 for sequence) was cloned into the retroviral MIGR1 GFP+ vector, as described in section 2.4.2, and the empty vector (EV) was used as a negative control. Transduced cells were FACS-sorted, using untransduced cells to establish the gating strategy as shown in **Appendix D Fig. S1**.

In order to identify the optimal time to harvest cells for downstream applications, a time course experiment was performed, during which RNA was collected from GFP+ flow-sorted cells on days 3, 5 and 7 post transduction and the expression of *MAF* and known MAF target genes was assessed (**Appendix D Fig. S2**). As expected, *MAF* expression was substantially increased. The effect on the expression of MAF target genes was most pronounced on days 5 and 7, so day 5 was chosen as optimal due to the higher MAF expression seen on that day as well as slightly higher expression of *ITGB7* and *CCND2*. A similar picture was seen in KMS12 BM, another t(11;14) MMCL, though the effect was generally less pronounced on the MAF

target genes and *CCND2* expression was not increased; therefore U266 were used for further experiments.

Target gene overexpression on day 5 following MIGR1 MAF transduction was confirmed in a series of replicate experiments (**Fig. 4.10A**). There was a marked increase in the expression of *CCR1* and *ITGB7*, however other targets such as *TLR4* did not reach statistical significance. In addition, there was only a modest effect on *CCND2*, perhaps due to the inhibitory genetic background of U266 overexpressing CCND1. MAF protein expression was also confirmed to be highly increased (**Fig. 4.10B & C**).

Transcriptome analysis revealed 1,445 genes were upregulated in MIGR MAF+ cells when compared to EV control, and 1,048 were downregulated (**Fig. 4.11**). Known MAF target genes such as *ITGB7* and *CCR1* were amongst the highest upregulated genes. Other known MAF associated genes, such as *CCND2*[28, 45, 68], *ABLIM1*[28, 67], *RND3*[68] and *RASSF4*[67], were also overexpressed. However, other known MAF targets, such as *TLR4* or *NUAK1* were not significantly differentially expressed. This could be perhaps due to inhibitory factors present in U266 or the lack of secondary signals possibly required for the expression of these genes, such as from the bone marrow microenvironment. In order to ensure that the transcriptome seen in MAF-overexpressing U266 was representative of MAF-translocated myeloma, I compared it to other known and in-house datasets (**Fig. 4.12**). Indeed, the transcriptome seen in MAF-transduced U266 cells was significantly enriched in genes known to be upregulated in the MF myeloma subgroup from GEP studies[67] (**Fig. 4.12A**). In addition, the MAF-overexpressing U266 transcriptome was significantly enriched in MAF target genes, as identified in ChIP-seq datasets against MAF on T helper cells[151] (**Fig. 4.12B**). Finally, over 1/4 of genes (27%) found differentially upregulated upon MAF overexpression, were also downregulated upon MAF knockdown in MM.1S (see section 3.4) (**Fig. 4.12C**). Taken together, these data confirmed that, despite some differences in transcriptomes, which would be expected given the background genetic differences, MAF-overexpressing U266 indeed recapitulated transcriptomic changes imposed by MAF translocation in MM.

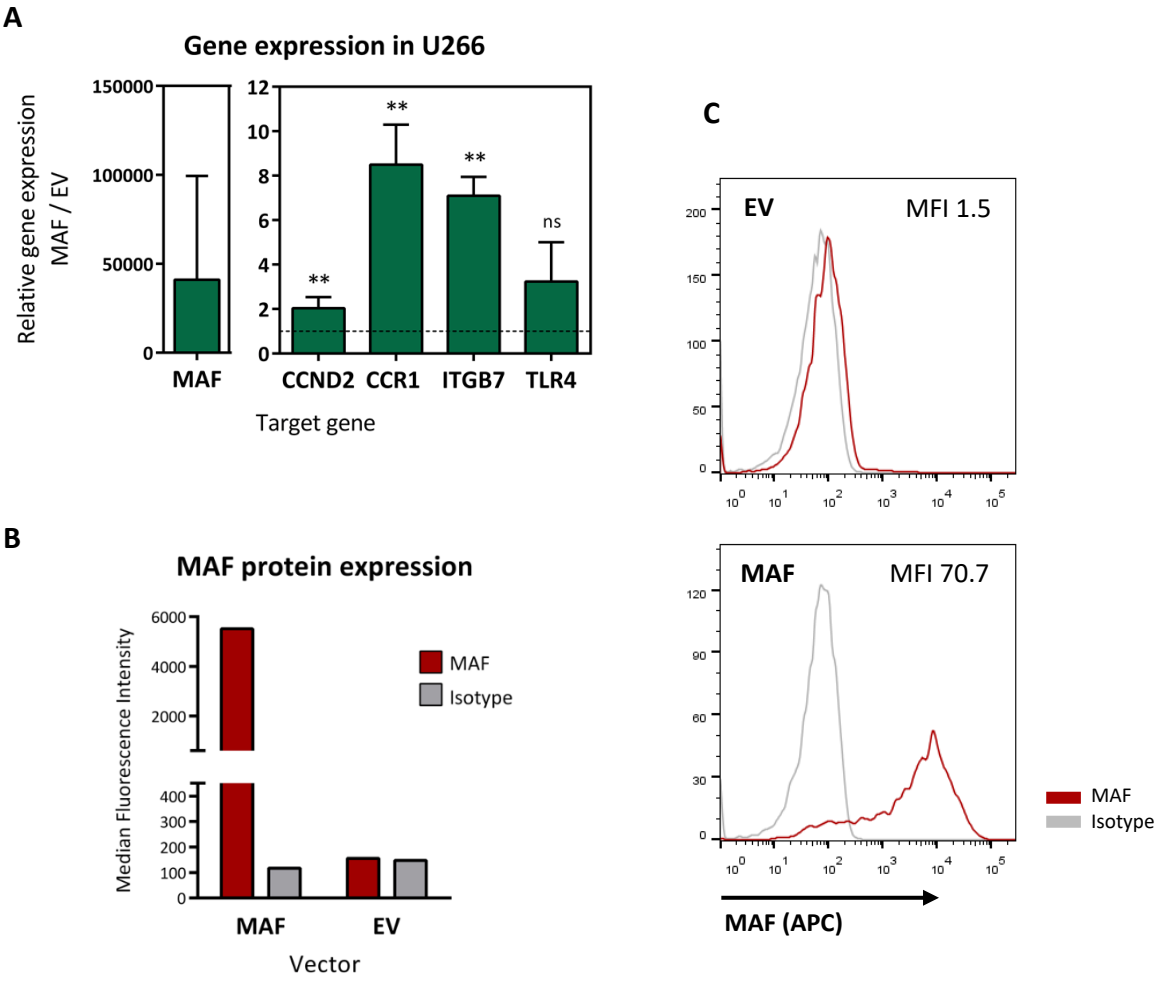


Figure 4.10: MAF overexpression in U266 cells

(A) Relative expression of MAF target genes upon MAF overexpression in the MAF-negative MMCL U266, obtained by qPCR. Expression normalised to *GAPDH* and expressed as ratio between MIGR1 MAF+ U266 and MIGR1 EV control U266 cells. Bars represent mean + SD of n=5 experiments. Horizontal line, relative expression MAF/EV =1. Statistical analysis comparing gene expression between MIGR MAF and MIGR EV cells performed using the non-parametric Mann-Whitney U test; **p<0.01, ns; non-significant. **(B)** Median Fluorescence intensity of MAF (red) vs isotype control (grey) in MIGR1 MAF+ transduced and EV control cells, as assessed by flow cytometry. Bars show mean of n=2 independent experiments. **(C)** Representative flow cytometry plots of MAF protein expression in transduced MIGR1 MAF+ and EV control cells. MFI; Median Fluorescence Intensity, expressed as ratio of MFI of MAF stain (red) over isotype control (grey). EV; empty vector.

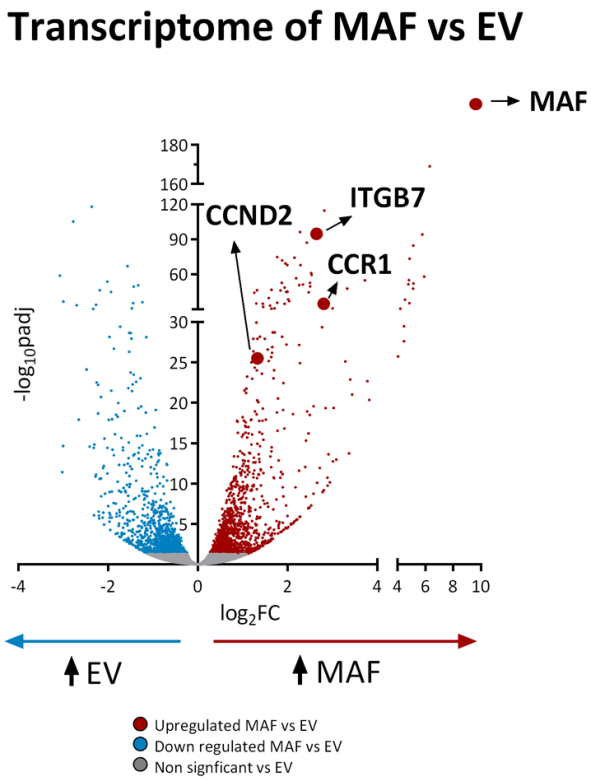


Figure 4.11: Transcriptome of MAF-overexpressing U266

Volcano plot of transcriptome following MAF-overexpression in the U266 myeloma cell line compared to empty vector control. *MAF* itself as well as indicative known MAF target genes highlighted. EV; empty vector

Having established that MAF-overexpressing U266 cells undergo changes in gene expression representative of the transcriptome characterising MAF-translocated myeloma, I used this experimental model to study the chromatin accessibility changes imposed by MAF overexpression at the "MAF-associated" regions.

The chromatin accessibility profile in MAF-overexpressing U266 when compared to EV control cells showed increased overall accessibility in EV cells rather than MAF+ cells. (**Fig. 4.13A**). By contrast however, when focusing on the "MAF-associated" regions (MAF-bound regions with differentially enhanced accessibility in MAF primary MM samples vs normal PC), the accessibility profile reversed, with enhanced overall accessibility observed in the MAF+ cells (**Fig. 4.13A**). Specifically, in the "MAF-associated" regions, 76% (1,254/1,661) acquired differential accessibility following MAF overexpression (**Fig. 4.13C**). Remarkably, 64% (805/1,254) of these showed enhanced accessibility when compared to EV control cells (**Fig. 4.13C**). Hypergeometric distribution testing, comparing the ratios of enhanced accessibility

over the total number of differentially accessible regions, observed in the "MAF-associated" vs genome-wide regions (49.1% enhanced accessibility in MAF+ cells genome-wide vs 64.2% in "MAF-associated" regions) indicated these observations were not random (p value $1.52x10^{-27}$, **Fig 4.13B**).

Of the MAF-associated regions, the regions of highest interest were those found to lack activatory epigenetic marks in B cells and / or normal tonsillar PC, the putative "MAF-activated" regions, as these regions suggest a possible activatory role of MAF during the development of MAF-translocated myeloma. Upon MAF overexpression in the MAF-negative MMCL U266, 382/726 "MAF-activated" regions acquired enhanced accessibility (hypergeometric distribution test, p value $1.23x10^{-46}$). Specifically, over half of the MAF-activated promoters (196/369) and enhancers (186/357) acquired enhanced accessibility, whereas only a very small proportion displayed reduced accessibility, supporting the role of MAF as a direct activator in these regions (**Fig. 4.14A**). Interestingly, the majority (65%, 248/382) of these regions displayed enhanced accessibility uniquely in the MAF MM subgroup.

Enrichment analysis on the validated "MAF-activated" regions, i.e. the 382 "MAF-activated" regions that acquired enhanced accessibility upon MAF-overexpression, using the online tool GREAT, which assigns regions to the nearest genes, showed the MF myeloma subgroup genes to be the most significantly enriched (**Appendix D Fig. S3**). This suggested that these regions regulate genes known to be associated with the MAF subgroup. In order to ascribe these regulatory regions to genes that more faithfully represent directly regulated MAF targets, I annotated the regulatory regions using the genes assigned to the ChIP-seq peaks following MAF knockdown in MM.1S (see section 3.6.1). Interestingly, some of the known MAF target genes, such as *CCR1*, *CX3CR1* and *NUAK1*, were predicted to be regulated by MAF-activated regulatory regions (**Fig. 4.14B**). Taken together this data suggests that MAF acts as an activator of regulatory regions, previously inactive in the normal PC, and mostly also during B cell maturation, and thus leading to upregulation of its target genes.

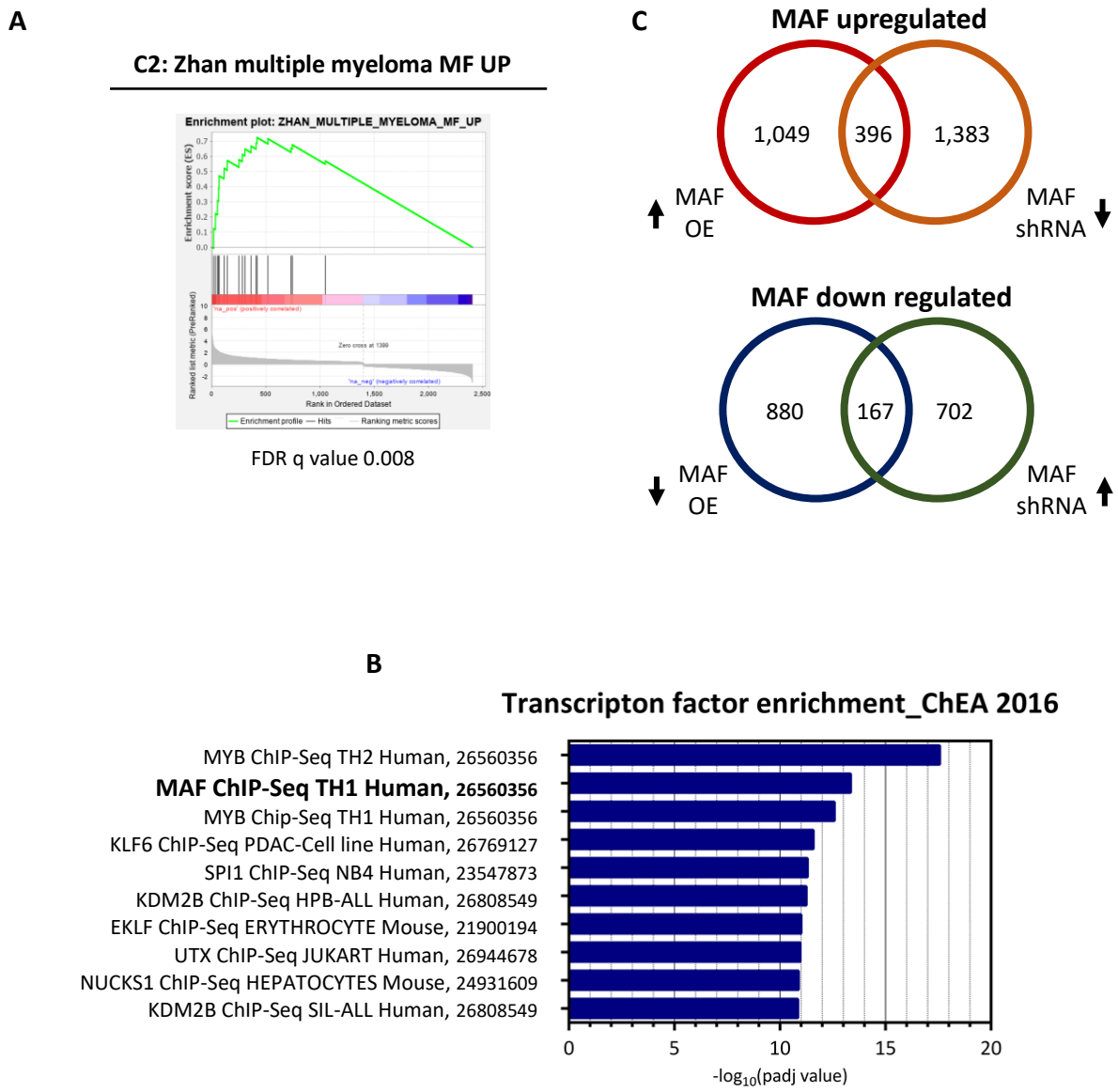


Figure 4.12: Comparative analysis of MAF overexpression in U266 with MAF transcriptome

(A) Enrichment analysis plot, produced using the GSEA software[104], showing significant enrichment of the MF MM subgroup genes from GEP studies[67] in the genes differentially expressed upon MAF-overexpression in the U266 myeloma cell line. **(B)** Top 10 significantly enriched transcription factor ChIP-seq datasets (showing pubmed manuscript ID, PMID), using the EnrichR online tool[105], highlighting MAF ChIP-seq in T helper cells as the second most significantly enriched in the MAF-overexpressing U266 transcriptome (MAF vs EV). **(C)** Comparison of the differential transcriptome obtained following MAF overexpression in U266 with that following MAF knockdown in MM.1S, showing common genes upregulated (top) and downregulated (bottom) by MAF. OE; overexpression, ChEA; ChIP-X Enrichment analysis, PDAC; pancreatic ductal adenocarcinoma, NBA; promyelocytic leukaemia cell line, HBP-ALL, Jurkat, SIL-all; T-cell acute leukaemia cell lines

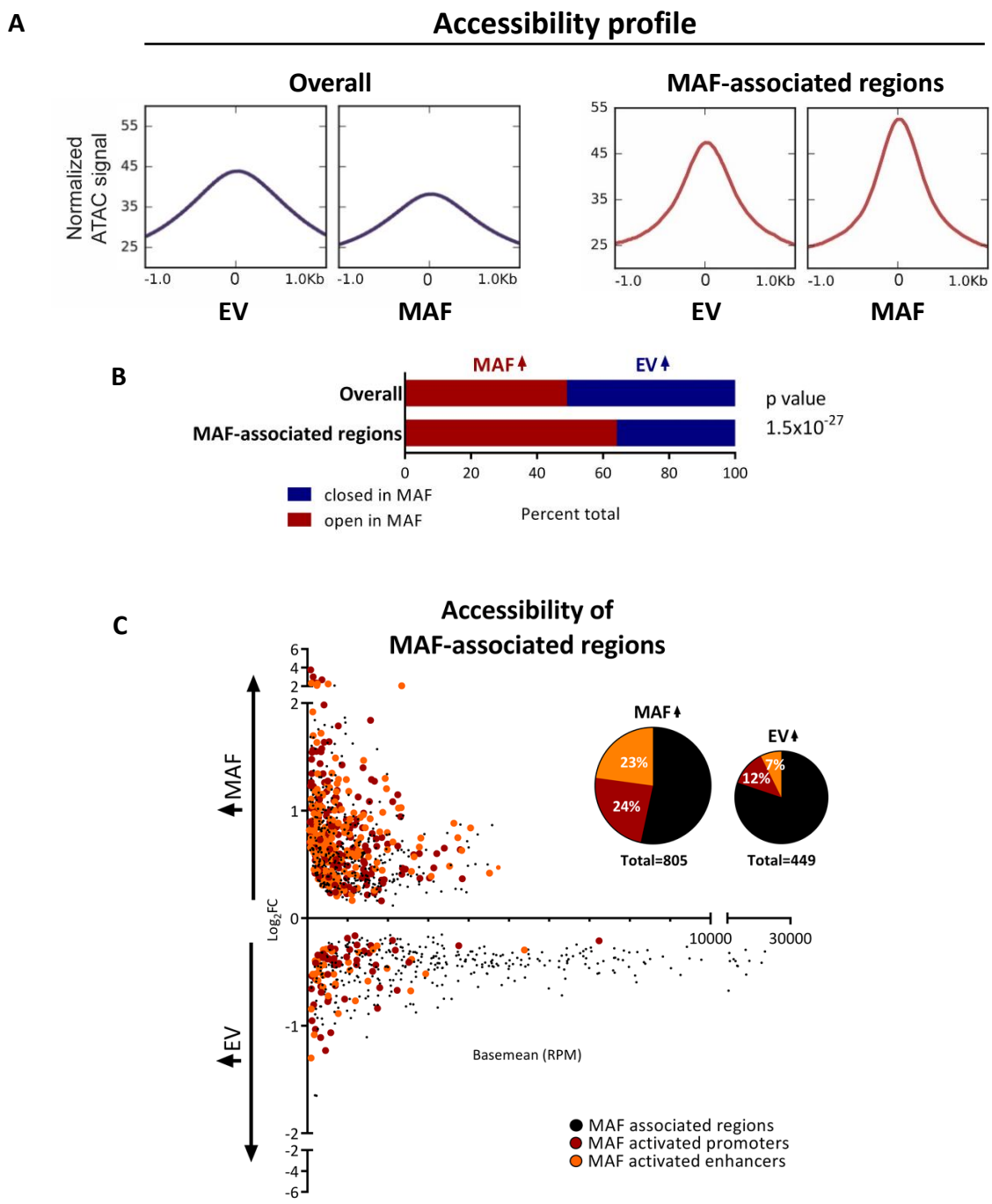


Figure 4.13: Accessibility of MAF-overexpressing U266

(A) Accessibility profiles seen in MAF-overexpressing U266 compared to EV control cells, genome-wide ("overall", in blue) and in "MAF-associated regions" (in red). **(B)** Percentages of "open" vs "closed" differential accessibility regions following MAF overexpression compared to EV, genome-wide and in "MAF-associated" regions, with p-value representing hypergeometric distribution test between genome-wide and MAF-associated regions. **(C)** MA plot of differential accessibility following MAF overexpression, focusing on the MAF-associated regions. Highlighted are the "MAF-activated" promoters and "MAF-activated enhancers", in red and orange respectively. Pie charts depicting number of "MAF-associated" regions with increased or reduced accessibility in MAF-overexpressing cells and percentage of MAF-activated promoters and enhancers in each. EV; empty vector.

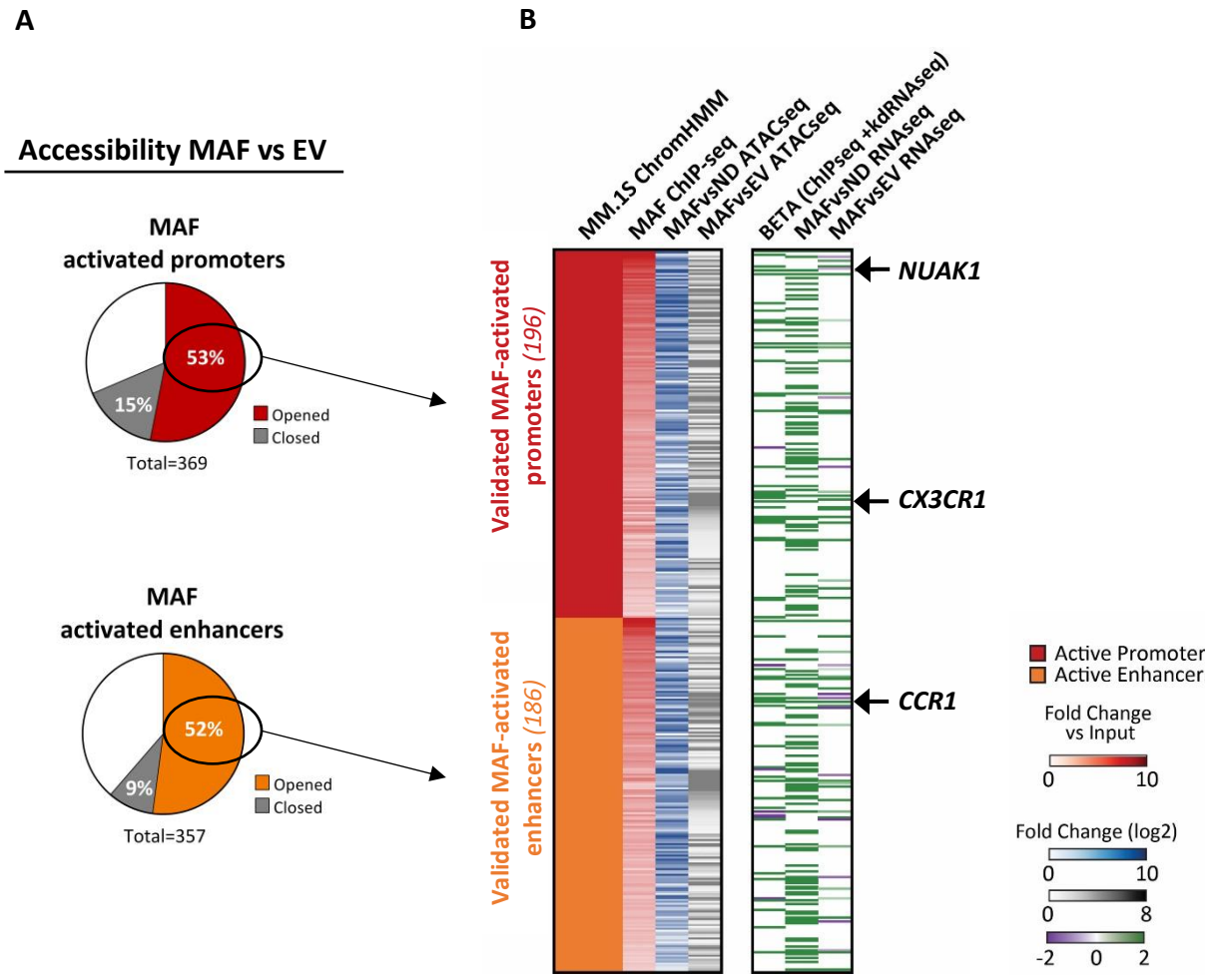


Figure 4.14: Validated MAF-activated promoters and enhancers

(A) Pie charts depicting percentage of MAF-activated promoters and enhancers with enhanced ("opened") or reduced ("closed") accessibility following MAF-overexpression in U266. **(B)** Heatmap depicting fold change in signal or corresponding gene expression for each of the "opened" MAF-activated promoters and enhancers following MAF overexpression. In order from left to right: MAF ChIP-seq; fold change over input in ChIP-seq against MAF in MM.1S, MAF vs ND ATAC-seq; log_2fold change in ATAC-seq signal of primary MAF MM samples when compared to normal plasma cells, MAF vs EV ATAC-seq; log_2fold change in ATAC-seq signal in MAF-overexpressing U266 when compared to empty vector control, Beta (ChIP-seq + kd RNAseq); Beta score of change in gene expression following MAF shRNA-mediated knockdown in MM.1S cells, MAF vs ND RNA-seq; log_2fold change in gene expression in primary MAF MM samples when compared to normal plasma cells; MAF vs EV RNA-seq assigned using the Beta-plus package; log_2fold change of gene expression in MAF-overexpressing U266 when compared to empty vector control.

4.5 Validation of the *CCR1* MAF-activated enhancer

Of the validated "MAF-activated" regulatory regions, the enhancer predicted to regulate *CCR1* was of particular interest, as *CCR1* is one of the direct MAF targets that has recently been shown to have potential clinical implications[84]. C-C Motif Chemokine Receptor 1 (CCR1) is involved in MM cell chemotaxis[45] and promotes cell proliferation upon interaction with bone marrow stromal cells[80]. More recently CCR1 was implicated in myeloma cell dissemination *in vivo*, with its inhibition leading to a reduction in myeloma cell dissemination both in bone marrow and peripheral blood[84]. My earlier results, discussed in section 3, showed *CCR1* to be a direct, MAF-bound target, which is in agreement with previous results showing *CCR1* downregulation upon MAF knockdown[45].

Study of the MAF-associated regulatory genome as described in the experiments above, suggested the presence of a MAF-bound enhancer peak downstream of *CCR1.* Inspecting the accessibility signal in MAF primary samples, revealed 3 putative enhancer peaks downstream of the *CCR1* gene (**Fig. 4.15**). Despite being MAF-bound, the most distal peak, "Peak 1", did not show statistically significant differential accessibility when compared to normal PC, whereas the two more proximal peaks, "Peak 2 and 3" showed enhanced accessibility vs normal PC, and therefore were included in the aforementioned analysis, as "MAF-associated" peaks. Of these two peaks, Peak 3 was marked by H3K4me1 in both normal PC and B cells. Peak 2 on the other hand, displayed no activatory chromatin marks in normal PC or indeed at any stage of B cell maturation, therefore forming part of the putative "MAF-activated" peaks. Indeed, upon MAF overexpression in U266, only Peak 2 and not Peak 3, acquired enhanced accessibility when compared to EV control cells, validating it as a "MAF-activated" enhancer. Since this peak (Peak 2) is inactive in normal PC and B cells and its activity is enhanced upon MAF-overexpression, it potentially drives the observed upregulation of *CCR1* transcription.

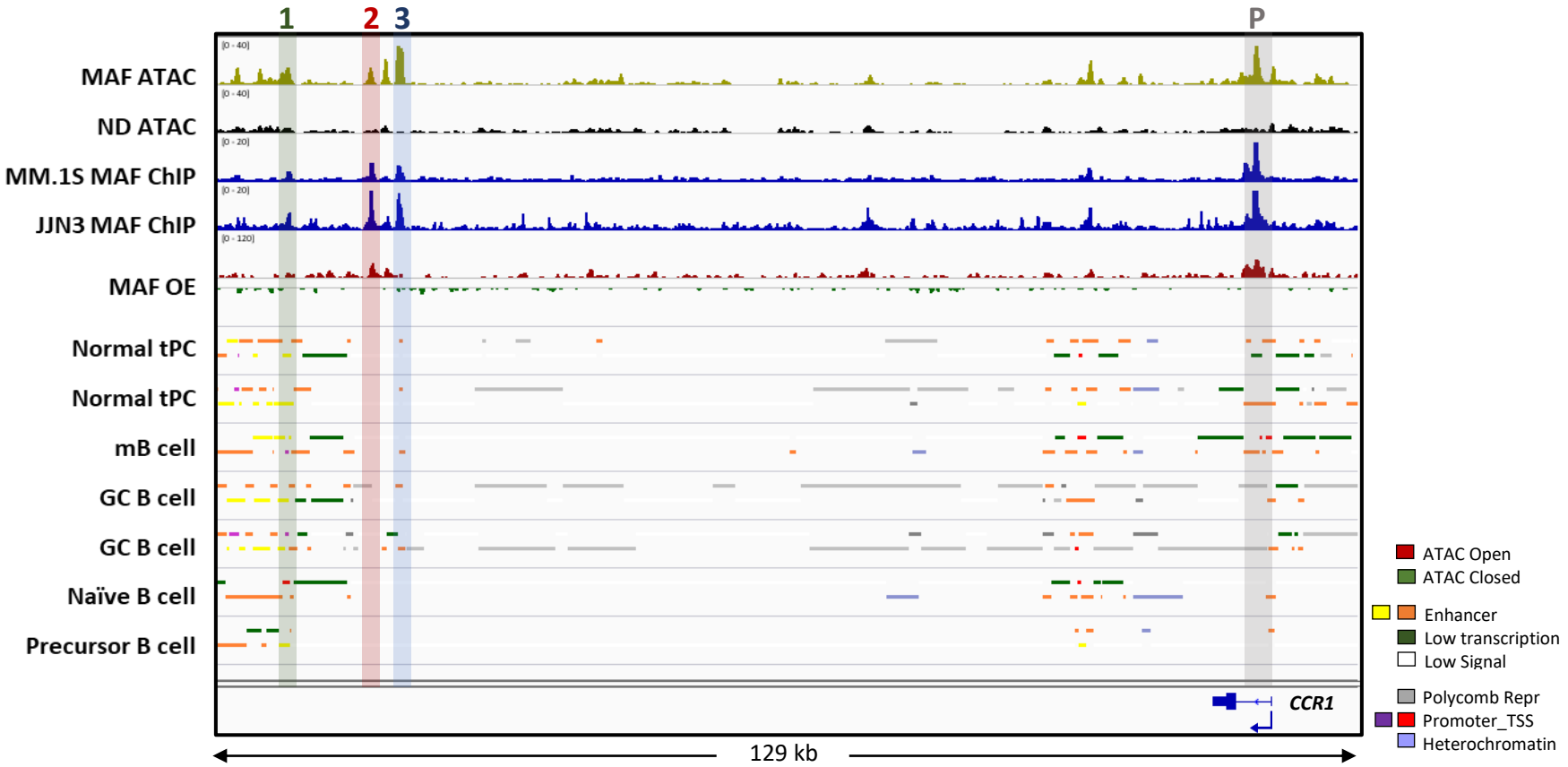


Figure 4.15: *CCR1* predicted enhancer

IGV plot depicting predicted enhancer downstream of *CCR1*. Putative enhancer peaks highlighted (peak 1, 2 and 3), as well as gene promoter (P). Tracks from top: MAF ATAC; average normalised ATAC-seq signal in MAF primary samples, ND ATAC; ATAC-seq signal in normal plasma cells, MM.1S MAF ChIP; ChIP-seq against MAF in MM.1S (FC over input), JJN3 MAF ChIP; ChIP-seq against MAF in JJN3 (FC over input), MAF OE; MAF overexpression in U266 (normalised ATAC-seq signal MAF vs empty vector) and ChromHMM chromatin states in normal tonsillar plasma cells (normal tPC), class-switched memory B cells (mB cell), germinal centre B cells (GC cell), naïve and precursor B cells (Blueprint Consortium data[115]).

In order to experimentally confirm the enhancer as a regulatory region of *CCR1*, I employed a CRISPRi method that enables repression of the transcriptional activity of enhancers of interest. For this purpose, I used an inducible lentiCRISPR-dCas9-KRABv2 vector, which upon doxycycline induction, expresses an inactive, "dead" CAS9 nuclease (dCAS9) fused to the Krüppel associated box (KRAB) domain, a transcriptional repressor. This method takes advantage of the bacterial derived CRISPR (clustered regularly interspaced palindromic repeats) system where short RNA sequences guide a Cas (CRISPR-associated) nuclease to target DNA, marking it for cleavage[152]. In the case of CRISPRi, an inactivated dCas9 protein, derived from *Streptococcus pyogenes*, is fused to a KRAB domain, allowing recruitment of transcriptional repressor complexes to DNA that is complementary to the guide RNA[153] (**Fig. 4.16**). Cell transduction with a CRISPRi-KRAB vector has been shown to lead to a 40-80% reduction in target gene expression with most guide RNAs (sgRNA), with few off-target effects[153, 154]. By contrast to RNAi methods, the CRISPRi technology, by employing a transcriptional repressor, lends itself to targeting non-coding regulatory regions[155], therefore making it an ideal tool studying putative MAF-dependent enhancers.

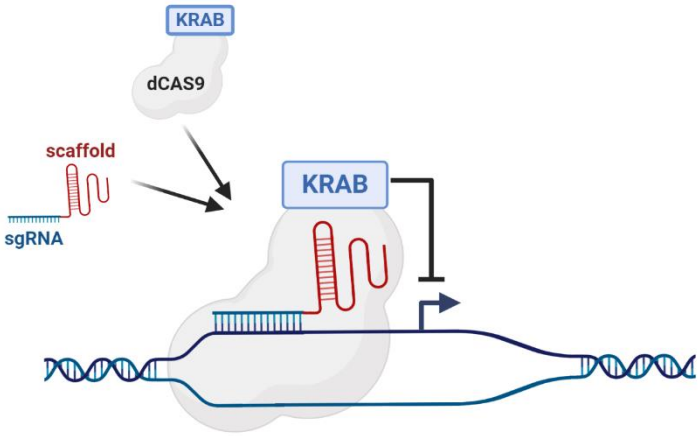


Figure 4.16: CRISPRi schematic representation

Schematic representation of CRISPRi function, depicting dCAS9 fused to a KRAB domain, which is guided to a target DNA sequence via a complementary guide RNA (sgRNA). Upon binding, the KRAB domain recruits repressive complexes in the DNA area of interest. Schematic created using online tool biorender.com.

Use of the lentiviral dCas9-KRAB vector was optimised in a series of pilot experiments as part of my work for a separate project. First, I established the concentration of doxycycline tolerated by the JJN3 MAF-translocated myeloma cell line, showing that doxycycline had no impact on cell viability at both 1μg/ml and 2μg/ml, when topped up every 3 days, every 48 hours or even daily (**Appendix D Fig. S4**). Having observed during the use of an inducible lenti-Cas9 vector that application of doxycycline every 72 hours, and to a lesser extent every 48 hours, led to fluctuant GFP levels (**Appendix D Fig. S5**), I compared induction levels of the dCas9-KRAB vector when applying doxycycline at different schedules. Indeed, more steady levels of induction, both in terms of fraction of induced cells and intensity of GFP fluorescence were achieved when applying doxycycline daily as opposed to 48-hourly (**Appendix D Fig. S6**). Finally, the concentration of doxycycline used (1μg/ml or 2μg/ml) had little effect on fraction of cells induced, induction levels over time, intensity of GFP or level of target gene expression following induction (**Appendix D Fig. S7**), therefore a concentration of 1μg/ml freshly reconstituted doxycycline, replenished daily, was used for subsequent experiments.

For the functional validation of the putative *CCR1* enhancer, two sgRNAs were designed per peak (**Fig. 4.17**). In addition, two guides targeting the *CCR1* promoter were used as positive controls and a non-targeting gal4 sgRNA as a negative control. Guides were designed and cloned into the lentiCRISPR-dCas9-KRABv2 vector as described in section 2.4.3. Following lentiviral transduction, transduced cells were selected with 5μg/ml puromycin and allowed to grow without selection pressure for 10-14 days, prior to induction with doxycycline, as described above. Post induction, cells were harvested for RNA, using FACS-sorting (gating strategy followed as per previous experiments, see **Appendix D Fig. S1**), on day 4, which was identified as the optimal time for cell harvest, following a time course experiment of 3 indicative guides (**Appendix D Fig. S8**). Transduction efficiency, measured by flow cytometry as percentage of live cells at 48 hours post puromycin, and induction efficiencies (measured as % GFP+ cells 24 hours post doxycycline) per sgRNA are outlined in **Appendix D Table S1**.

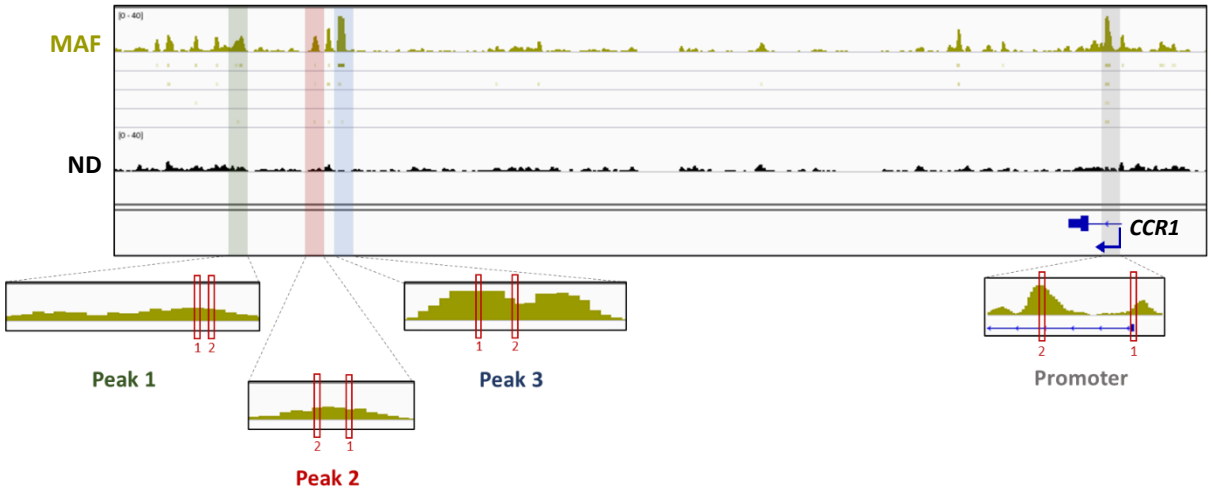


Figure 4.17: CCR1 putative enhancer sgRNA target sites

Target sites of sgRNAs in three peaks of the putative CCR1 enhancer. Two sgRNAs (1 and 2, in red) were designed per peak and 2 further sgRNAs were directed against the CCR1 promoter as positive controls.

Targeting the *CCR1* promoter with the CRISPRi constructs, as expected, significantly reduced *CCR1* mRNA expression as assessed by qPCR (**Fig. 4.18**). Of the putative enhancer peaks, repression of the "MAF-activated" peak 2, the only "MAF-associated" peak to acquire enhanced accessibility following MAF overexpression, had the most pronounced effect on *CCR1* expression, similar to that of the promoter-targeting guides. A smaller effect was seen when targeting peak 1, which is active throughout B cells and normal PC and displays accessibility in normal PC. No effect on *CCR1* expression was seen when targeting peak 3, which was also unaffected following MAF overexpression. Reassuringly, the magnitude of the effect of each sgRNA was independent of the transduction or induction efficiency (**Appendix D Table S1**). Thus, the *CCR1* enhancer peak shown experimentally to be directly MAF-activated following MAF overexpression and which has no activatory epigenetic marks in PC and B cells, was the one to have the most significant impact on *CCR1* expression in MAF-translocated myeloma cells. This observation is also in agreement with the observed very low *CCR1* expression during B cell development (**Fig. 4.19**).

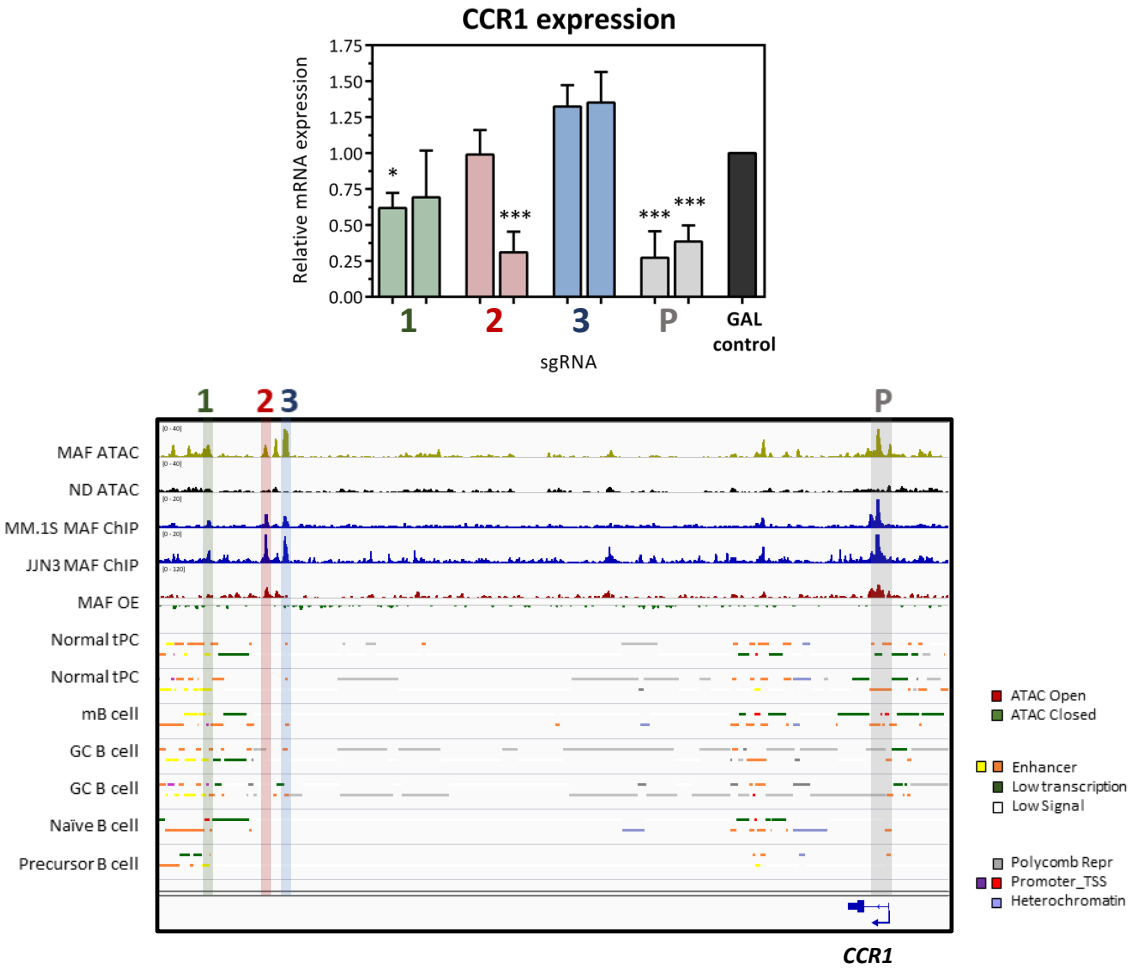


Figure 4.18: *CCR1* expression following CRISPRi targeting of enhancer peaks

CCR1 expression, as measured by qPCR (top), following CRISPRi targeting of each of the enhancer peaks, 1, 2 and 3 and promoter peak (P), as shown in IGV plot (bottom). JJN3 cells were transduced with a dCas9-KRAB vector with a guide targeting each one of the peaks, and RNA was harvested on day 4 post induction. Two guides were used per peak and compared to a non-targeting gal4 control guide. Expression normalised to *GAPDH* and gal4 control. Bars represent mean + SD, n=3. *p < 0.05, ***p < 0.001, 1-way ANOVA with Dunnett's post-hoc multiple comparisons correction. MAF ATAC; normalised ATAC-seq signal in primary MAF MM samples, ND ATAC; normalised ATAC-seq signal in normal BM plasma cell samples, MAF overexpression in U266 (normalised ATAC-seq signal MAF vs empty vector), MAF ChIP; ChIP-seq against MAF in MM.1S and JJN3 (FC over input), Chromatin states depicted for: normal tPC; normal tonsillar plasma cells, mB cell; class-switched memory B cells, GC B cell; germinal centre B cells (Blueprint consortium data[115]).

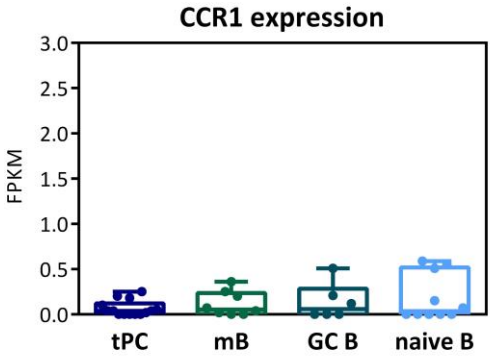


Figure 4.19: *CCR1* expression across B cell maturation

CCR1 expression, in FPKM, across B maturation stages and normal tonsillar plasma cells (Blueprint consortium data[120]). tPC; normal tonsillar plasma cells, mB; class-switched memory B cells, GC B; germinal centre B cells, naïve B; naïve B cells.

Exploring this validated "MAF-activated" peak 2 across different myeloma subgroups revealed it to have enhanced accessibility uniquely in the MAF subgroup (**Fig. 4.20A & B**), with the same being true for the *CCR1* promoter. Conversely, peak 3 also showed enhanced accessibility in the CCND1 subgroup, although to a much lesser extent than seen in MAF samples. However, my CRISPRi experiments revealed that repressing peak 3 did not have a significant effect on *CCR1* expression, suggesting this peak has a redundant or even entirely separate role. Taken together, these results suggest that enhanced accessibility at the *CCR1* promoter as well as *CCR1* enhancer (peak 2) account for the upregulated *CCR1* expression observed in the MAF subgroup compared to normal PC, an event unique to MAF myeloma compared to other MM genetic groups and a direct result of MAF upregulation (**Fig. 4.20C & D**).

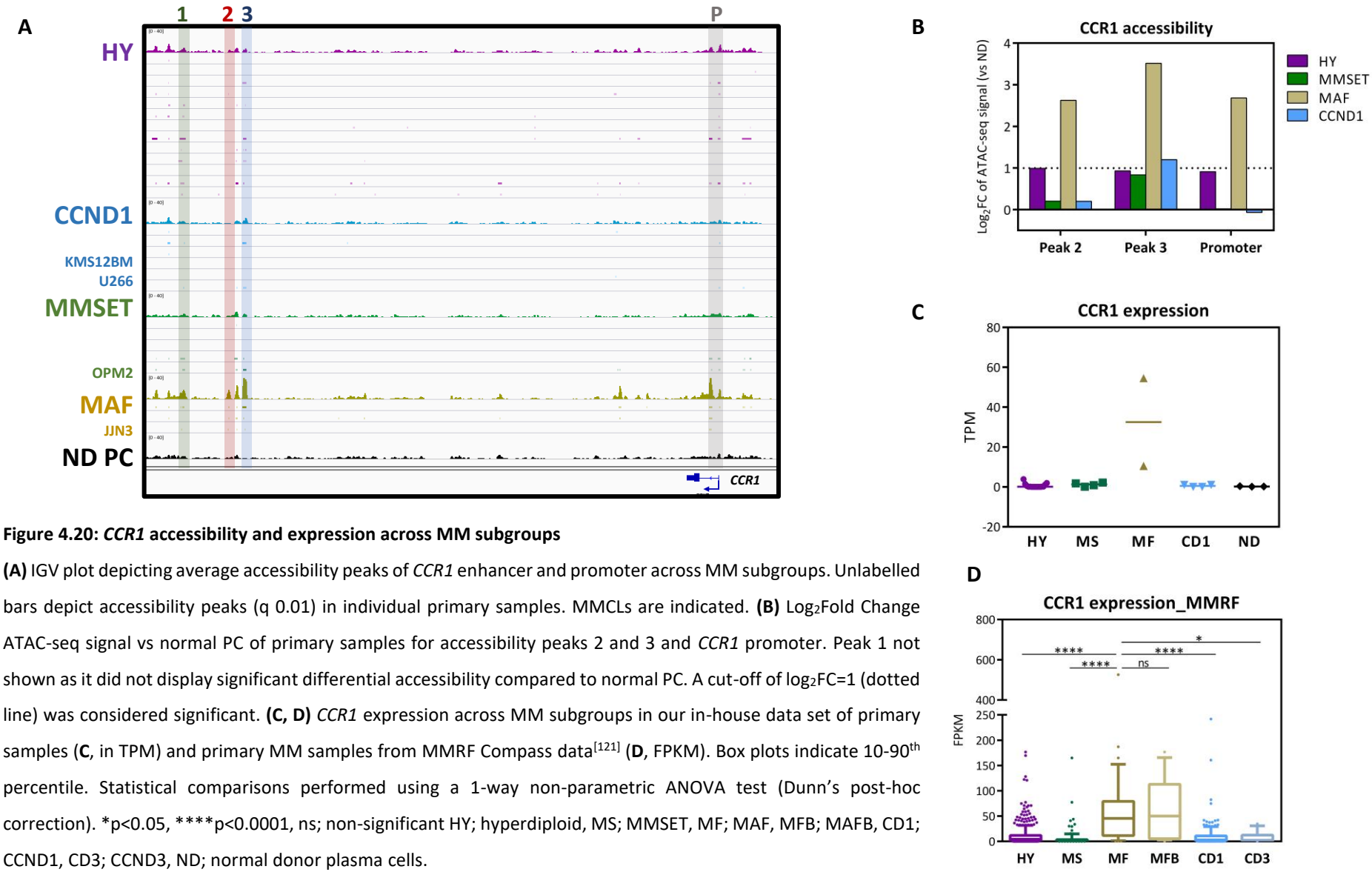


Figure 4.20: *CCR1* accessibility and expression across MM subgroups

(A) IGV plot depicting average accessibility peaks of *CCR1* enhancer and promoter across MM subgroups. Unlabelled bars depict accessibility peaks (q 0.01) in individual primary samples. MMCLs are indicated. **(B)** Log_2Fold Change ATAC-seq signal vs normal PC of primary samples for accessibility peaks 2 and 3 and *CCR1* promoter. Peak 1 not shown as it did not display significant differential accessibility compared to normal PC. A cut-off of log_2FC=1 (dotted line) was considered significant. **(C, D)** *CCR1* expression across MM subgroups in our in-house data set of primary samples (**C**, in TPM) and primary MM samples from MMRF Compass data[121] (**D**, FPKM). Box plots indicate 10-90th percentile. Statistical comparisons performed using a 1-way non-parametric ANOVA test (Dunn's post-hoc correction). *p<0.05, ****p<0.0001, ns; non-significant HY; hyperdiploid, MS; MMSET, MF; MAF, MFB; MAFB, CD1; CCND1, CD3; CCND3, ND; normal donor plasma cells.

Using the same analysis, other putative enhancer peaks, directly activated by MAF were identified, such as in the case of *CX3CR1*. C-X3-C Motif Chemokine Receptor 1 (CX3CR1) is a chemokine receptor implicated in MM cell survival[137] and myeloma-induced angiogenesis[138] and its expression has been consistently associated with the MF molecular group in GEP studies of MM patients[67, 68]. In addition, in my studies of the MAF transcriptome, I found *CX3CR1* to be a direct MAF target, MAF-bound and downregulated upon MAF-knockdown (**Fig. 3.26**). However, in addition to MAF-bound peaks upstream and downstream of the gene, which display accessibility in all MM subgroups and are constitutively active in normal PC and developing B cells, the above analysis also identified a further downstream accessibility peak (**Appendix D Fig. S9A**). This peak is also MAF bound, but does not display any activatory marks in normal PC and is uniquely differentially accessible in MAF MM samples. Upon MAF-overexpression, this peak displayed enhanced accessibility, indicating it is a "MAF-activated" regulatory region, and suggesting its activation may have a role in the upregulation of *CX3CR1* expression seen uniquely in the MAF samples (**Appendix D Fig. S9B & C**). In order to confirm this, a CRISPRi approach, similar to that employed for CCR1, would be required to experimentally validate the regulatory region's role in *CX3CR1* expression.

Not all identified putative "MAF-activated" regions were validated as directly MAF-activated upon MAF overexpression, despite being MAF bound in the context of t(14;16) MM. Such an example is a regulatory region downstream of *TLR4*. Stimulation by Toll-like receptor (TLR) ligands has been shown to increase MM cell proliferation by inducing autocrine secretion of IL6 and activation of the NF-κB pathway[139, 140]. In addition, TLRs have been implicated in immune cell evasion in the bone marrow microenvironment[141]. *TLR4* has previously been identified as one of the genes upregulated in the MF group of MM patients[67], and my analysis of the MAF transcriptome identified *TLR4* as a directly MAF-regulated gene, both MAF-bound and downregulated upon knockdown. The *TLR4* promoter is MAF-bound and shows enhanced accessibility in comparison to normal PC, though not uniquely in the MAF subgroup. The analysis of the MAF cistrome and chromatin accessibility profile also identified a peak downstream of *TLR4*, representing a putative "MAF-activated" region, MAF-bound and lacking activatory marks in B cells and normal PC (**Fig. 4.21A**). CRISPRi experiments indeed confirmed the enhancer peak, "peak 1" as a regulatory region of *TLR4* (**Fig. 4.21B**). This peak

was unique to MAF myeloma, with no accessibility seen in the other MM genetic subgroups. Indeed, *TLR4* expression followed a similar pattern, being highly upregulated in MAF samples (**Fig. 4.21C & D**), potentially suggesting that this regulatory region may responsible for the overexpression of *TLR4* in this genetic group. However, upon MAF overexpression neither the promoter nor enhancer peaks showed differential accessibility (**Fig. 4.21A**). Therefore, at least in the context of U266 MMCL, MAF overexpression alone failed to modify the *TLR4* enhancer accessibility. Consistently, *TLR4* expression was not increased in the MAF-overexpressing U266 (padj >0.99). These observations could either be due to inhibitory factors present in U266, a CCND1-translocated MMCL, or alternatively, additional factors may be required in order to activate these regions, such as binding of another transcription factor not present in U266 or environmental signalling such as for example from the bone marrow microenvironment or T cells.

Overall, my results so far are consistent with the notion that oncogenic MAF can activate in a specific manner, regulatory genomic regions that are inactive in normal plasma cells thus activating transcription of target genes. The majority (65%, 248/382) of these validated "MAF-activated" regions display enhanced accessibility uniquely in the MAF subgroup, thus explaining the mechanism by which some of the target genes are selectively upregulated in this MM genetic subgroup. However, it is not a prerequisite that the regulatory peaks need to be exclusive to the MAF subgroup, as it is possible that the same peaks are active in other MM subgroups via different mechanism.

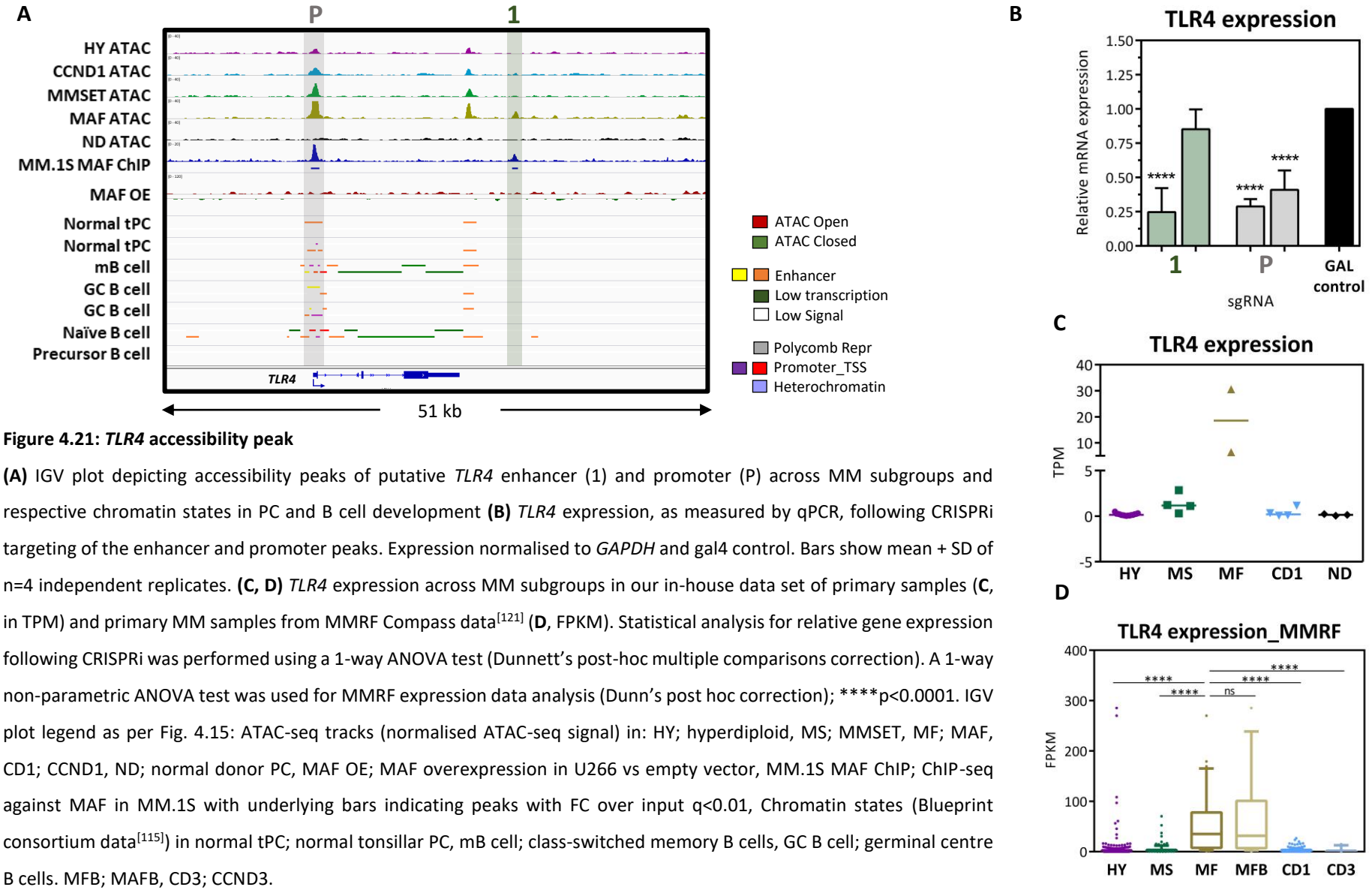


Figure 4.21: *TLR4* accessibility peak

(A) IGV plot depicting accessibility peaks of putative *TLR4* enhancer (1) and promoter (P) across MM subgroups and respective chromatin states in PC and B cell development **(B)** *TLR4* expression, as measured by qPCR, following CRISPRi targeting of the enhancer and promoter peaks. Expression normalised to *GAPDH* and gal4 control. Bars show mean + SD of n=4 independent replicates. **(C, D)** *TLR4* expression across MM subgroups in our in-house data set of primary samples (**C**, in TPM) and primary MM samples from MMRF Compass data[121] (**D**, FPKM). Statistical analysis for relative gene expression following CRISPRi was performed using a 1-way ANOVA test (Dunnett's post-hoc multiple comparisons correction). A 1-way non-parametric ANOVA test was used for MMRF expression data analysis (Dunn's post hoc correction); ****p<0.0001. IGV plot legend as per Fig. 4.15: ATAC-seq tracks (normalised ATAC-seq signal) in: HY; hyperdiploid, MS; MMSET, MF; MAF, CD1; CCND1, ND; normal donor PC, MAF OE; MAF overexpression in U266 vs empty vector, MM.1S MAF ChIP; ChIP-seq against MAF in MM.1S with underlying bars indicating peaks with FC over input q<0.01, Chromatin states (Blueprint consortium data[115]) in normal tPC; normal tonsillar PC, mB cell; class-switched memory B cells, GC B cell; germinal centre B cells. MFB; MAFB, CD3; CCND3.

4.6 Identification and characterisation of the *CCND2* enhancer

Integrating the MAF cistrome with the differentially open regions found in MAF primary samples maps the epigenome influenced by MAF. By juxtaposing this against the chromatin landscape of normal PC and B cells and validating a significant proportion of those regions by confirming enhanced accessibility following MAF overexpression identifies directly MAF-activated regulatory regions, previously inactive in the normal cellular counterpart. Many of those regions are found uniquely in MAF MM. However, the combination of the MAF cistrome with primary MAF sample accessibility can also identify regions where MAF may still have a regulatory role to play, even if these are not differentially accessible from normal plasma cells. One such example is the *CCND2* enhancer.

Cyclin D2 (CCND2) is a protein involved in cell cycle regulation and cell proliferation and its depletion results in cell cycle arrest and cell death in MM cells[156]. Previous studies have shown *CCND2* downregulation upon MAF knockdown and have reported the presence of the MAF motif in the *CCND2* promoter[45]. *CCND2* has also been found to be upregulated specifically in the MF MM subgroup in GEP studies[68]. Indeed, my earlier results confirmed *CCND2* to be a direct MAF target (section 3.6).

As part of a larger collaborative project comparing overall chromatin accessibility in primary MM and normal PC, putative enhancer peaks were identified upstream of the *CCND2* gene[100]. Overall accessibility upstream of *CCND2* combined with *CCND2* expression separated the CCND1-translocated MM samples from the other subgroups[100]. As *CCND2* is one of the main targets of MAF and given that cyclin D proteins are dysregulated in virtually all MM, subclassifying it in CCND1-, CCND2- or CCND3-expressing[157], I explored further the role of these enhancer peaks and MAF binding in *CCND2* expression, in the context of MAF MM.

Inspecting the MAF binding sites, revealed multiple MAF-bound peaks upstream of the gene, in addition to binding of the promoter. Indeed, these enhancer peaks were found to be

accessible in the primary MAF samples (**Fig. 4.22A**). In addition, juxtaposing the MAF-bound peaks against chromatin marks in MM.1S (described in section 2.11.3) revealed them to be active enhancers (**Fig. 4.22A**). Only one of the peaks displayed differential accessibility compared to normal PC (peak 2), whereas other MAF-bound accessible peaks did not display enhanced accessibility in comparison to normal PC (peak 1 and 3). Finally, these enhancer peaks were also accessible in the MMSET and hyperdiploid MM subgroups, despite expression of *CCND2* being lower in these groups (**Fig. 4.23**).

In order to investigate if the upregulation of *CCND2* expression in the MAF subgroup was regulated by MAF-bound enhancer peaks, in addition to direct binding to the promoter, I investigated the effect of these peaks on *CCND2* expression using CRISPRi repression of peaks 1, 2, 3 and 4 (the latter not bound by MAF). Downregulation of *CCND2* expression was seen upon inhibition of the most distal peak 1 and most proximal peak 4, with little effect seen when targeting the other peaks (**Fig. 4.22B**), confirming the former as regulatory regions of *CCND2*. In fact, the effect seen on *CCND2* expression when targeting those regions was similar if not more pronounced than the downregulation caused by the *CCND2* promoter guides. Taken together, these results validated the presence of a *CCND2* enhancer in multiple myeloma.

Interestingly, one of the peaks with the most pronounced regulatory effect on *CCND2*, the most distal peak 1, was also the one with the most avid MAF binding, whereas the peaks with little effect on *CCND2* expression as well as the promoter itself displayed much weaker binding of the transcription factor. These results suggest that MAF binds avidly to specific peaks of the *CCND2* enhancer and upregulates *CCND2* expression over and above that seen in other MM samples (**Fig. 4.23**).

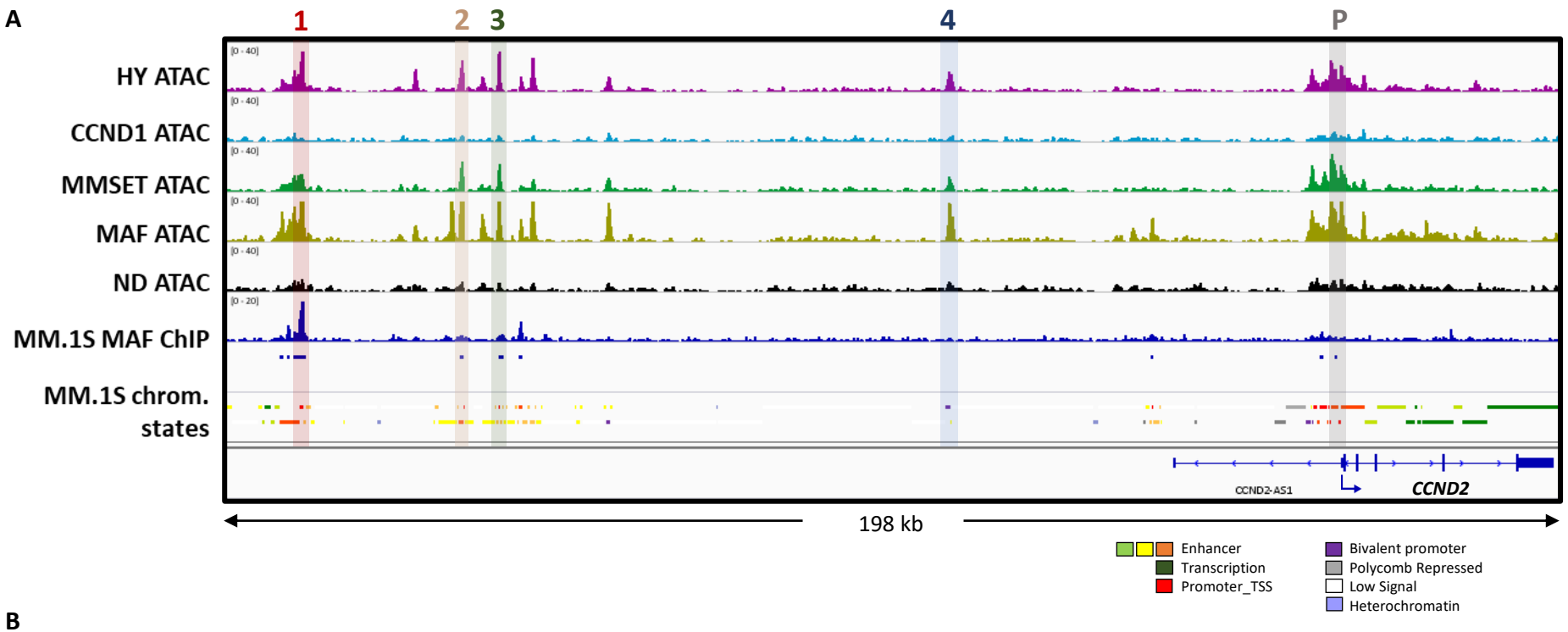


Figure 4.22: *CCND2* enhancer

(A) IGV plot depicting accessibility peaks (average normalised ATAC-seq signal) of putative *CCND2* enhancer across MM subgroups, as identified as part of a larger collaborative work[100], as well as ChIP against MAF and respective chromatin states **(B)** *CCND2* expression, as measured by qPCR, following CRISPRi targeting of each of the peaks 1, 2, 3, 4 and promoter peak (P). Expression normalised to *GAPDH* and gal4 control. Bars show mean + SD of n=3 independent experiments. *p<0.05, ***p<0.001, ****p<0.0001, 1-way ANOVA with Dunnett's post-hoc multiple comparisons correction. MM.1S MAF ChIP; ChIP-seq against MAF in MM.1S (FC over input, with bars indicating peaks q<0.01), HY; hyperdiploid, ND, normal donor bone marrow PC, MM.1S chrom. states; Chromatin states in MM.1S (ENCODE data[117]).

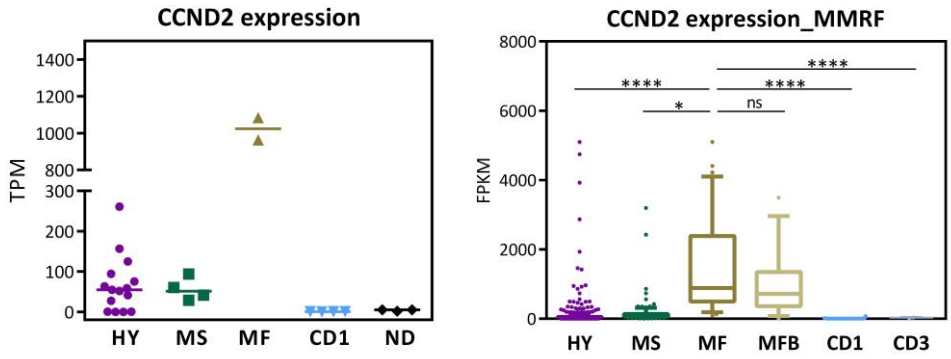


Figure 4.23: *CCND2* expression across myeloma subgroups

CCND2 expression across MM subgroups in our in-house data set of primary samples (left, in TPM) and primary MM samples from MMRF Compass data[121] (right, in FPKM). Statistical analysis performed using a 1-way non-parametric ANOVA test with Dunn's post-hoc multiple comparisons correction. *p<0.05, ****p<0.0001, ns; non-significant, HY; hyperdiploid, MS; MMSET, MF; MAF, MFB; MAFB, CD1; CCND1, CD3; CCND3, ND; normal donor PC.

4.7 Discussion

In addition to genetic changes, cancers are often characterised by extensive alterations in their epigenetic and regulatory genomic landscape[146]. ATAC-seq is a technique requiring a small number of cells, thus allowing the mapping of chromatin accessibility in primary normal and malignant cells which cannot be collected in large numbers. ATAC-seq data have been used in solid organ cancers, as well as haematological malignancies such as CLL, to show that chromatin accessibility alone can cluster samples in different malignant subgroups[147, 148]. In CLL, integration of chromatin accessibility with histone marks and comparison with the normal B cell counterpart allowed the identification of superenhancers, critical in the regulation of the malignant transcriptional programme[148].

In the case of MM, one previous study described the epigenetic changes seen in MM as compared to PC and demonstrated decondensation of the MM heterochromatin[144]. The authors used H3K27ac marks in primary MM plasma cells obtained from 29 treatment-naïve MM patients and from normal PC to define differentially active enhancers in MM, such as for example enhancers located upstream or on the gene bodies of *MCL1*, *CXCR4* and *TET1*. They then compared the differential ATAC-seq signal in those samples with chromatin states in GM12878 cells (EBV-transformed B cell line) and H3K27ac marks in primary myeloma plasma cells observing that 25% of active enhancers in MM arise from heterochromatic / low signal B cell regions. Crucially however, normal plasma cells used as the comparator in this study were *in vitro* differentiated from splenic B cells and the primary MM sample dataset did not contain any MAF-translocated samples.

Having established during my earlier investigations of the MAF cistrome that MAF preferentially binds intergenic and intronic regions, suggesting a predilection for enhancers, I proceeded to investigate the epigenetic alterations induced by MAF-translocation in MM, a primary genetic event in the process of myelomagenesis.

4.7.1 Chromatin accessibility profile in MAF-translocated MM

First, I compared chromatin accessibility between 2 MAF samples and 3 normal bone marrow PC, to find an overall enhanced accessibility profile in the MAF samples, which was associated with gene expression upregulation. Although my sample size is small, genes annotated to the differentially accessible MAF peaks were significantly enriched in genes associated with the MF subgroup in GEP studies, suggesting my findings are representative of MAF myeloma as a whole. Enhanced accessibility was seen in all myeloma genetic subgroups, compared to normal PC but this was more pronounced in the MAF subgroup than any other group. A third of the total differentially accessible peaks were unique to MAF MM, almost triple the number found in the next highest subgroup. Indeed, therefore the chromatin accessibility profile of the MAF samples is distinctly different, reflected in its ability to cluster MAF myeloma separately to the other MM subgroups.

Keeping in mind the differences and similarities in chromatin accessibility between MAF MM and the other subgroups, I elected to use all differentially accessible peaks in MAF myeloma for further analysis, rather than focus on the MAF-unique regions. The reason for that was two-fold. First, peaks shared amongst MAF and other subgroups may still be directly linked to the MAF transcription factor in the case of MAF MM, as distinct pathways may have led to their differential accessibility in other subgroups. Second, MAF expression is also found to a lesser extent in the MMSET group, as well as occasionally in some hyperdiploid myelomas, making it harder to separate the MAF from non MAF-linked peaks in those groups. This would have therefore led to the exclusion of regions that are MAF-associated both in MAF and MMSET or even hyperdiploid myeloma.

In order therefore to ensure a direct link between the differentially enhanced peaks in MAF myeloma and the MAF transcription factor, I integrated chromatin accessibility with the MAF cistrome in MM.1S. MM.1S was used in preference to the MAF cistrome obtained in JJN3, due to the extensive epigenetic characterisation of MM.1S publicly available. A quarter of the differentially accessible MAF peaks were MAF-bound, identifying the group of “MAF-associated” regions. The majority of the differential chromatin accessibility peaks in MAF samples were located in intronic and intergenic regions. The same was true for the “MAF-

associated" regions; 85% were identified as enhancers by genomic annotation. Epigenomic annotation, using histone mark maps of MM.1S, assigned a significant proportion of those to promoters; i.e. some of those regions displayed chromatin marks more in keeping with promoter activation despite residing outside an annotated transcription start site, as had been determined by genomic annotation. However, 35% of "MAF associated" regions were identified as enhancers by epigenomic annotation as well. These results are similar to those observed in other cancers, where differences in chromatin accessibility between various cancer types centre around distal regulatory elements rather than promoters, where chromatin accessibility was less distinguishing[147]. Finally, motif analysis of the "MAF-associated" regions revealed they are also bound by members of the IRF family of transcription factors, all of which share very similar motifs, similar to what was seen previously in the overall MAF cistrome in MM.1S, JJN3 as well as H929 MMCLs.

4.7.2 MAF as an activator of specific regulatory regions

Overlaying the "MAF-associated" regions with the chromatin mark maps across development in normal B cells and tonsillar PC, where MAF is not expressed, revealed that 30-40% of these regions were inactive during the development of the normal PC. This suggests these regions became activated during the process of myelomagenesis. A similar conclusion was reached in a comprehensive study of the CLL epigenome[149]. The authors, by comparing ATAC-seq and H3K27ac ChIP-seq signals between primary CLL and normal B cells, identified a group of regions displaying differential activity in the malignant cells. Most of these regions were found to display low or heterochromatic signals during the stages of B cell development, suggesting the presence of *de novo* activated enhancers in CLL. In the study of the myeloma enhanceome described above, approximately 25% of regions found accessible in primary non-MAF MM samples, were associated with heterochromatic marks / low signal in GM12878 B cells[144]. However, regions that were already active in the primary memory B cells and in the *in vitro* developed PC were not excluded, therefore a proportion of these peaks may have become activated during normal development of a B cell to a PC. Given these findings are similar to those observed in the "MAF-associated" regions, my findings could have reflected the differences of MM with normal PC in general, and a similar pattern may have been

observed in other MM subgroups. However, comparison of the "MAF-associated" regions with U266 ChromHMM states revealed that a similar proportion of the "MAF-associated" regions were also inactive in the t(11;14) MMCL. This, in conjunction with the observation that 37% of these regions are uniquely differentially accessible in MAF myeloma when compared to other subgroups, suggests that although the overall pattern of enhanced activation over the normal cellular counterpart can be found across MM, the regions described here are characteristic of the MAF subgroup.

Using the above analysis, I identified 726 *de novo* activated enhancers and epigenetically designated promoters in the MAF samples, previously inactive in the normal PC and in the majority of cases throughout B cell development. Indeed, of these putative "MAF-activated" regions just over half (371/726) were unique to the MAF subgroup, suggesting the *de novo* activation of these regions is not generic to MM as a whole. A similar number of regions, *de novo* activated in CLL compared to the normal B cell developmental stages were found in the aforementioned study: 498 out of 7,121 regions with differential activity between CLL and normal B cells, displayed low or heterochromatic signal in the normal cellular counterpart[149].

All of the "MAF-activated" regions were MAF-bound, however, in order to experimentally determine which of these putative "MAF-activated" regions are directly activated by the transcription factor, I overexpressed MAF in a non-MAF expressing MMCL, the U266. Upon MAF overexpression, 53% (382/726) of the putative "MAF-activated" regions acquired enhanced accessibility (53% of "MAF-activated" promoters and 52% of "MAF-activated" enhancers), validating these regions as directly activated by the transcription factor in this setting. Interestingly, this included regions proximal to some of the known MAF associated genes, such as *CCR1*, *CX3CR1* and *NUAK1*, also shown to be direct MAF target genes from my earlier experiments. In order to confirm these findings, and further support the hypothesis of MAF as a direct activator of these regulatory regions, future experiments would involve ChIP-seq against chromatin marks such as H3K27ac, H3K4me1 and H3K4me3 in the MAF overexpressing cells, with the EV U266 as a control, to confirm an increased signal in these marks following MAF-overexpression. In addition, chromatin immunoprecipitation against MAF in the same cells would indeed confirm direct binding of the TF upon overexpression, to

confirm binding in the same regions as identified in MM.1S. These experiments are currently under way.

Overall, my data suggest that MAF is associated with specific aberrant regulatory region activation, 30-40% of which appear to acquire *de novo* activation compared to normal PC. A significant proportion of those regulatory regions appear to be directly activated by MAF, and my results to date support the hypothesis of MAF as an activator of a specific subset of regulatory regions during myelomagenesis. Similar findings to the above, though in context of normal constitutive expression of MAF, rather than aberrant overexpression, were described in macrophages[158]. Here the authors stimulated resting macrophages with IFN-γ, and concomitantly with gene expression downregulation, observed disassembly of the assigned enhancers as evidenced by a reduction in H3K27ac and ATAC-seq signals. This however was not observed in the corresponding promoters of the downregulated genes. MAF downregulation as well as enrichment of the MAF motif in the "closed" upon IFN-γ stimulation enhancers suggested the involvement of MAF in this process. Indeed ChIP-seq before and after macrophage stimulation, revealed 53% of the 642 disassembled enhancers lost MAF binding upon IFN-γ exposure. These findings were recapitulated upon MAF-knockdown, suggesting a direct role of the transcription factor in this process. The authors concluded that MAF functions to maintain chromatin accessibility in a specific group of enhancers regulating signature genes in the resting macrophages, and that suppression of the transcription factor leads to inactivation and loss of chromatin accessibility in these enhancers.

4.7.3 The example of the *CCR1* enhancer, a validated "MAF-activated" region

CCR1 is a chemokine receptor, previously shown to be involved in MM cell adhesion, proliferation and chemotaxis, facilitating the growth and dissemination of MM cells[80, 81, 84]. As a proof of concept, given the known role of CCR1 in MM, I elected to validate this directly MAF-activated peak, as an enhancer of *CCR1*. Inspection of the ATAC-seq data identified 3 peaks downstream of the *CCR1* gene representing putative enhancers of *CCR1*. One of these peaks had been identified as a "MAF-activated" peak (MAF-bound, over-accessible in MAF vs

normal PC and inactive in normal PC, whilst active in MAF MM) and was validated as acquiring enhanced accessibility upon MAF overexpression.

In order to assign these peaks as enhancers of *CCR1*, I used a CRISPRi method to repress the activity of each individual peak. This method has previously been used for the functional annotation of enhancers, such as for example in the case of a *BCL2* enhancer found in a breast cancer cell line, where the authors observed a 20-50% reduction in *BCL2* expression[147]. Indeed, CRISPRi experiments validated the "MAF-activated" region as the peak with the most significant effect on *CCR1* expression, akin to the effect seen when targeting the promoter region, whereas a much smaller, if at all, effect was seen upon targeting the other two peaks. Using this example, I demonstrate the direct activating capacity of MAF in a subset of enhancers that were previously found to be silent in the normal plasma cell. Future experiments involving chromatin immunoprecipitation of activatory histone marks and MAF itself, as described above, would confirm MAF binding in this region and activation upon MAF overexpression. The direct role of MAF binding on the *CCR1* enhancer could also be confirmed by mutagenising the MAF binding site located in this peak using CRISPR and observing the effect of this on *CCR1* expression. A complementary future experiment would be to perform CRISPRi repression in the MAF-overexpressing U266 and investigate whether this would abrogate the enhanced expression of the target genes. Given the inducibility of the CRISPRi system used, the same MAF-overexpressing cells could be used as their own control by measuring *CCR1* expression before and after induction of the CRISPRi vector. Finally, Capture-C experiments[159] in MAF-translocated cell lines or indeed t(14;16) primary MM samples would confirm interactions of the gene promoter with enhancer peaks.

4.7.4 Limitations of using a myeloma cell line

Overall, in this analysis I have focused on the regions activated by MAF in the context of being inactive in normal PC, in order to determine if MAF may act as an activator of regulatory regions during myelomagenesis. A schematic of this analysis is shown in **Fig. 4.24**. In order to determine which of the putative "MAF-activated" regions are indeed directly activated by MAF, I have ectopically expressed MAF in a t(11;14) MM cell line. However, U266 do not

represent normal PC and using an established myeloma background to investigate the effects ectopic MAF may have once overexpressed in a normal cell, introduces possible confounders.

First, a t(11;14) MM cell may not possess a permissive background for the effects of MAF on the regulatory genome to be fully appreciable, especially given that in MM, MAF overexpression even as a secondary event is not observed in conjunction with t(11;14). Indeed, not all of the putative "MAF-activated" enhancers were validated as directly MAF-activated in my experiments. The regions not "opened" by MAF mostly demonstrated no change in accessibility following MAF overexpression, and a very small proportion displayed reduced accessibility. One such example, was the regulatory region distal to *TLR4*, another of the MAF target genes characteristically overexpressed in MAF myeloma. This peak is MAF-bound and previously inactive in normal PC, and although, using CRISPRi it was validated as a regulatory region of *TLR4* expression, it did not demonstrate enhanced accessibility upon MAF overexpression. This could be due to inhibitory factors specifically present in U266, a CCND1-translocated MMCL. Another explanation would be that in these regions, MAF alone is not sufficient for their activation, and perhaps requires the presence of other regulatory factors not present in U266 or additional stimulation, for example by the bone marrow microenvironment. This could be the case especially given that, as demonstrated in my earlier experiments, many of the MAF target genes are involved in cell to cell interactions.

Second, although MAF ectopic expression in U266 lead to activation of over 50% of putative "MAF-activated" regulatory regions, this does not necessitate the same would be observed in the cellular environment of a normal PC, which could have a less permissive environment that an established malignant cell. In order to overcome possible inhibitory environments imposed by a different primary translocation, experiments could also include MAF overexpression in a non-primary translocated MMCL, such as the AMO-1, which are currently under way. However, in order to more faithfully recapitulate the events during MAF-translocation and demonstrate directly the activatory role MAF has in these regions, overexpression experiments would need to be conducted in normal PC. In practice, the feasibility of such an endeavour is challenging, given the difficulty in obtaining normal PC, especially in sufficient numbers for such experiments, and the difficulties in transducing them.

As an alternative, future experiments could involve MAF overexpression in *in vitro* derived PC from peripheral blood B cells, which are much more easily obtainable.

4.7.5 Possible mechanisms underpinning MAF-induced epigenetic activation

The mechanism by which MAF activates regulatory regions is unclear and requires further investigation. One possibility involves PC-specific priming of these regions, following which ectopic MAF binds to and activates them. This theory has previously been proposed by Sur and Taipale, where a transcription factor specific to a cell type renders the region accessible to an oncogenic transcription factor which, when present, activates the enhancer and induces gene expression[160]. This could be possible in the case of MAF myelomagenesis. However, although the "MAF-associated" regions were derived from differential accessibility between MAF MM and normal PC, the majority of these regions displayed minimal accessibility in normal PC, so pre-existing priming of these regions would need to occur with minimal detectable accessibility.

The second mechanism of epigenetic activation induced by MAF is via pioneering action in these regions. Pioneer transcription factors initiate chromatin accessibility by their unique ability to bind nucleosomal DNA, prior to other transcription factors, rendering it accessible to subsequent binding of other factors [161]. To establish whether a transcription factor has pioneering action two criteria have to be met [161]. First, the pioneer factor needs to bind first, prior to the binding of other transcription factors, to initiate chromatin reorganisation. Second, pioneers access previously inaccessible DNA regions, through their ability to bind nucleosomes. In the context of a pioneer transcription factor, previously "closed" chromatin, when the transcription factor is not expressed, becomes *de novo* accessible following transcription factor expression and binding [162].

Although one could envisage MAF acting as a pioneer transcription factor in MM based on the above findings, the data is not sufficient to support or refute this hypothesis and further experiments are required. In the case of the validated "MAF-activated" regulatory regions, *de novo* MAF expression has been shown to increase accessibility in MAF-expressing U266,

where MAF was not expressed before. These regions were previously inactive in the normal PC, as shown by the chromatin states. Not many of these regions were previously heterochromatic, though heterochromatin is thought to be inaccessible even to pioneer transcription factors [162]. However, direct MAF binding has been extrapolated from the MAF cistrome in the MM.1S cells, and has not been directly demonstrated following ectopic MAF overexpression. In addition, although these regions were inactive in normal PC, in order to demonstrate pioneering action, MAF would need to be shown to directly bind and increase accessibility in inactive genomic regions. One way to accomplish that, would be to overexpress MAF in normal PC and investigate accessibility changes in areas already demonstrated to be inactive by chromatin marks. However, as previously discussed, such experiments would be challenging. Pioneering action could also be demonstrated in the context of MAF overexpressing U266 with some further experiments. Integrating my above results with the U266 chromatin marks available from ENCODE suggests that 71% of the validated "MAF-activated" regions indeed didn't display chromatin activation marks in U266. In order to demonstrate pioneering action in this context, chromatin immunoprecipitation experiments against MAF as well as chromatin marks such as H3K27ac and H3K4me1 before and after MAF expression, would demonstrate direct MAF binding of these regions and further validate the activation of these enhancers following MAF expression. In addition, low or high throughput MNAse assays in U266 before and after MAF overexpression would investigate the possibility of nucleosomal occupancy before MAF expression and MAF's ability to displace nucleosomes following its expression. As an example, such a study was performed demonstrating PU.1 as a pioneer factor in myeloid precursors using a PU.1 inducible system [111]. The authors performed ChIP-seq against PU.1 and H3K4me1 to demonstrate that in subset of enhancers, following PU.1 induction, H3K4me1 was gained upon PU.1 binding. Subsequently, MNAse-seq was used to demonstrate depletion of nucleosomal occupancy upon PU.1 induction in the same sites. This data was used to establish PU.1 as a pioneer factor in myeloid precursors [162]. These experiments are currently under way in the U266 MAF-expressing system.

4.7.6 Other MAF regulatory regions – the *CCND2* enhancer

As part of a larger collaborative project, analysis of the accessibility profile of primary MM cells, allowed the identification of a putative MM enhancer upstream of *CCND2*, present in all MM subgroups apart from CCND1-translocated myeloma[100]. However, CCND2 is selectively upregulated in MAF myeloma and has been validated as a MAF target both in my datasets as well as previous studies[45]. As a separate analysis therefore, I investigated the role of this putative enhancer on *CCND2* expression, focusing on MAF-bound peaks, in order to determine the effect of MAF binding.

Integration of the ATAC-seq data with the MAF cistrome, allowed identification of MAF-bound peaks, with a potential role in augmenting *CCND2* expression. Indeed, following CRISPRi repression, one of the peaks with the largest effect on *CCND2* expression was the one most avidly bound by MAF. In fact, targeting this peak produced a more pronounced reduction in *CCND2* expression than targeting the gene's promoter region. However, in addition to being MAF-bound and accessible in MAF MM, this peak demonstrated accessibility in other non-MAF MM subgroups as well as normal PC. Therefore, avid MAF binding to one of the most influential *CCND2* enhancer peaks, in addition to binding of the promoter, might explain the selective upregulation of *CCND2* in MAF MM compared to other MM subgroups and normal PC.

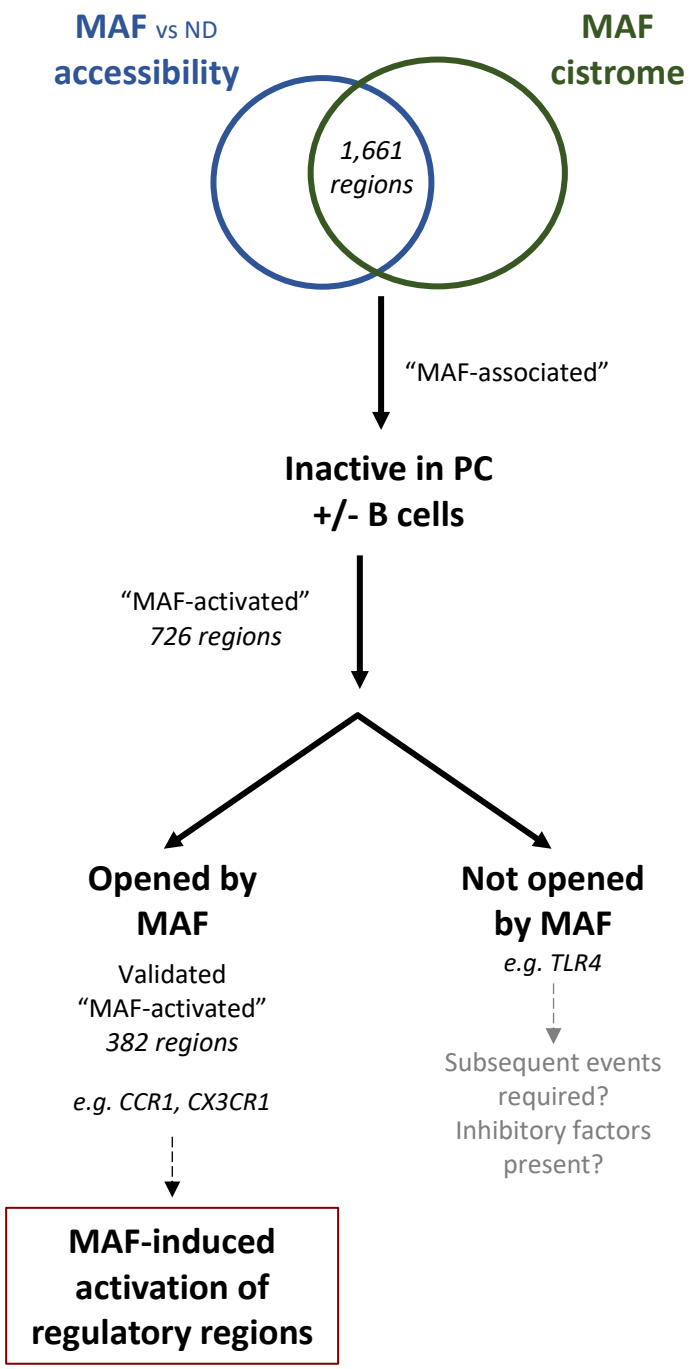


Figure 4.24: Schematic of epigenomic analysis and proposed MAF activatory function

Schematic depicting the analysis employed in this chapter to support the role of MAF as an activatory of specific regulatory regions. ND; normal plasma cells, PC; plasma cells

Chapter 5

MAF cooperating factors

5.1 Introduction

Transcription factors (TF) cooperate to regulate gene expression and their interactions can modify the specificity and selectivity of DNA target binding. TF interactions are broadly classified into two categories[163]. The first involves direct protein to protein interactions, in the form of homo- or heterodimerisation. This can occur either spontaneously in the absence of DNA binding or, more commonly, proteins may require DNA binding in order to dimerise. In this case, the presence of the specific DNA sequence enables the interaction between TF, for example by inducing a conformational change or mediating physical proximity between the interacting proteins[163]. However, only a minority of TF partnerships are secondary to direct homo- or heterodimerisation, with only 5% of interactions predicted to be mediated by direct protein to protein binding[164]. The second mechanism is entirely dependent on DNA binding and involves no direct protein to protein interaction, rather TF co-bind a DNA segment, which acts as a platform for their cooperative function. This would be for example the case when binding of one TF induces conformational changes to the DNA structure allowing the binding of other factors[163].

TF of the bZIP family, of which MAF is a member of, form dimers in solution via their leucine zipper domain[163]. MAF has been shown to bind the half MAF recognition element (MARE) in DNA as a monomer[32], but is also able to bind the MARE as a homodimer and a heterodimer *in vitro*, for example together with JUN and FOS TF[28, 30, 31, 33, 165], with mutations to the leucine zipper domain disrupting these interactions[33, 165]. In addition, heterodimers have been shown to have different specificity for DNA binding to homodimers[33]. There is however, overall, limited data on potential partners of MAF, with none having been described in multiple myeloma.

Using protein microarrays of bZIP TF, Newman and Keating were able to demonstrate *in vitro* interaction of MAF with ATF-4, BACH1, MAFB and homodimerisation with itself, with the latter two displaying the strongest interactions[36]. Indeed, MAF was shown to cooperate with MAFB during keratinocyte differentiation, when double knockdown of both proteins produced a much more pronounced downregulation of keratinocyte differentiation genes

than knockdown of each TF separately, as well as preventing upregulation of maturation markers in tissue[42]. Just over 1,900 genes were found to be bound by both TF in differentiated keratinocytes and ChIP-re-ChIP assays confirmed co-binding in select loci. The authors concluded that the two TFs cooperate, and via physical proximity regulate genes for keratinocyte differentiation.

c-MYB is another TF found to interact with MAF, with some studies demonstrating direct physical association between the two. MAF and c-MYB physically interact in the later stages of myeloid development to regulate transcription of the gene encoding the surface glycoprotein CD13/APN[35]. c-MYB activates expression of *CD13/APN*, but increasing levels of MAF in more mature myeloid cells reduces the ability of c-MYB to transactivate the gene's promoter, ultimately reducing *CD13/APN* expression. MAF alone was unable to bind to the *CD13/APN* promoter. However, *in vitro* assays demonstrated direct binding of c-MYB to the MAF protein, and the complex, detected in lysates of maturing myeloid cells, was then able to bind the gene's promoter. Cooperation of MAF with c-MYB has also been shown in CD4+ T cells[34], where MAF forms a complex with c-MYB in stimulated T cells, preventing c-MYB binding and transactivation of the *BCL2* promoter, with a consequent reduction in *BCL2* expression.

Having described the transcriptome and epigenetic changes imposed by MAF in t(14;16) MM, I proceeded to investigate potential TF partners and chromatin binding proteins which may be cooperating with MAF in this process.

5.2 IRF4 co-occupancy

Inspecting the MAF ChIP-seq data obtained in myeloma cell lines (described in section 3.5) revealed the IRF4 motif as the second most enriched in both MAF-translocated cell lines MM.1S and JJN3 (**Fig. 5.1**). IRF4 is key TF in plasma cell development, whose upregulation drives the transcriptional programme necessary to promote plasma cell (PC) differentiation from B cells[15, 166], and is essential for normal plasma cell as well as myeloma cell survival[166,

[167]. Given that IRF4 has been shown to interact with other transcription factors, including members of the bZIP TF family[15, 168, 169], it constituted a good candidate MAF partner to investigate further.

Comparing the MAF cistrome with that of IRF4 in MM.1S cells, obtained from the Young Lab, (GEO accession number GSM1195560), revealed a total of 13,398 sites co-bound by both transcription factors (**Fig.5.2**). MAF and IRF4 were considered co-bound if the binding sites contained at least one overlapping base pair. Impressively, this suggested IRF4 co-binds 72% of the total MAF binding sites (**Fig. 5.2**). This included key MAF transcription targets, such as *ITGB7*, *CCR1* and *TLR4*, found to be direct MAF targets in my earlier experiments in MAF-translocated cell lines and are also bound by MAF in the MAF-overexpressing cell line H929 (**Fig.5.3**). IRF4 binding was also confirmed by in-house ChIP-qPCR in MM.1S cells (**Fig. 5.3**). Focusing on the "MAF-associated" regions, which are both MAF-bound and display enhanced accessibility in MAF primary cells compared to normal PC, IRF4 co-binding was determined when there was at least one bp overlap between the IRF4 and ATAC peaks. The majority of the "MAF-associated" regions, 1,430/1,661 (86%) were found to be bound by IRF4 (**Fig. 5.4**). Of the "MAF-associated" regions that were inactive in normal PC and B cells, i.e. the "MAF-activated" regions, 87% (630/726) were IRF4-bound. Approximately the same percentage (88%, 337/382) was found to be IRF4-bound in the validated "MAF-activated" regions, i.e. regions that acquired enhanced accessibility following ectopic MAF expression in U266 cells (**Fig. 5.4**). For example, IRF4 binding was seen in both the *CCR1* and *TLR4* regulatory peaks (**Fig. 5.5, 5.6**). Closer inspection of IRF4 binding on the *CCR1* enhancer peaks revealed that the two peaks seen to have an effect on *CCR1* expression upon CRISPRi targeting (Peak 1 and 2) were both co-bound by the two TF (**Fig. 5.5**). By contrast, Peak 3, which had no effect on *CCR1* expression, was not found to be significantly bound by IRF4 (**Fig. 5.5**). However, peak 1 was not significantly differentially accessible between t(14;16) primary samples and normal plasma cells, indicating that MAF here may act as a transcriptional regulator but does not modify chromatin accessibility. Taken together, these findings suggest that IRF4, cooperates with MAF globally and does not specifically target regions where MAF enhances chromatin accessibility.



IRF4 motif

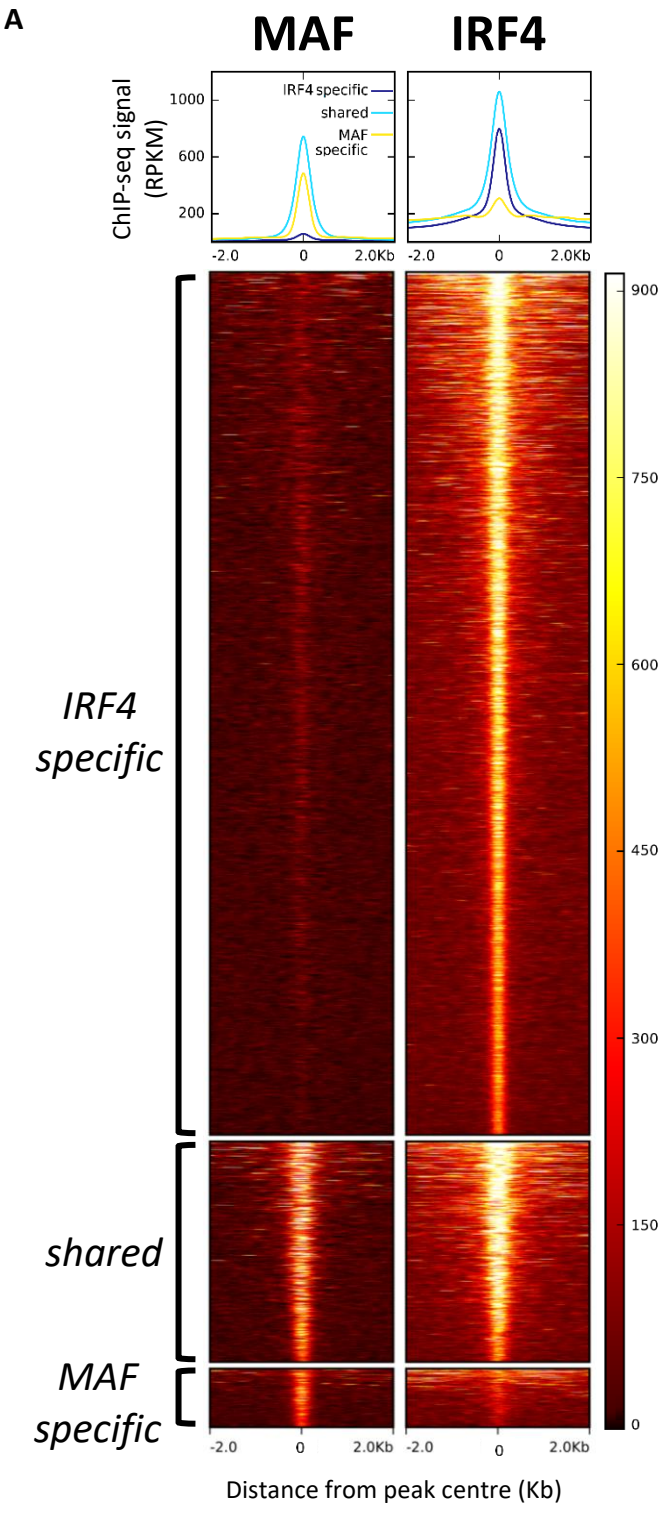
MM.1S P value ($-\log_{10}$) 702
JJN3 P value ($-\log_{10}$) 342

Figure 5.1: IRF4 motif the second most enriched motif in MAF translocated cells

Weblogo of the IRF4 motif, indicating the significance of enrichment ($-\log_{10}$P value) obtained during motif analysis of ChIP-seq against MAF in MM.1S and JJN3 myeloma cells.

To further delineate the co-localisation of the two transcription factors, I explored the distance between the binding of MAF and IRF4, by subtracting the summit of the closest IRF4 binding peak from the summit of the MAF binding site, focusing on co-bound sites, both genome wide and in the 1,661 "MAF-associated" regions (**Fig. 5.7**). In both cases, the median absolute distance between the binding of the two TF was the same; namely 74bp in the former and 75bp in the latter. Within the limits of ChIP-seq resolution, this indicates that the binding sites of the two transcription factors are located very close to one another though the TFs may not be linearly directly adjacent.

ChIP re-ChIP qPCR experiments were performed in order to experimentally validate the co-binding of the two TF, focusing at specific loci of interest. Co-binding of the two TF was seen in both directions, IRF4>MAF and MAF>IRF4, as evidenced by enrichment in the *ITGB7* promoter, *TLR4* promoter and Peak 2 of the *CCR1* enhancer (**Fig. 5.8**).



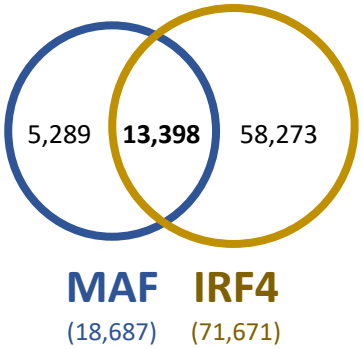


Figure 5.2: MAF and IRF4 cistrome

(A) Heatmap and histogram of MAF and IRF4 ChIP-seq in MM.1S myeloma cells, indicating peaks shared and specific to each TF. ChIP-seq peak intensity in RPKM, distance from peak centre on x axis (Kb). IRF4 ChIP-seq available from the Young Lab (GSM1195560). **(B)** Venn diagram of number of binding sites for each TF, indicating number of shared peaks between MAF and IRF4.

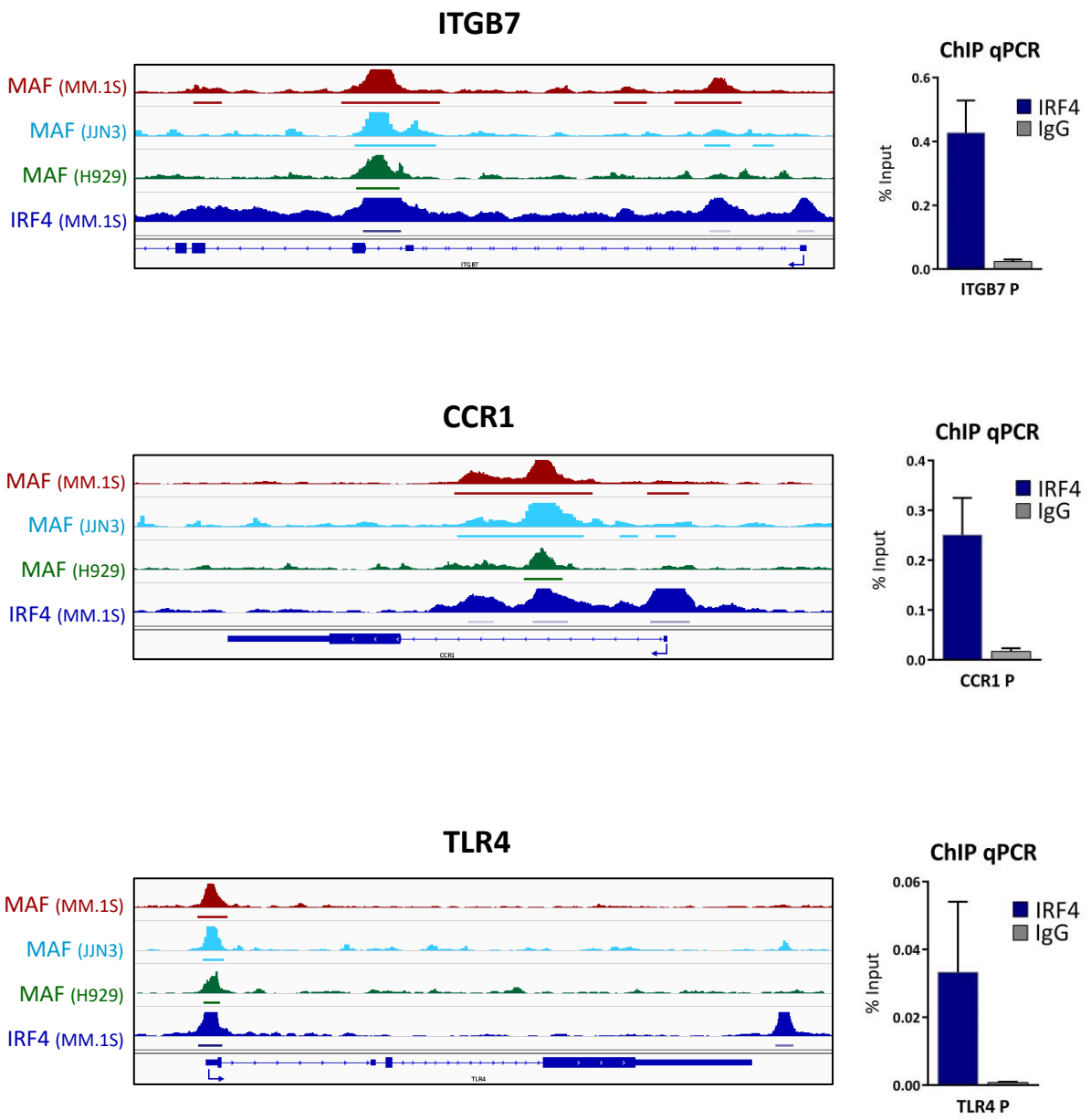


Figure 5.3: IRF4 binding on key MAF target genes

IGV plots of key MAF target genes showing binding of MAF in MM.1S (red), JJN3 (light blue) and H292 (green) as well as IRF4 in MM.1S (dark blue), as obtained from the Young Lab (GSM1195560). MAF ChIP-seq tracks represent FC over input, with bars indicating peaks q<0.01. IRF4 ChIP-seq tracks represent normalised ChIP-seq signal (bars showing peaks q<0.01). On the right, in-house ChIP-qPCR against IRF4 of the same target genes in MM.1S cells, against IgG control, expressed as percent input DNA. Bars represent mean + SEM of n=4 experiments.

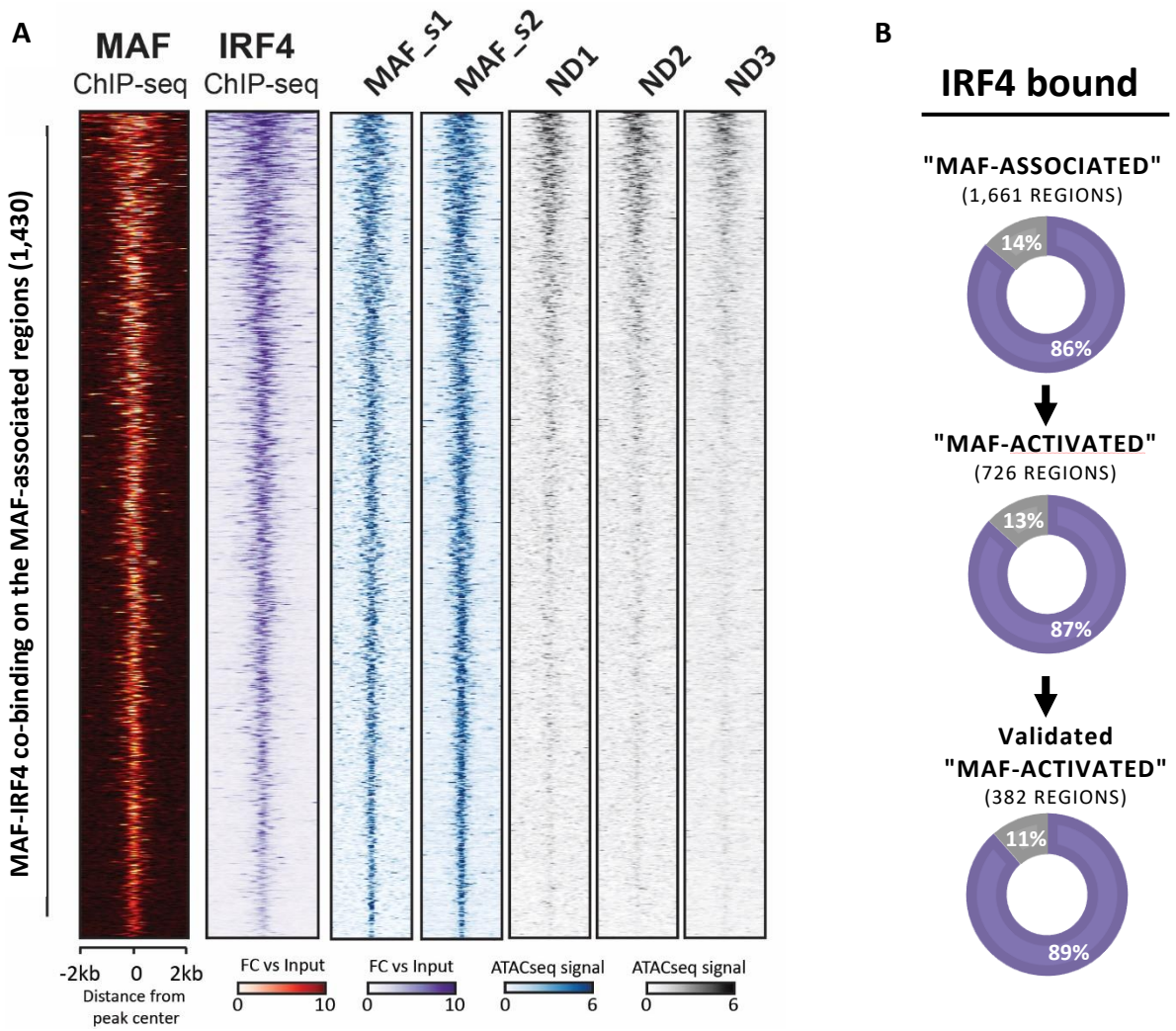


Figure 5.4: IRF4 binding in MAF-associated regions

(A) Heatmap illustrating the "MAF-associated" regions found to be co-bound by IRF4. IRF4 ChIP-seq data in MM.1S obtained from the Young Lab (GSM1195560), and illustrated in purple (IRF4 ChIP-seq). In red, MAF binding on the MAF-associated regions in MM.1S (MAF ChIP-seq). In blue, the ATAC-seq peaks obtained from the two primary MAF myeloma samples (MAF_s1, MAF_s2), compared to the peaks found in normal plasma cells (ND1, ND2, ND3, in grey). **(B)** Pie charts depicting percent of "MAF-associated", "MAF-activated" and the validated by overexpression experiments "MAF-activated" regions co-bound by IRF4 (purple).

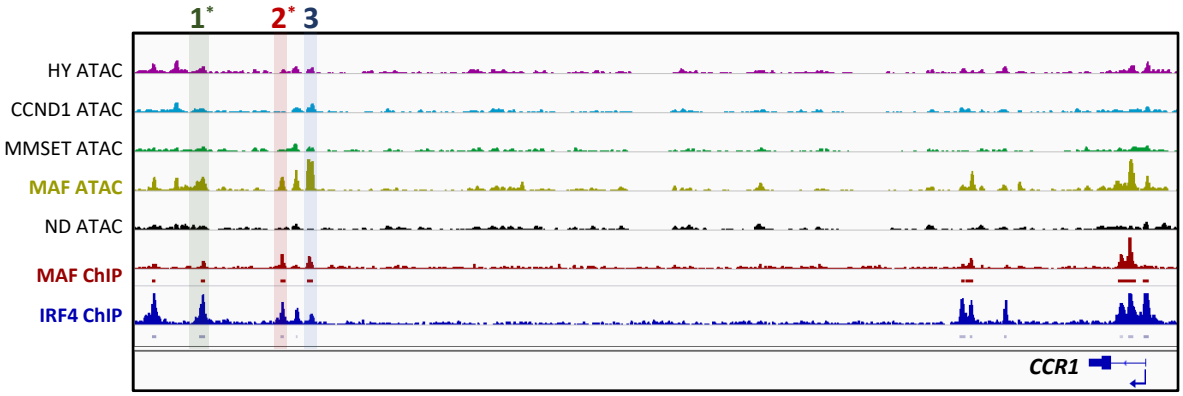


Figure 5.5: Binding of IRF4 on *CCR1* enhancer

IGV plot depicting MAF and IRF4 binding, as identified by ChIP-seq in MM.1S, on *CCR1* enhancer peaks. Highlighted are the enhancer peaks found in primary MAF samples following ATAC-seq (average normalised ATAC-seq signal, MAF ATAC). Stars next to numbered peaks indicate peaks shown to have an effect on *CCR1* expression by CRISPRi. For comparison, peaks obtained in hyperdiploid primary MM (HY ATAC), CCND1-translocated MM (CCND1 ATAC), MMSET-translocated MM (MMSET ATAC) and normal plasma cells (ND ATAC) are also shown. MAF ChIP-seq tracks represent FC over input (bars indicating peaks q<0.01), IRF4 ChIP-seq tracks show normalised ChIP-seq signal (bars show peaks q<0.01).

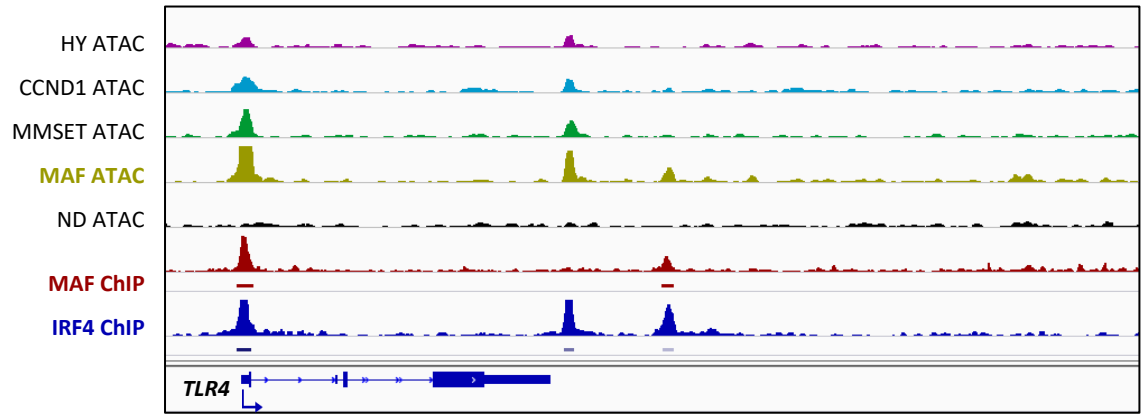


Figure 5.6: Binding of IRF4 on *TLR4* enhancer

IGV plot depicting MAF and IRF4 binding, as identified by ChIP-seq in MM.1S, on the *TLR4* gene and regulatory peak. The ATAC-seq peaks obtained in primary MAF samples (average normalised ATAC -seq signal, MAF ATAC) are also shown, as well as hyperdiploid primary MM (HY ATAC), CCND1-translocated MM (CCND1 ATAC), MMSET-translocated MM (MMSET ATAC) and normal plasma cells (ND ATAC). MAF ChIP-seq tracks represent FC over input (bars indicating peaks q<0.01), IRF4 ChIP-seq tracks show normalised ChIP-seq signal (bars show peaks q<0.01).

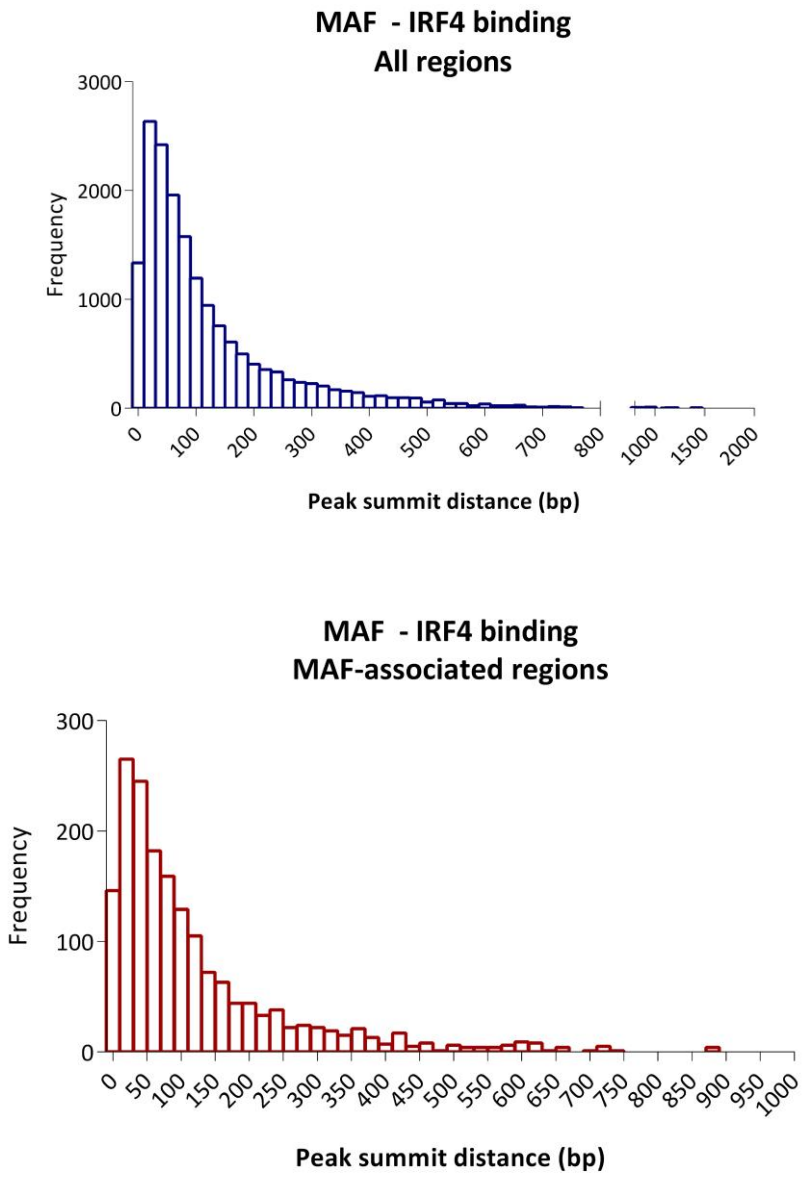


Figure 5.7: Distance between MAF and IRF4 binding

Histograms of the absolute distance between the summits of MAF and IRF4 binding sites, genome-wide in all MAF-IRF4 co-bound regions (top) and the 1,661 co-bound MAF-associated regions (bottom) illustrating the distribution of distance the between the two bound transcription factors in MM.1S cells.

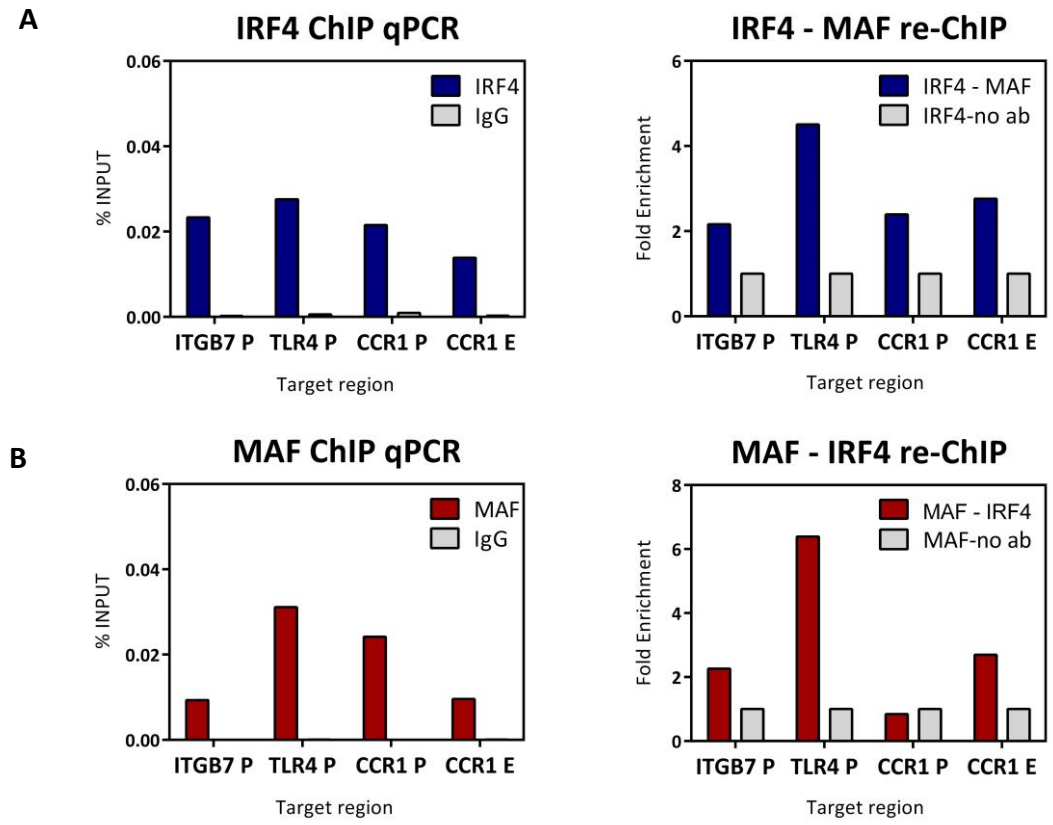


Figure 5.8: ChIP re-ChIP assays for MAF and IRF4

ChIP re-ChIP qPCR for IRF4>MAF **(A)** and MAF>IRF4 **(B)** showing co-occupancy of the two transcription factors in select loci. Graphs on the left depict the ChIP qPCR against each transcription factor and graphs on the right the re-ChIP assay in each direction. ChIP qPCR enrichment expressed as % input DNA and ChIP re-ChIP assays shown as fold enrichment to control reaction (transcription factor followed by no antibody). Bars represent mean of n=2 independent experiments. ITGB7 P; *ITB7* promoter, TLR4 P; *TLR4* promoter, CCR1 P; *CCR1* promoter, CCR1 E; *CCR1* enhancer peak referring to peak 2 as indicated on Fig. 5.5.

5.3 IRF4 and MAF self- and cross-regulation

Having determined that MAF and IRF4 may be co-operating transcription factors in the context of MAF-translocated myeloma, preliminary evidence suggested the two proteins may also be involved in regulating each other.

First, analysis of the IRF4 ChIP-seq data in MM.1S cells showed binding of the IRF4 TF on the *MAF* promoter (**Fig. 5.9**). This was confirmed by in-house ChIP-qPCR in both t(14;16) cell lines

(**Fig. 5.9**). The converse was also true; ChIP-seq and ChIP-qPCR confirmed MAF binding on the *IRF4* enhancer in the two MAF-translocated cell lines, MM.1S and JJN3 (**Fig.5.10**). Furthermore, shRNA-mediated MAF knockdown led to the downregulation of *IRF4* expression in both MM.1S and JJN3, (**Fig. 5.11**), further supporting the possibility of a role of MAF in regulating *IRF4* expression. Interestingly, both transcription factors were also found to be bound to their own regulatory regions, suggesting the possibility of auto-regulation, in addition to cross-regulation between MAF and IRF4 (**Fig. 5.9, 5.10**). However, comparing IRF4 expression across the MM subgroups, revealed that IRF4 expression was not higher in the MAF samples, and indeed only marginally higher than normal PC, with the range of IRF4 expression in other MM samples overlapping with the range of expression seen in normal bone marrow PC (**Fig. 5.12**). This suggests, as expected, that regulation of IRF4 in the MAF-samples is complex, with other factors likely involved, which would require further investigation in order to delineate.

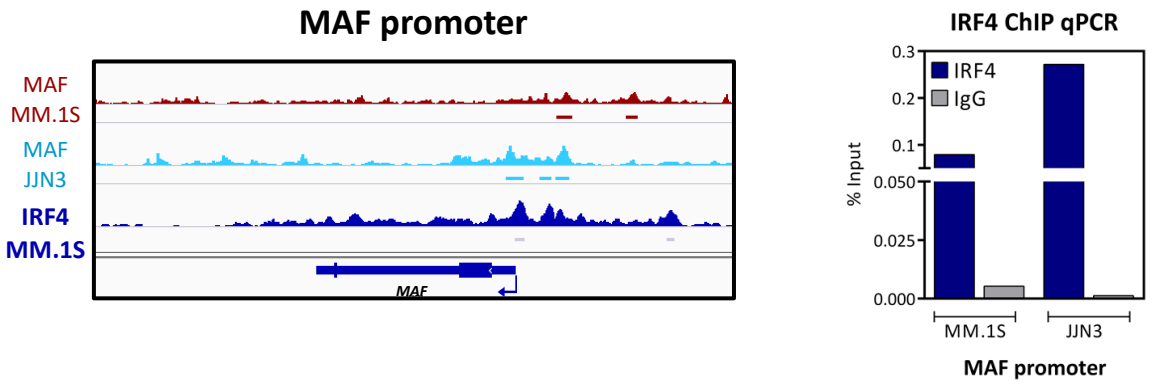


Figure 5.9: IRF4 binding on *MAF* promoter

IGV plot (left) of ChIP-seq data, showing binding of MAF on the *MAF* promoter in MM.1S cells (red) and JJN3 (light blue). Highlighted, in dark blue, ChIP-seq against IRF4 in MM.1S (data from the Young Lab, GSM1195560), showing binding of IRF4 on the *MAF* promoter. On the right, ChIP-qPCR showing IRF4 binding in MM.1S and JJN3 cell lines in the *MAF* promoter, expressed as percentage of input DNA and compared to IgG control. Bars represent mean of n=2 independent experiments. MAF ChIP-seq tracks represent FC over input (bars indicating peaks q<0.01), IRF4 ChIP-seq tracks show normalised ChIP-seq signal (bars show peaks q<0.01).

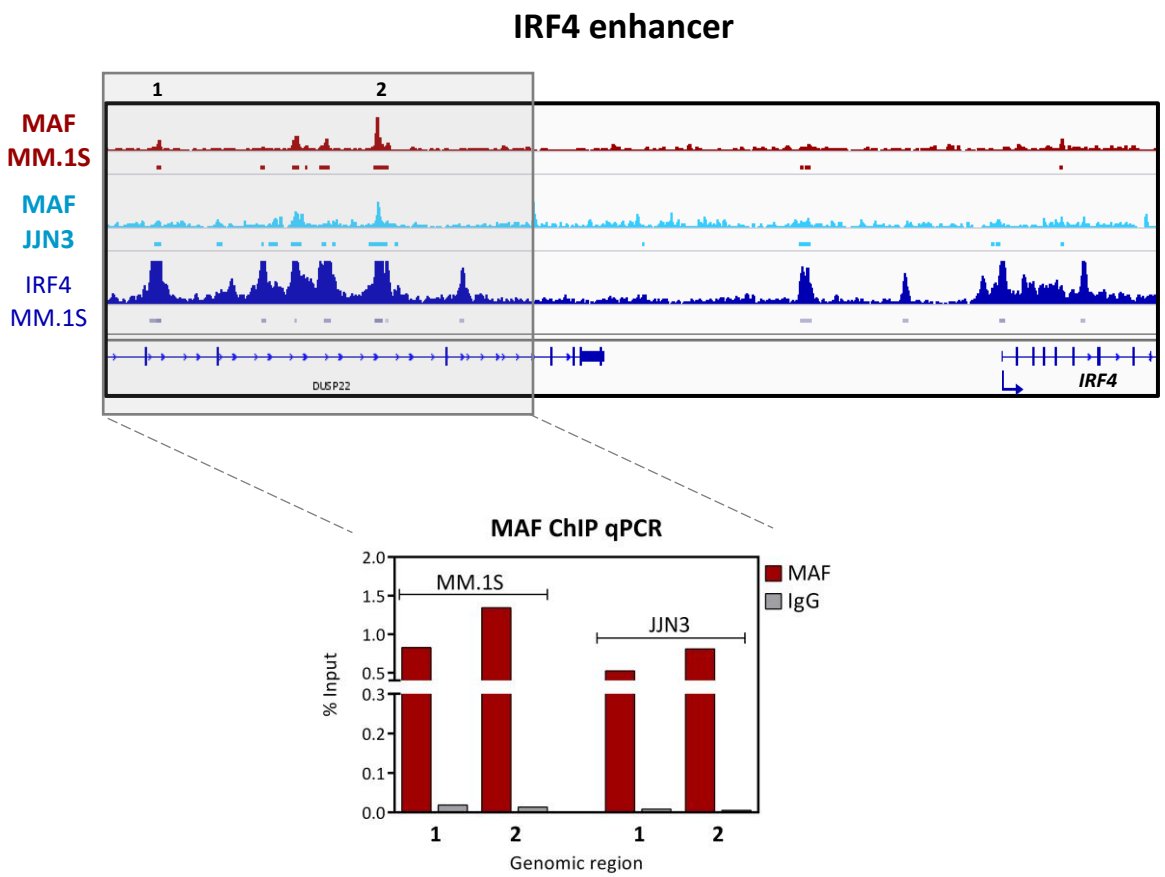


Figure 5.10: MAF binding on *IRF4* enhancer

IGV plot (top) of ChIP-seq data, showing binding of MAF on the *IRF4* superenhancer (*IRF4/DUSP22* superenhancer location as previously described[170, 171]) in MM.1S cells (red) and JJN3 (light blue). In dark blue, ChIP-seq against IRF4 in MM.1S (data from the Young Lab, GSM1195560). Below, ChIP-qPCR showing MAF binding in MM.1S and JJN3 cell lines in the two *IRF4* enhancer peaks indicated, expressed as percentage of input DNA and compared to IgG control. Bars represent mean of n=2 independent experiments. MAF ChIP-seq tracks represent FC over input (bars indicating peaks q<0.01), IRF4 ChIP-seq tracks show normalised ChIP-seq signal (bars show peaks q<0.01).

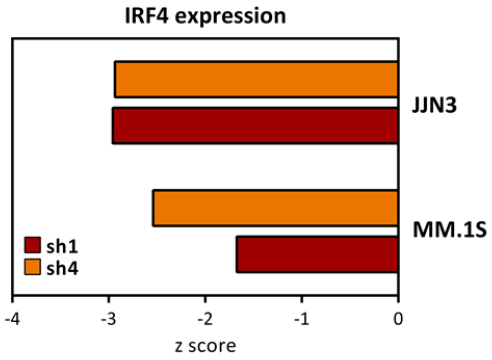


Figure 5.11: *IRF4* expression upon MAF knockdown

IRF4 expression on RNA-seq following MAF shRNA-mediated knockdown in MM.1S and JJN3 cells, using two different shRNA constructs, sh1 and sh4. Expression shown as z-score compared to scrbl control.

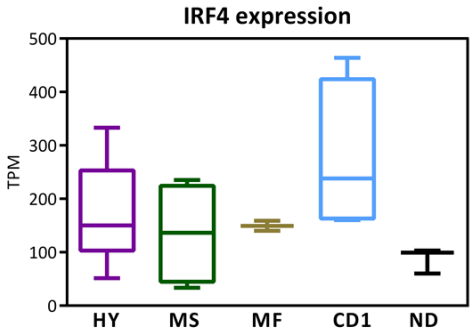


Figure 5.12: *IRF4* expression across MM subgroups

RNA-seq data showing *IRF4* expression (in TPM) across the MM subgroups, compared with normal bone marrow plasma cells. HY; hyperdiploid subgroup, MS; MMSET-translocated MM, MF; MAF-translocated MM, CD1; CCND1-translocated MM, ND; normal donor plasma cells.

5.4 Interaction of MAF with IRF4 and other cooperating factors investigated by chromatin proteomics

In order to identify other candidate partners of MAF, such as other transcription factors or chromatin binding proteins, I employed Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins (RIME)[172, 173]. This technique enables identification of the protein component of chromatin pull-down, by performing on-bead trypsinisation of the immunoprecipitate, followed by protein purification and mass spectrometry (**Fig. 5.13**). For this purpose, as large quantities of antibody were required and sufficient quantities of the original anti-MAF antibody used for ChIP-seq were unavailable, I first validated a suitable anti-MAF antibody, by using ChIP-qPCR in JJN3 cells against known MAF-bound loci (**Appendix E Fig. S1**). Two hundred million JJN3 cells were subjected to RIME followed by mass spectrometry performed by Active Motif. MAF reactions were performed in two technical replicates and IgG reactions were used as a control for each immunoprecipitation. Analysis of the mass spectrometry output was performed by Active Motif, using the PEAKS X (Bioinformatics Solutions Inc.) software.

A total of 1,424 proteins were identified following peptide spectrum match at FDR 1.5%. Candidate proteins were subsequently filtered for background levels of detection, identifying 404 proteins present in any of the reactions. In order to ensure high accuracy of results, proteins were only considered significantly enriched if they were identified in both MAF reactions and were present below the background threshold in both IgG control reactions, as previously described[173]. This resulted in 72 proteins being identified as MAF-associated (**Fig. 5.14A**). In support of my previous findings, IRF4 was included in the potential MAF-interactors, at an 11-fold change increase in relative abundance over IgG background (**Fig. 5.14B**).

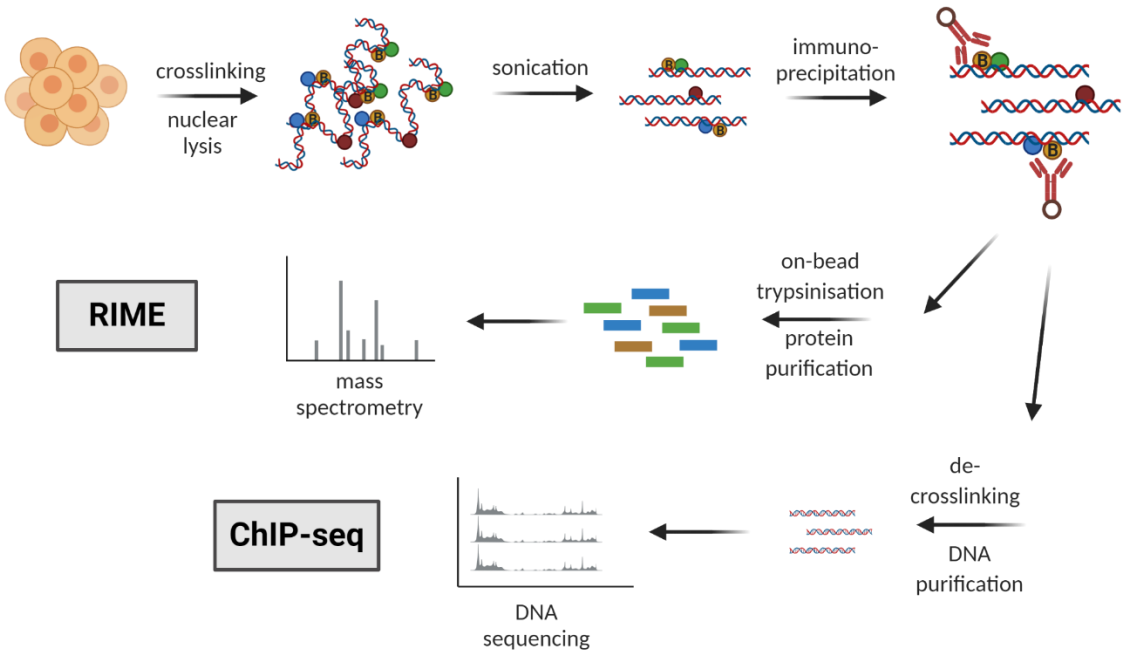


Figure 5.13: Schematic of Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins (RIME)
Schematic depicting the process of RIME. The technique allows identification of proteins interacting with a target – bait – protein, enriching for on-chromatin partners. This can be coupled with ChIP-seq experiments, to identify the bait's DNA binding sites.

Reassuringly, 66/72 (92%) of the proteins were nuclear proteins, where MAF, being a transcription factor, would be primarily found. Gene ontology (GO) terms were used to annotate the biological and molecular function of the potential MAF interactors. As expected, 57/72 (79%) proteins were confirmed to have a known role in the regulation of gene

expression (enrichment -$\log_{10}$FDR 19.59). Within the MAF interactome, several proteins were found to be RNA binding factors (37/72), whereas 25/72 interactors, were DNA binding and processing proteins (**Fig. 5.15**). Enrichment analysis revealed that the top most enriched functions of MAF interactors were related to RNA processing and regulation (**Fig. 5.16**), this included Heterogeneous Nuclear Ribonucleoprotein L (HNRPL) and Cell Cycle And Apoptosis Regulator 2 (CCAR2), both amongst the most highly enriched proteins in the MAF interactome (2nd and 5th most enriched respectively). However, of special interest were 14 proteins known to be involved in chromatin organisation, including 6 with nucleosomal binding properties and 2 proteins involved in binding supercoiled DNA (**Fig. 5.16, 5.17**). In addition, 9 proteins have a known role in the epigenetic regulation of gene expression (**Fig. 5.16, 5.17**). Taken together, these findings further support the role of MAF as an epigenetic regulator, capable of influencing chromatin accessibility.

Within these groups of proteins known to be involved in chromatin organisation and epigenetic gene expression regulation, a couple of good candidates stand out for further validation as potential cooperating factors with MAF in its role as transcriptional activator and regulator of chromatin accessibility. High mobility group protein B2 (HMGB2) belongs to a family of chromatin binding proteins, involved in chromatin remodelling and inducing DNA conformational change[174, 175]. HMGB2 has been shown to increase the binding affinity of steroid receptors to DNA, resulting in upregulation of target gene transcription[176], and *in vitro* studies have confirmed its ability to bind nucleosomes and facilitate nucleosomal sliding[177]. In addition, HMGB proteins have been shown to promote DNA looping, thereby enabling interactions between promoters and regulatory regions[178]. These properties make HMGB2 a good potential candidate partner of MAF, capable of promoting its function in epigenetic activation. Another interesting MAF-partner protein candidate is Protein Arginine Methyltransferase 1 (PRMT1), an arginine methyltransferase known to methylate histones H2, H3 and H4, amongst others, specifically mediating the methylation of H4R3me1 and H4R3me2a leading to transcriptional activation.

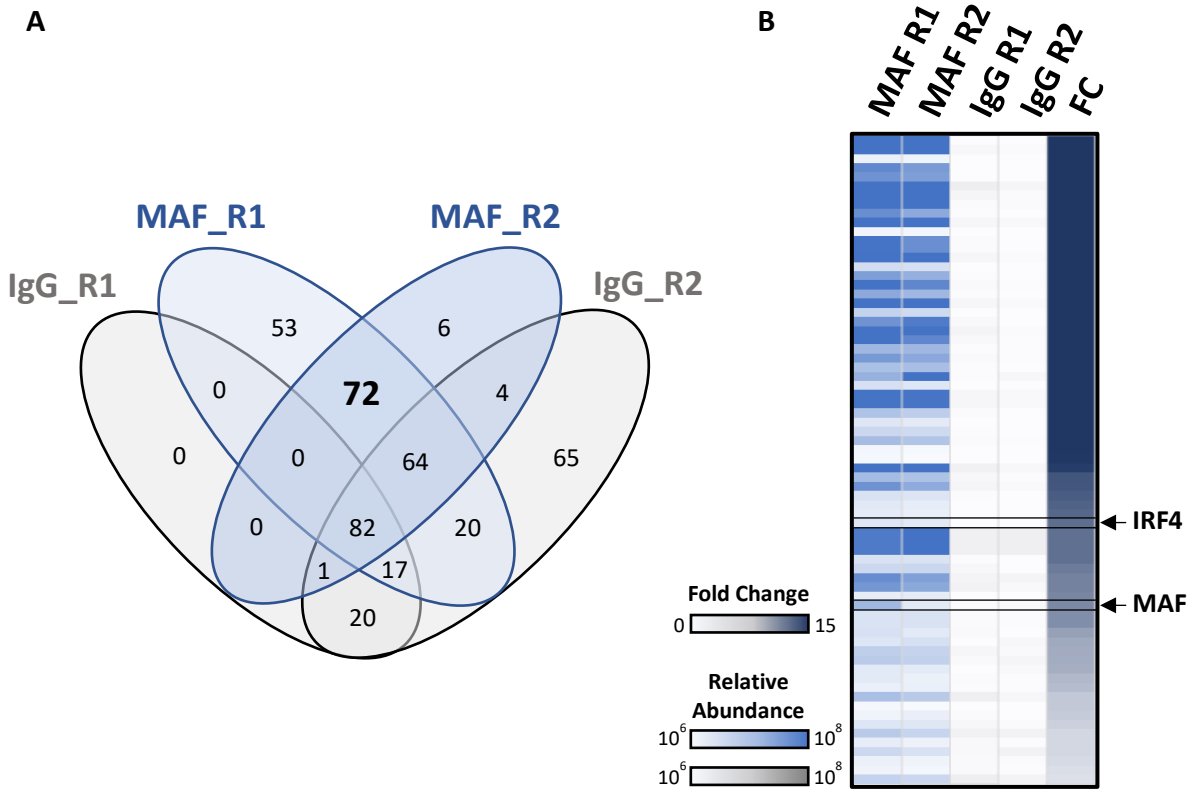


Figure 5.14: MAF RIME results

(A) Venn diagram representing number of proteins detected in each reaction above background threshold. (B) Heatmap of the relative abundance of the enriched proteins and fold change enrichment, averaged between the two MAF reactions over IgG. Relative abundance of each protein is the average intensity of mass spectrometry signal of the top 3 most abundant peptides relative to the total sample protein abundance. Arrows indicate enrichment of MAF itself and IRF4. MAF R1; MAF immunoprecipitation reaction replicate 1, IgG R1; IgG reaction replicate 1, MAF R2; second replicate of MAF immunoprecipitation reaction, IgG R2, IgG reaction replicate 2, FC; fold change of relative abundance.

PRMT1 has been shown to be essential for the malignant transformation and maintenance of MLL-translocated acute myeloid leukaemia, with PRMT1 shown to co-immunoprecipitate with the fusion product and bind MLL-fusion target genes leading to increased H4R3me2a at target gene loci and upregulation of their expression[179]. Thus, cooperation with PRMT1 could mechanistically explain MAF's role in epigenetic activation of transcription.

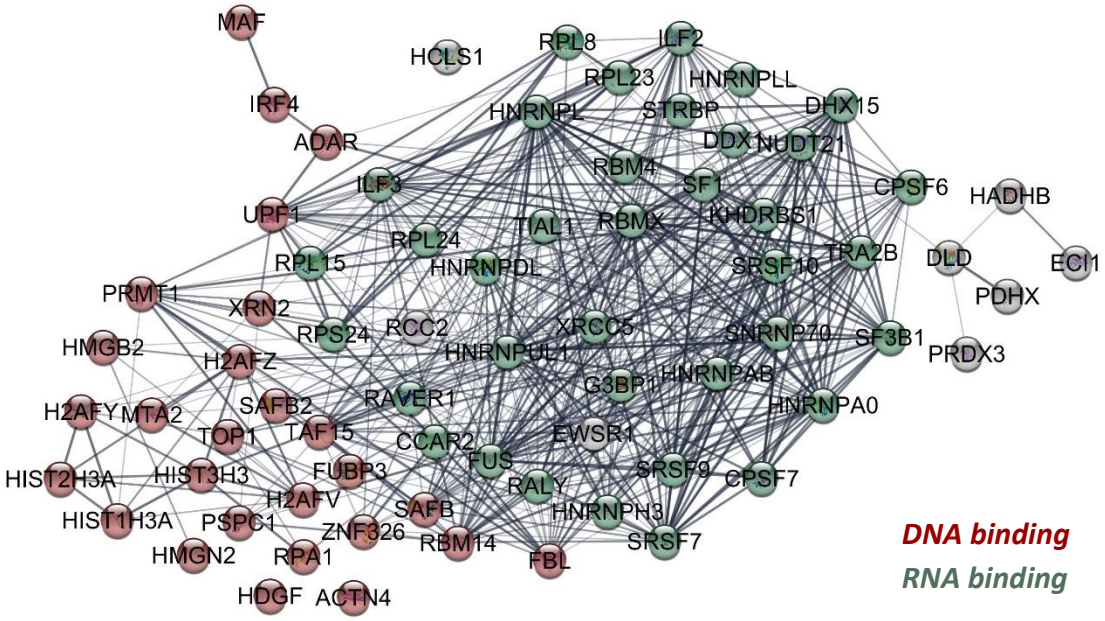


Figure 5.15: String network of MAF interactors

String network of MAF interactors, visualised using the Cytoscape software[122], highlighting proteins that are known to be DNA or RNA binding.

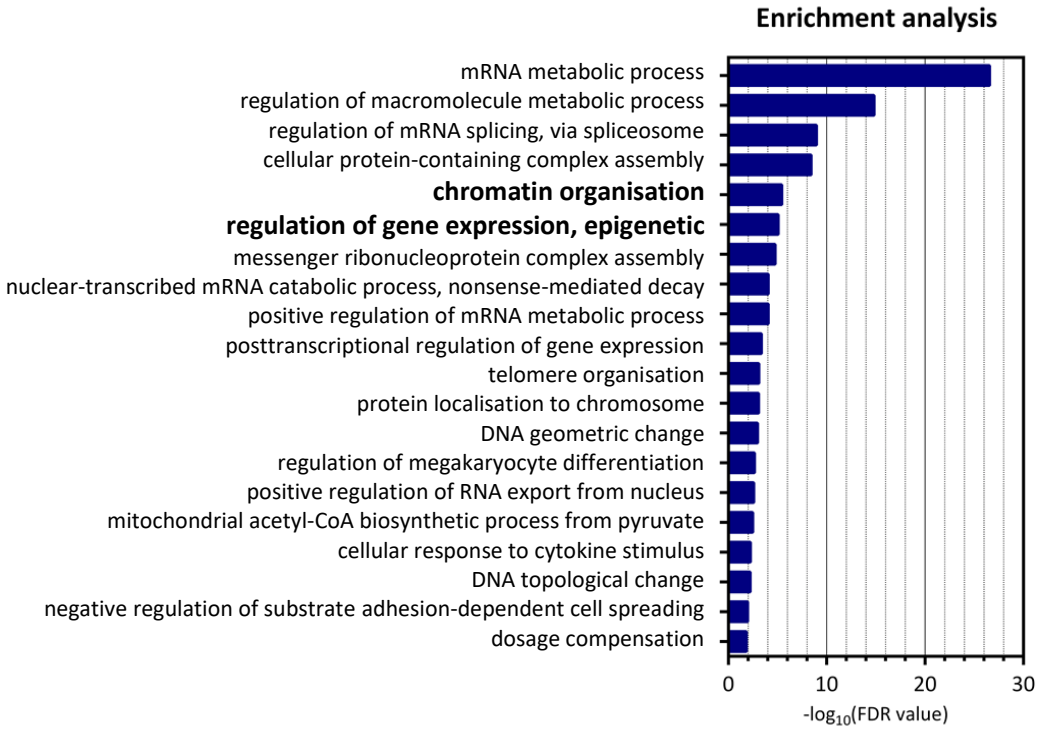


Figure 5.16: Biological functions of MAF interactors

Enrichment analysis of the biological function of MAF interactors, showing top 20 most significant, using Gene Ontology (GO) terms.

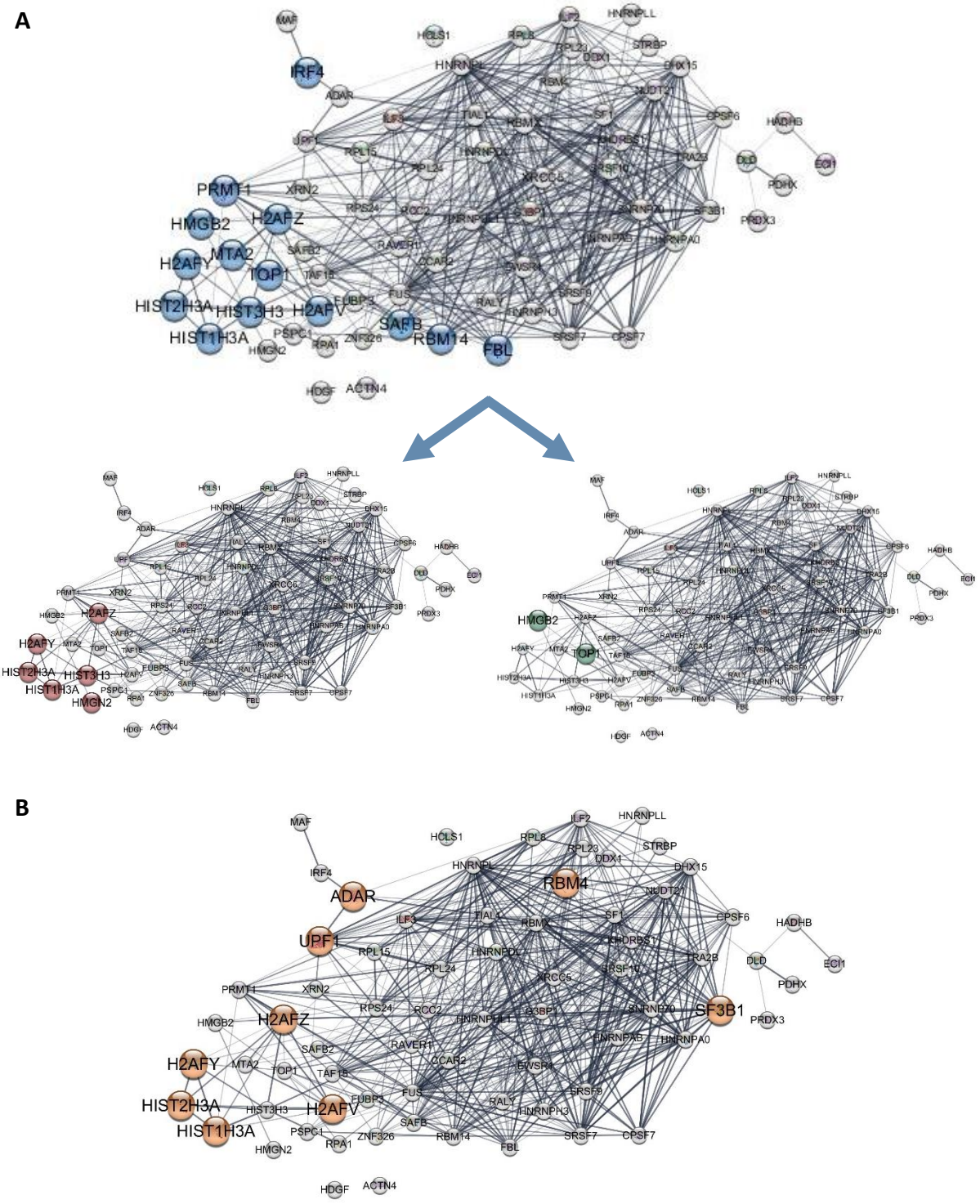


Figure 5.17: Molecular functions of MAF interactors

(A) String network of MAF interactors, highlighting in blue proteins involved in chromatin organisation. Some of these proteins have nucleosomal binding properties (red) and two proteins are known to bind supercoiled DNA (green). **(B)** Proteins with a known role in epigenetic regulation of gene expression (orange).

5.5 Discussion

The experiments and analysis in this chapter have provided some insights into potential MAF interacting and coopering factors.

Starting from the motif analysis of the MAF ChIP-seq data in t(14;16) MMCLs I identified IRF4 as a putative MAF-cooperating transcription factor. IRF4 is a TF, belonging to the IRF family of TFs, that is expressed in B and T cells upon stimulation of the B and T cell receptor respectively [168]. It is also one of the key TFs in plasma cell development[15, 166]. IRF4 has a DNA binding domain as well as an IRF Association Domain, thought to mediate protein-protein interactions with TFs of various families, which enhance its own weak binding to DNA[168]. Its interaction for example with PU.1 results in a conformational change in IRF4 allowing avid DNA binding and enabling the heterodimer to transactivate enhancers such as the Igκ and IL-1β enhancers[168]. Irf4 has also been shown to cooperate with members of the bZIP family of transcription factors, such as batf and jun[169]. Following the observation of enrichment of AP-1 motifs in irf4 ChIP-seq in mouse T cells, the authors demonstrated co-localisation of jun, batf and irf4, with the complex localising on the *IL-10* locus among others. Although the authors were unable to consistently co-immunoprecipitate irf4 with any of the two transcription factors, and were only able to co-immunoprecipitate jun-batf, knockdown of either of the proteins, reduced binding of the other TF by ChIP-seq, demonstrating cooperative DNA binding of the transcription factors[169].

There is little literature on the cooperation of MAF and IRF4, despite both transcription factors being involved in T follicular helper (Tfh) cell development and function. Rengarajan et al demonstrated that ectopically expressing MAF and IRF4, combined with NFATc2, can exponentially increase expression of *IL-4*, leading to the conclusion that the TFs act synergistically to upregulate IL-4[180]. In Tfh cells, co-expression of MAF and IRF4 could transactivate the *IRF4* promoter, but neither of the two transcription factors could do so individually, suggesting the formation of a cooperating complex between the two proteins enhancing *IRF4* expression[51].

In the context of t(14;16) MM, due to the ectopic overexpression of MAF, the two TFs co-exist. Here, using publicly available IRF4 ChIP-seq data in MM.1S, I was able to demonstrate co-localisation of the two transcription factors, genome-wide as well as in the "MAF-associated" regions. This was confirmed by in-house ChIP-qPCR in loci of interest. In addition, IRF4 was found to be co-bound with MAF on the *CCR1* validated "MAF-activated" enhancer peak. Co-binding was subsequently experimentally validated by ChIP re-ChIP assays in select loci.

Further experiments should be performed, to validate the above findings, which are currently underway. First, an in-house IRF4 ChIP-seq, would confirm the analysis results from the publicly available data. However, chromatin immunoprecipitation experiments cannot determine if MAF forms a complex with IRF4 or instead the transcription factors facilitate each other's DNA-binding and function. IP co-IP assays on endogenous proteins could help elucidate this further. Finally, although in this chapter I have demonstrated co-localisation of IRF4 on MAF binding sites and target regions, I have not delineated the functional consequences of this. In order to experimentally determine if the two transcription factors indeed co-regulate MAF target genes, IRF4 knockdown would be required followed by RNA-seq with subsequent integration with the MAF shRNA-mediated knockdown transcriptome, in order to assess the extent of their overlap and establish the co-regulated network of MAF-IRF4.

IRF4 knockdown experiments have previously been performed in MMCL, though not in a MAF-translocated cell line, demonstrating the dependence of MM cells on IRF4[167]. Comparing gene expression profiling following IRF4 knockdown with gene expression in normal B cells and plasma cells, revealed that IRF4 regulated a more extensive transcriptional programme in MM that either in PC or B cells. This observation prompted the authors to hypothesise that this could be due to epigenetic changes in MM, allowing IRF4 to bind to and upregulate loci, previously inactive in normal PC[167]. It is plausible that in the context of t(14;16) MM, MAF overexpression is responsible for these epigenetic changes that subsequently potentiate the action of IRF4, especially for example in the "MAF-activated" regions. The MAF-overexpression system lends itself well to testing such a hypothesis. ChIP-qPCR against the two transcription factors in MAF-overexpressing vs EV control U266 cells,

(which express IRF4 but not MAF) would help to establish if MAF binds to genome areas pre-marked by IRF4 or instead MAF binding facilitates and potentiates IRF4 binding.

In this chapter I have also examined the linear distance between the binding sites of the two transcription factors, in genomic regions where co-binding has been observed. Genome-wide and in the "MAF-associated" regions, the median distance was 75bp. This analysis is however limited by the resolution of ChIP-seq assays, where sonication generates 250-500bp long DNA fragments, with sequencing acquiring 50bp reads. With this limitation in resolution in mind, a 75bp linear distance might imply the two transcription factors are not in direct contact. However, although the transcription factors may not be linearly directly adjacent, this doesn't preclude IRF4 and MAF being in direct contact via DNA looping. In addition, the two transcription factors could be part of the same protein complex, as is the case of the polymerase II pre-initiation complex, spanning a DNA length of 70bp[181]. Finally, it is possible that the two TFs cooperate without physical contact, by for example inducing DNA conformational change upon binding of one transcription factor that facilitates and enhances the binding of the second; a mechanism of TF cooperation that has previously been proposed[163]. In order to delineate the exact mechanism of cooperation between MAF and IRF4, further investigation is required. To address the ChIP-seq resolution constraints, higher resolution techniques could be applied, such as ChIP-exo where following crosslinking, exposed chromatin is further cut by an exonuclease thus providing single bp resolution[182]. This would provide a more precise picture of the distance between the TF binding. IP co-IP assays could establish whether the proteins are in physical contact or not, which could be further confirmed using luminometry assays, that are based on chemical labelling of antibodies that subsequently luminesce upon close contact. Alternatively, a proximity ligation assay, PLA, could be employed, where the proximity of primary antibodies can be detected by secondary antibodies labelled with "proximity probes", oligonucleotides which can form continuous DNA fragments when in close proximity, which can then be amplified and detected[183].

Finally, in this chapter I present preliminary data identifying other potential MAF-cooperating proteins, using RIME. This technique enables the identification of the interactome of a bait protein, following crosslinking and immunoprecipitation. It's a technique requiring a relatively

small number of cells and enables detection of endogenous proteins without the need for chemical modifications. However, it cannot distinguish between direct and indirect interactions and is unable to determine if proteins identified are in physical contact. In addition, co-localising but perhaps non-interacting complexes would not be distinguished. Further, locus-dependent variations in interacting proteins and the specific interactome associated with a particular enhancer cannot be elucidated. This may well be crucial for a transcription factor like MAF whose function involves activating specific regulatory regions. Finally, by excluding proteins found in the IgG reaction, false negatives could ensue, as it is conceivable that a protein interacting with the bait protein could also be found in small amounts in the IgG reaction.

Despite those limitations, some interesting candidate proteins were identified, which if validated, could offer insight into the mechanism of action of MAF in the context of MAF-myeloma. IRF4 was identified as a partner of MAF, providing complementary evidence to my earlier results. Other candidates of the MAF interactome, such as HMGB2 or PRMT1, could provide a mechanistic explanation of MAF's role as in transcriptional activation and regulation of chromatin accessibility. HMGB2 is a chromatin remodeller, previously shown to bind to and enhance the mobility of nucleosomes, as well as enhancing contact between promoters and regulatory regions by promoting DNA looping. PRMT1 on the other hand, is an arginine methyltransferase, shown to activate transcription via histone methylation. The properties of either of these proteins could explain the mechanisms by which MAF epigenetically activates enhancer regions, especially in regions found to be previously inactive.

These preliminary data would require further investigation and validation. In the first instance, RIME in a second MAF-translocated MMCL would ensure consistency of the results, thereby increasing their reliability. Validatory experiments thereafter would depend on the target protein of interest. In the case of HMGB2, chromatin immunoprecipitation against HMGB2 and MAF in the MAF overexpressing U266 compared to EV control cells and IP co-IP experiments would demonstrate co-binding and direct physical contact respectively. Knockdown of HMGB2 followed by MAF ChIP would further confirm or refute co-operation between the two proteins. Similar experiments would be required to support PRMT1's role as a MAF-cooperating or enabling protein, such as ChIP experiments in MAF-overexpressing

cells, especially in comparison to EV control cells, to demonstrate co-localisation and increase in H4R3 methylation following MAF overexpression. Knockdown or PRMT1 inhibitor experiments would be required to investigate the effect of PRMT1 inhibition on MAF targets and MAF binding. Separately, other future experiments could focus on identifying enhancer-specific interacting proteins, for example by using CRISPR Affinity Purification in situ of Regulatory Elements (CAPTURE) where a targeting a specific locus using a biotinylated dCAS9 enables identification of proteins interacting with the regulatory region of interest[93].

In summary, in this section I have identified IRF4 as a putative MAF-cooperating transcription factor and have attempted to delineate the compendium of other MAF partnering proteins. These results provide a mechanistic explanation for my previous findings supporting MAF's role as an epigenetic activator of regulatory regions in the context of t(14;16) MM. MAF and IRF4 could be cooperating directly in enhancer regulation, via direct physical contact. On the other hand, indirect cooperation between TFs has previously been extensively described, either via co-localisation and binding of a common protein complex or via inducing chromatin remodelling, enabling the binding of the second TFs by displacing nucleosomes or changing chromatin structure to increase the affinity of binding of the partner TFs and other protein complexes[184]. In either case, other co-bound factors may be enabling MAF's epigenetic activating role, and potential candidates have been identified in this chapter. Confirmatory experiments are required to further delineate the MAF interactome and validate specific partnering proteins. Such work has potential future applications in providing candidate targets for drug inhibition of protein-protein interactions. Pellegrino et al. for example, designed peptide inhibitors against the leucine zipper domain of MAF and demonstrated *in vitro* their ability to disrupt MAF dimerisation[29]. Disrupting a specific and critical protein – protein interaction as a method of disrupting MAF function, is likely to be more selective in its approach and MM-specific and thus makes for are an attractive drug target.

Chapter 6

Conclusions & Future work

Multiple myeloma (MM) is an incurable malignancy of plasma cells (PC) and the MAF translocation to the *IgH* enhancer is one of the primary oncogenic events[14, 19, 21, 22] accounting for 3-5% of MM cases[18-20]. Little is known about the function of MAF in MM and the consequences of its ectopic overexpression in PC. Using primarily myeloma cell lines, some studies have identified a handful of MAF's downstream target genes[45, 79, 85]. In addition, gene expression studies in MM patients have identified genes selectively upregulated in t(14;16) MM when compared to other MM subtypes[28, 67, 68, 88]. However, the full compendium of molecular targets directly bound by and regulated by MAF have not been previously described. In addition, chromatin accessibility and epigenetic changes imposed by oncogenic MAF have never been elucidated, and the only study exploring the epigenome in MM did not contain any MAF-translocated samples due to the rarity of the disease[144].

The overall aim of the project is to determine the impact of ectopic MAF expression on MM PC, by investigating the transcriptome changes imposed by the translocation as well as the influence of the transcription factor on the regulatory genome. Using normal PC as reference whenever possible, enabled me to identify changes occurring in t(14;16) MM and describe the molecular mechanisms by which MAF promotes myelomagenesis.

In the first chapter, I determined the transcriptome profile of primary MAF-translocated PC as this compares to normal BM PC, identifying both genes uniquely differentially expressed in MAF MM as well as those shared with other MM genetic subgroups. The transcriptome profile of MAF however was distinctly different and able to cluster MAF MM away from other MM subtypes. In order to determine the transcriptome changes imposed by the transcription factor and causally link the two, I used shRNA mediated knockdown in t(14;16) MMCL to define the genome-wide functional targets of MAF. MAF depletion lead to MM cell death, confirming the t(14;16) MM cells' dependency on the transcription factor. In order to determine the MAF's direct gene targets, the cistrome of MAF was defined using ChIP-seq against MAF in t(14;16) MMCLs, identifying for the first time the genome-wide binding profile of the transcription factor. Integrating the MAF cistrome and the transcriptome following MAF knockdown in t(14;16) MMCL MM.1S, identified a group of 2,112 directly regulated targets of MAF, the majority of which (1,436 genes) are upregulated by the transcription factor. In addition to confirming genes previously linked to t(14;16) MM in GEP studies as

direct targets of MAF (such as *CCR1*, *TLR4*, *CX3CR1*), and confirming known targets of MAF as directly bound by the transcription factor (such as *ITGB7*, *CCND2*, *NUAK1*), it also identified new, previously unrecognised MAF gene targets, such as *CD28*, *IRF4*, *FLI1*. Reflecting this gene list back on the differential transcriptome of the primary t(14;16) MM cells identified a small list of just over 100 genes, which likely represent the core transcriptional programme directly activated by MAF in MM PC. This revealed key pathways affected by ectopic MAF expression and included, in addition to cell cycle regulation, cell to cell interaction pathways influencing the cross-talk between MM cells and the bone marrow microenvironment, such as cell adhesion, migration, cytokine and chemokine signalling pathways.

Analysis of the MAF cistrome revealed that the transcription factor preferentially binds enhancers and regulatory regions. In the next chapter therefore, I investigated the impact of MAF on genomic regulatory regions, focusing specifically on a select group of regions previously inactive in the normal PC in order to determine the effect of ectopic MAF expression on the PC *en route* to myelomagenesis. When compared to normal PC, MAF primary samples displayed overall enhanced chromatin accessibility. This was more pronounced than other MM subgroups, with 33% of regions being differentially accessible uniquely in MAF MM. Indeed, the distinct chromatin accessibility profile of MAF MM was able to cluster it separately from the other MM genetic groups. One quarter of these MAF over-accessible regions were bound by the transcription factor, the "MAF-associated" regions, the majority of which were located outside TSS. Annotating these regions using MM.1S chromatin marks and juxtaposing them against chromatin mark maps of normal PC and B cells of different maturation stages, revealed that 43% of the "MAF-associated regions" displayed activatory marks in t(14;16) cells but were inactive in normal PC (of which over half were over-accessible uniquely in MAF MM). This identified a group of 726 epigenetically annotated promoters and enhancers previously inactive in the normal PC, the putative "MAF-activated" regions. These regions are MAF-bound, differentially accessible in MAF MM compared to normal PC and display activatory chromatin marks in t(14;16) cells whilst being inactive in normal PC and mostly previous B cell developmental stages. The activation of these regulatory regions is therefore linked to the transcription factor. In order to determine the extent of MAF's direct role in the observed activation of these regions, I ectopically overexpressed the TF in a non-MAF MMCL. Over 50% of the putative "MAF-activated" regions acquired

enhanced accessibility upon MAF overexpression. Therefore MAF, in the context of t(14;16) MM, acts as an activator of specific genomic regulatory regions, enhancing the expression of key MAF target genes. As an example, using CRISPRi I have identified and validated an enhancer of *CCR1*, one of the direct MAF target genes.

Separately, combining the cistrome of MAF with the accessibility profile of primary MAF myeloma cells enabled the identification of a novel *CCND2* enhancer, one of the key target genes of MAF. CRISPRi experiments functionally validated the *CCND2* enhancer and indeed identified one of the most avidly MAF-bound peaks to have a pronounced effect on the gene's expression.

Finally, in the last chapter, I identified possible protein partners, cooperating with MAF in t(14;16) MM. Analysis of the MAF cistrome identified IRF4 as a possible co-localising TF. This was verified using a publicly available IRF4 ChIP-seq, demonstrating that over 70% of genome-wide MAF binding sites and over 80% of "MAF-associated" and "MAF-activated" regions are co-bound by IRF4, suggesting global cooperation between the two TFs. Co-localisation of MAF and IRF4 was separately confirmed in select MAF-bound loci using ChIP re-ChIP assays. In addition, the observation that the two TFs bind each other's regulatory regions and the reduction of *IRF4* expression upon MAF depletion, suggested possible cross-regulation of the two. Finally, using a proteomics approach, 72 proteins, including IRF4, were identified as being part of the MAF interactome. Candidate MAF-cooperating proteins, such as HMGB2, a chromatin binding protein involved in chromatin remodelling and nucleosomal sliding, and PRMT1, an arginine methyltransferase involved in histone methylation, provide insights into possible mechanisms underpinning MAF's role as an activator of regulatory regions.

In summary, this work has provided insights into the role of oncogenic MAF in multiple myeloma. First, it provides a comprehensive description of MAF's transcriptome and cistrome, identifying direct MAF target genes and key pathways regulated by the transcription factor in t(14;16) MM. Second it details the regulatory genomic changes associated with ectopic expression of MAF in MM. This led to identification of a number of regulatory regions directly activated by oncogenic MAF, having previously been quiescent in the normal PC, establishing the role of MAF as an activator of specific regulatory regions in this cellular

context. Finally, it identifies IRF4 as a potential MAF-cooperating transcription factor, whilst also suggesting other possible candidates of the MAF oncogenic interactome.

Overall, results presented here give an insight into events related to ectopic MAF expression in the plasma cell during the process of myelomagenesis. They also however raise many more questions, that require additional work in order to be addressed. One question that remains to be fully addressed is whether MAF acts as a pioneer factor in the regulatory areas it has been found to activate or instead depends on pre-existing binding of other transcription factors, and experiments trying to address this are currently under way. In addition, further exploration of the exact nature of IRF4 and MAF cooperation is warranted, as well as validatory experiments of other potential MAF partners, such as through IP co-IP assays, ChIP, ChIP re-ChIP and knockdown experiments. Although I have used primary MAF-translocated and normal plasma cells to investigate the transcriptome and chromatin accessibility associated with MAF myeloma, most of the experiments presented here have been done on MMCLs and results have been used to extrapolate what may be happening in the normal plasma cell during MAF-induced myelomagenesis. However, in order to further determine the precise molecular events occurring upon MAF translocation, overexpression experiments would need to be conducted in *in vitro* developed plasma cells, due to the limitations and constraints in obtaining and experimentally handling normal plasma cells.

Finally, this body of work has focused on furthering knowledge on mechanisms underpinning the function of oncogenic MAF in MM, rather than investigating new therapeutic approaches. However, targeting a primary genetic event in MM, to which the malignant clone remains dependent on, may overcome the emergence of resistant subclones. Based on the knowledge acquired regarding the molecular function of the MAF TF, one could envisage the possibility of therapeutic interventions arising from targeting a MM MAF-specific enhancer or interrupting its cooperation or binding to a partner protein, thus targeting all MAF-expressing clones and avoiding unwanted effects on other non-malignant cells. Precise identification of disease-associated enhancers and proteins involved in their regulation enables their manipulation for therapeutic purposes. This for example has been shown following identification of the enhancer upstream of the *globin* genes involved in suppressing fetal haemoglobin during development; upregulation of fetal haemoglobin gene expression was

achieved following *in vitro* knockout of the enhancer, with therapeutic implications for haemoglobinopathies[93].

References

1. UK, C.R.; Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma.
2. Kyle, R.A. and S.V. Rajkumar, *Multiple Myeloma.* New England Journal of Medicine, 2004. **351**(18): p. 1860-1873.
3. Rajkumar, S.V., et al., *International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma.* Lancet Oncol, 2014. **15**(12): p. e538-48.
4. Swerdlow SH, C.E., Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (Revised 4th edition)* 2017, Lyon: IARC.
5. Therneau, T.M., et al., *Incidence of monoclonal gammopathy of undetermined significance and estimation of duration before first clinical recognition.* Mayo Clinic proceedings, 2012. **87**(11): p. 1071-1079.
6. Kyle, R.A., et al., *Monoclonal gammopathy of undetermined significance (MGUS) and smoldering (asymptomatic) multiple myeloma: IMWG consensus perspectives risk factors for progression and guidelines for monitoring and management.* Leukemia, 2010. **24**(6): p. 1121-7.
7. Weinstock, M. and I.M. Ghobrial, *Extramedullary multiple myeloma.* Leuk Lymphoma, 2013. **54**(6): p. 1135-41.
8. Kumar, S.K., et al., *Improved survival in multiple myeloma and the impact of novel therapies.* Blood, 2008. **111**(5): p. 2516-20.
9. Rajkumar, S.V., *Multiple myeloma: 2016 update on diagnosis, risk-stratification, and management.* Am J Hematol, 2016. **91**(7): p. 719-34.
10. Terpos, E. and I. Ntanasis-Stathopoulos, *Multiple Myeloma: Clinical Updates From the American Society of Hematology Annual Meeting 2018.* Clin Lymphoma Myeloma Leuk, 2019. **19**(7): p. e324-e336.
11. Kumar, S.K., et al., *Continued improvement in survival in multiple myeloma: changes in early mortality and outcomes in older patients.* Leukemia, 2014. **28**(5): p. 1122-8.
12. Palumbo, A., et al., *Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2015. **33**(26): p. 2863-2869.
13. Burrows, P.D. and M.D. Cooper, *B cell development and differentiation.* Curr Opin Immunol, 1997. **9**(2): p. 239-44.
14. Bergsagel, P.L. and W.M. Kuehl, *Chromosome translocations in multiple myeloma.* Oncogene, 2001. **20**(40): p. 5611-22.
15. Nutt, S.L., et al., *The generation of antibody-secreting plasma cells.* Nat Rev Immunol, 2015. **15**(3): p. 160-71.
16. Roco, J.A., et al., *Class-Switch Recombination Occurs Infrequently in Germinal Centers.* Immunity, 2019. **51**(2): p. 337-350.e7.
17. Kuehl, W.M. and P.L. Bergsagel, *Early genetic events provide the basis for a clinical classification of multiple myeloma.* Hematology Am Soc Hematol Educ Program, 2005: p. 346-52.
18. Morgan, G.J., B.A. Walker, and F.E. Davies, *The genetic architecture of multiple myeloma.* Nat Rev Cancer, 2012. **12**(5): p. 335-48.
19. Chesi, M. and P.L. Bergsagel, *Molecular pathogenesis of multiple myeloma: basic and clinical updates.* International journal of hematology, 2013. **97**(3): p. 313-323.
20. Fonseca, R., et al., *Genetics and cytogenetics of multiple myeloma: a workshop report.* Cancer Res, 2004. **64**(4): p. 1546-58.
21. Avet-Loiseau, H., et al., *14q32 translocations and monosomy 13 observed in monoclonal gammopathy of undetermined significance delineate a multistep process for the oncogenesis of multiple myeloma. Intergroupe Francophone du Myelome.* Cancer Res, 1999. **59**(18): p. 4546-50.
22. Pawlyn, C. and G.J. Morgan, *Evolutionary biology of high-risk multiple myeloma.* Nat Rev Cancer, 2017. **17**(9): p. 543-556.
23. Manier, S., et al., *Genomic complexity of multiple myeloma and its clinical implications.* Nat Rev Clin Oncol, 2017. **14**(2): p. 100-113.

24. Walker, B.A., et al., *Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma.* Blood, 2018.
25. Lohr, J.G., et al., *Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy.* Cancer Cell, 2014. **25**(1): p. 91-101.
26. Ledergor, G., et al., *Single cell dissection of plasma cell heterogeneity in symptomatic and asymptomatic myeloma.* Nat Med, 2018. **24**(12): p. 1867-1876.
27. Blank, V. and N.C. Andrews, *The Maf transcription factors: regulators of differentiation.* Trends Biochem Sci, 1997. **22**(11): p. 437-41.
28. Eychene, A., N. Rocques, and C. Pouponnot, *A new MAFia in cancer.* Nat Rev Cancer, 2008. **8**(9): p. 683-93.
29. Pellegrino, S., et al., *Molecular insights into dimerization inhibition of c-Maf transcription factor.* Biochim Biophys Acta, 2014. **1844**(12): p. 2108-15.
30. Yang, Y. and A. Cvekl, *Large Maf Transcription Factors: Cousins of AP-1 Proteins and Important Regulators of Cellular Differentiation.* Einstein J Biol Med, 2007. **23**(1): p. 2-11.
31. Kataoka, K., *Multiple mechanisms and functions of maf transcription factors in the regulation of tissue-specific genes.* J Biochem, 2007. **141**(6): p. 775-81.
32. Yoshida, T., et al., *The 5'-AT-rich half-site of Maf recognition element: a functional target for bZIP transcription factor Maf.* Nucleic Acids Res, 2005. **33**(11): p. 3465-78.
33. Kataoka, K., M. Noda, and M. Nishizawa, *Maf nuclear oncoprotein recognizes sequences related to an AP-1 site and forms heterodimers with both Fos and Jun.* Mol Cell Biol, 1994. **14**(1): p. 700-12.
34. Peng, S., et al., *c-Maf interacts with c-Myb to down-regulate Bcl-2 expression and increase apoptosis in peripheral CD4 cells.* Eur J Immunol, 2007. **37**(10): p. 2868-80.
35. Hedge, S.P., et al., *c-Maf interacts with c-Myb to regulate transcription of an early myeloid gene during differentiation.* Mol Cell Biol, 1998. **18**(5): p. 2729-37.
36. Newman, J.R. and A.E. Keating, *Comprehensive identification of human bZIP interactions with coiled-coil arrays.* Science, 2003. **300**(5628): p. 2097-101.
37. Qiang, Y.W., et al., *MAF protein mediates innate resistance to proteasome inhibition therapy in multiple myeloma.* Blood, 2016. **128**(25): p. 2919-2930.
38. Xu, Y., et al., *The ubiquitin-conjugating enzyme UBE2O modulates c-Maf stability and induces myeloma cell apoptosis.* J Hematol Oncol, 2017. **10**(1): p. 132.
39. Wang, S., et al., *Inhibition of the deubiquitinase USP5 leads to c-Maf protein degradation and myeloma cell apoptosis.* Cell Death Dis, 2017. **8**(9): p. e3058.
40. Zhang, C. and Z.M. Guo, *Multiple functions of Maf in the regulation of cellular development and differentiation.* Diabetes Metab Res Rev, 2015. **31**(8): p. 773-8.
41. Si, N., et al., *A novel MAF missense mutation leads to congenital nuclear cataract by impacting the transactivation of crystallin and noncrystallin genes.* Gene, 2019. **692**: p. 113-118.
42. Lopez-Pajares, V., et al., *A LncRNA-MAF:MAFB transcription factor network regulates epidermal differentiation.* Dev Cell, 2015. **32**(6): p. 693-706.
43. Wende, H., S.G. Lechner, and C. Birchmeier, *The transcription factor c-Maf in sensory neuron development.* Transcription, 2012. **3**(6): p. 285-289.
44. Wende, H., et al., *The transcription factor c-Maf controls touch receptor development and function.* Science, 2012. **335**(6074): p. 1373-6.
45. Hurt, E.M., et al., *Overexpression of c-maf is a frequent oncogenic event in multiple myeloma that promotes proliferation and pathological interactions with bone marrow stroma.* Cancer Cell, 2004. **5**(2): p. 191-9.
46. Natkunam, Y., et al., *Characterization of c-Maf transcription factor in normal and neoplastic hematolymphoid tissue and its relevance in plasma cell neoplasia.* Am J Clin Pathol, 2009. **132**(3): p. 361-71.
47. Ho, I.C., et al., *The proto-oncogene c-maf is responsible for tissue-specific expression of interleukin-4.* Cell, 1996. **85**(7): p. 973-83.
48. Kim, J.I., et al., *The transcription factor c-Maf controls the production of interleukin-4 but not other Th2 cytokines.* Immunity, 1999. **10**(6): p. 745-51.
49. Andris, F., et al., *The Transcription Factor c-Maf Promotes the Differentiation of Follicular Helper T Cells.* Front Immunol, 2017. **8**: p. 480.
50. Jogdand, G.M., S. Mohanty, and S. Devadas, *Regulators of Tfh Cell Differentiation.* Frontiers in immunology, 2016. **7**: p. 520-520.

51. Sahoo, A., et al., *Batf is important for IL-4 expression in T follicular helper cells.* Nat Commun, 2015. **6**: p. 7997.
52. Wheaton, J.D., C.-H. Yeh, and M. Ciofani, *Cutting Edge: c-Maf Is Required for Regulatory T Cells To Adopt RORγt(+) and Follicular Phenotypes.* Journal of immunology (Baltimore, Md. : 1950), 2017. **199**(12): p. 3931-3936.
53. Kusakabe, M., et al., *c-Maf plays a crucial role for the definitive erythropoiesis that accompanies erythroblastic island formation in the fetal liver.* Blood, 2011. **118**(5): p. 1374-85.
54. Pouponnot, C., et al., *Cell context reveals a dual role for Maf in oncogenesis.* Oncogene, 2006. **25**(9): p. 1299-310.
55. Pavlovic, M., et al., *Enhanced MAF Oncogene Expression and Breast Cancer Bone Metastasis.* J Natl Cancer Inst, 2015. **107**(12): p. djv256.
56. Morito, N., et al., *Overexpression of c-Maf contributes to T-cell lymphoma in both mice and human.* Cancer Res, 2006. **66**(2): p. 812-9.
57. Morito, N., et al., *A novel transgenic mouse model of the human multiple myeloma chromosomal translocation t(14;16)(q32;q23).* Cancer Res, 2011. **71**(2): p. 339-48.
58. Walker, B.A., et al., *Characterization of IGH locus breakpoints in multiple myeloma indicates a subset of translocations appear to occur in pregerminal center B cells.* Blood, 2013. **121**(17): p. 3413-9.
59. Avet-Loiseau, H., et al., *Translocation t(14;16) and multiple myeloma: is it really an independent prognostic factor?* Blood, 2011. **117**(6): p. 2009-11.
60. Avet-Loiseau, H., et al., *Cytogenetic, interphase, and multicolor fluorescence in situ hybridization analyses in primary plasma cell leukemia: a study of 40 patients at diagnosis, on behalf of the Intergroupe Francophone du Myelome and the Groupe Francais de Cytogenetique Hematologique.* Blood, 2001. **97**(3): p. 822-5.
61. Wei, G., et al., *Clinical implications of c-maf expression in plasma cells from patients with multiple myeloma.* Exp Hematol Oncol, 2017. **6**: p. 16.
62. Annunziata, C.M., et al., *A mechanistic rationale for MEK inhibitor therapy in myeloma based on blockade of MAF oncogene expression.* Blood, 2011. **117**(8): p. 2396-404.
63. Moreaux, J., et al., *TACI expression is associated with a mature bone marrow plasma cell signature and C-MAF overexpression in human myeloma cell lines.* Haematologica, 2007. **92**(6): p. 803-11.
64. Fonseca, R., et al., *Clinical and biologic implications of recurrent genomic aberrations in myeloma.* Blood, 2003. **101**(11): p. 4569-75.
65. Shah, V., et al., *Prediction of outcome in newly diagnosed myeloma: a meta-analysis of the molecular profiles of 1905 trial patients.* Leukemia, 2018. **32**(1): p. 102-110.
66. Kuiper, R., et al., *A gene expression signature for high-risk multiple myeloma.* Leukemia, 2012. **26**(11): p. 2406-13.
67. Zhan, F., et al., *The molecular classification of multiple myeloma.* Blood, 2006. **108**(6): p. 2020-8.
68. Broyl, A., et al., *Gene expression profiling for molecular classification of multiple myeloma in newly diagnosed patients.* Blood, 2010. **116**(14): p. 2543-53.
69. Nair, B., et al., *Superior results of Total Therapy 3 (2003-33) in gene expression profiling-defined low-risk multiple myeloma confirmed in subsequent trial 2006-66 with VRD maintenance.* Blood, 2010. **115**(21): p. 4168-73.
70. Boyd, K.D., et al., *A novel prognostic model in myeloma based on co-segregating adverse FISH lesions and the ISS: analysis of patients treated in the MRC Myeloma IX trial.* Leukemia, 2012. **26**(2): p. 349-55.
71. Walker, B.A., et al., *Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma.* J Clin Oncol, 2015. **33**(33): p. 3911-20.
72. Lagana, A., et al., *Integrative network analysis identifies novel drivers of pathogenesis and progression in newly diagnosed multiple myeloma.* Leukemia, 2018. **32**(1): p. 120-130.
73. Walker, B.A., et al., *APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma.* Nat Commun, 2015. **6**: p. 6997.
74. Avet-Loiseau, H., et al., *Oncogenesis of multiple myeloma: 14q32 and 13q chromosomal abnormalities are not randomly distributed, but correlate with natural history, immunological features, and clinical presentation.* Blood, 2002. **99**(6): p. 2185-91.
75. Wu, P., et al., *A gene expression-based predictor for myeloma patients at high risk of developing bone disease on bisphosphonate treatment.* Clin Cancer Res, 2011. **17**(19): p. 6347-55.
76. Ziogas, D.C., M.A. Dimopoulos, and E. Kastritis, *Prognostic factors for multiple myeloma in the era of novel therapies.* Expert Rev Hematol, 2018. **11**(11): p. 863-879.

77. Kervoëlen, C., et al., *Dexamethasone-induced cell death is restricted to specific molecular subgroups of multiple myeloma.* Oncotarget, 2015. **6**(29): p. 26922-26934.
78. Mao, X., et al., *A chemical biology screen identifies glucocorticoids that regulate c-maf expression by increasing its proteasomal degradation through up-regulation of ubiquitin.* Blood, 2007. **110**(12): p. 4047-54.
79. Neri, P., et al., *Integrin β7-mediated regulation of multiple myeloma cell adhesion, migration, and invasion.* Blood, 2011. **117**(23): p. 6202-6213.
80. Oba, Y., et al., *MIP-1alpha utilizes both CCR1 and CCR5 to induce osteoclast formation and increase adhesion of myeloma cells to marrow stromal cells.* Exp Hematol, 2005. **33**(3): p. 272-8.
81. Vandyke, K., et al., *HIF-2alpha Promotes Dissemination of Plasma Cells in Multiple Myeloma by Regulating CXCL12/CXCR4 and CCR1.* Cancer Res, 2017. **77**(20): p. 5452-5463.
82. Vallet, S. and K.C. Anderson, *CCR1 as a target for multiple myeloma.* Expert Opin Ther Targets, 2011. **15**(9): p. 1037-47.
83. Menu, E., et al., *Role of CCR1 and CCR5 in homing and growth of multiple myeloma and in the development of osteolytic lesions: a study in the 5TMM model.* Clin Exp Metastasis, 2006. **23**(5-6): p. 291-300.
84. Zeissig, M.N., et al., *Therapeutic Targeting of CCR1 to Prevent Dissemination of Multiple Myeloma Plasma Cells.* Blood, 2019. **134**(Supplement_1): p. 3099-3099.
85. Suzuki, A., et al., *ARK5 is transcriptionally regulated by the Large-MAF family and mediates IGF-1-induced cell invasion in multiple myeloma: ARK5 as a new molecular determinant of malignant multiple myeloma.* Oncogene, 2005. **24**(46): p. 6936-44.
86. Hou, X., et al., *A new role of NUAK1: directly phosphorylating p53 and regulating cell proliferation.* Oncogene, 2011. **30**(26): p. 2933-42.
87. Suzuki, A., et al., *ARK5 is a tumor invasion-associated factor downstream of Akt signaling.* Mol Cell Biol, 2004. **24**(8): p. 3526-35.
88. Mattioli, M., et al., *Gene expression profiling of plasma cell dyscrasias reveals molecular patterns associated with distinct IGH translocations in multiple myeloma.* Oncogene, 2005. **24**(15): p. 2461-73.
89. Moffat, J., et al., *A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen.* Cell, 2006. **124**(6): p. 1283-98.
90. Concordet, J.P. and M. Haeussler, *CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens.* Nucleic Acids Res, 2018. **46**(W1): p. W242-w245.
91. Hsu, P.D., et al., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nat Biotechnol, 2013. **31**(9): p. 827-32.
92. Haeussler, M., et al., *Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR.* Genome Biol, 2016. **17**(1): p. 148.
93. Liu, X., et al., *In Situ Capture of Chromatin Interactions by Biotinylated dCas9.* Cell, 2017. **170**(5): p. 1028-1043.e19.
94. Sanjana, N.E., O. Shalem, and F. Zhang, *Improved vectors and genome-wide libraries for CRISPR screening.* Nat Methods, 2014. **11**(8): p. 783-784.
95. Chen, C. and H. Okayama, *High-efficiency transformation of mammalian cells by plasmid DNA.* Mol Cell Biol, 1987. **7**(8): p. 2745-52.
96. Schmittgen, T.D. and K.J. Livak, *Analyzing real-time PCR data by the comparative C(T) method.* Nat Protoc, 2008. **3**(6): p. 1101-8.
97. Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nat Methods, 2013. **10**(12): p. 1213-8.
98. Caputo, V.S., et al., *Mechanism of Polycomb recruitment to CpG islands revealed by inherited disease-associated mutation.* Hum Mol Genet, 2013. **22**(16): p. 3187-94.
99. Furlan-Magaril, M., H. Rincón-Arano, and F. Recillas-Targa, *Sequential Chromatin Immunoprecipitation Protocol: ChIP–reChIP*, in *DNA-protein interactions : principles and protocols*, T. Moss and B. Leblanc, Editors. 2009, Humana Press: New York.
100. Trasanidis, N., et al., *Distinct Chromatin Accessibility Changes, Aberrant Transcription Factor Networks Combined with Novel Oncogenic Enhancers Characterise Myeloma-Initiating Genetic Events.* Blood, 2019. **134**(Supplement_1): p. 1769-1769.
101. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
102. Liao, Y., G.K. Smyth, and W. Shi, *The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote.* Nucleic Acids Res, 2013. **41**(10): p. e108.

103. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome biology, 2014. **15**(12): p. 550-550.
104. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
105. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.* Nucleic Acids Res, 2016. **44**(W1): p. W90-7.
106. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
107. Broad Institute. *Picard Tools*. Available from: http://broadinstitute.github.io/picard/.
108. Feng, J., et al., *Identifying ChIP-seq enrichment using MACS.* Nature protocols, 2012. **7**(9): p. 1728-1740.
109. Jalili, V., et al., *Using combined evidence from replicates to evaluate ChIP-seq peaks.* Bioinformatics, 2015. **31**(17): p. 2761-9.
110. Ramirez, F., et al., *deepTools: a flexible platform for exploring deep-sequencing data.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W187-91.
111. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Mol Cell, 2010. **38**(4): p. 576-89.
112. McLean, C.Y., et al., *GREAT improves functional interpretation of cis-regulatory regions.* Nat Biotechnol, 2010. **28**(5): p. 495-501.
113. Ernst, J. and M. Kellis, *ChromHMM: automating chromatin-state discovery and characterization.* Nature methods, 2012. **9**(3): p. 215-216.
114. Blueprint Epigenome project. *ChromHmm Segmentation of ChIP-Seq data*. Available from: http://dcc.blueprint-epigenome.eu/#/md/secondary_analysis/Segmentation_of_ChIP-Seq_data_20140811.
115. Stunnenberg, H.G. and M. Hirst, *The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery.* Cell, 2016. **167**(5): p. 1145-1149.
116. Hoffman, M.M., et al., *Integrative annotation of chromatin elements from ENCODE data.* Nucleic Acids Research, 2012. **41**(2): p. 827-841.
117. Dunham, I., et al., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
118. Kundaje, A., et al., *Integrative analysis of 111 reference human epigenomes.* Nature, 2015. **518**(7539): p. 317-30.
119. Allhoff, M., et al., *Detecting differential peaks in ChIP-seq signals with ODIN.* Bioinformatics, 2014. **30**(24): p. 3467-3475.
120. *BLUEPRINT / EPICO data analysis portal 1.0.5*. 2016; Available from: http://blueprint-data.bsc.es/release_2016-08/#!/#%2F.
121. *MMRF Researcher Gateway*. Available from: https://research.themmrf.org/.
122. Doncheva, N.T., et al., *Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data.* J Proteome Res, 2019. **18**(2): p. 623-632.
123. Aschenbrenner, D., et al., *An immunoregulatory and tissue-residency program modulated by c-MAF in human TH17 cells.* Nat Immunol, 2018. **19**(10): p. 1126-1136.
124. Kawano, Y., et al., *Multiple myeloma cells expressing low levels of CD138 have an immature phenotype and reduced sensitivity to lenalidomide.* International journal of oncology, 2012. **41**(3): p. 876-884.
125. Bataille, R., et al., *The phenotype of normal, reactive and malignant plasma cells. Identification of "many and multiple myelomas" and of new targets for myeloma therapy.* Haematologica, 2006. **91**(9): p. 1234-40.
126. Paiva, B., et al., *Utility of flow cytometry immunophenotyping in multiple myeloma and other clonal plasma cell-related disorders.* Cytometry B Clin Cytom, 2010. **78**(4): p. 239-52.
127. Jelinek, T., et al., *Current applications of multiparameter flow cytometry in plasma cell disorders.* Blood Cancer J, 2017. **7**(10): p. e617.
128. Kumar, S., T. Kimlinger, and W. Morice, *Immunophenotyping in multiple myeloma and related plasma cell disorders.* Best practice & research. Clinical haematology, 2010. **23**(3): p. 433-451.
129. Brynjolfsson, S.F., et al., *Long-lived plasma cells in human bone marrow can be either CD19(+) or CD19().* Blood Adv, 2017. **1**(13): p. 835-838.
130. Halliley, J.L., et al., *Long-Lived Plasma Cells Are Contained within the CD19(-)CD38(hi)CD138(+) Subset in Human Bone Marrow.* Immunity, 2015. **43**(1): p. 132-45.

131. Piccaluga, P.P., et al., *Gene expression analysis of angioimmunoblastic lymphoma indicates derivation from T follicular helper cells and vascular endothelial growth factor deregulation.* Cancer Res, 2007. **67**(22): p. 10703-10.
132. Howland, K.C., et al., *The Roles of CD28 and CD40 Ligand in T Cell Activation and Tolerance.* The Journal of Immunology, 2000. **164**(9): p. 4465.
133. Paysan, L., et al., *Rnd3 in Cancer: A Review of the Evidence for Tumor Promoter or Suppressor.* Mol Cancer Res, 2016. **14**(11): p. 1033-1044.
134. Wang, S., et al., *Target analysis by integration of transcriptome and ChIP-seq data with BETA.* Nat Protoc, 2013. **8**(12): p. 2502-15.
135. Cortés, J.R. and T. Palomero, *The curious origins of angioimmunoblastic T-cell lymphoma.* Current opinion in hematology, 2016. **23**(4): p. 434-443.
136. De Matos Simoes, R., et al., *Systematic Characterization of Genes Representing Preferential Molecular Vulnerabilities for Myeloma Cells Compared to Other Neoplasias - Implications for the Biology and Therapeutic Targeting of Myeloma.* Blood, 2019. **134**(Supplement_1): p. 4407-4407.
137. Wada, A., et al., *Role of chemokine CX3CL1 in progression of multiple myeloma via CX3CR1 in bone microenvironments.* Oncol Rep, 2015. **33**(6): p. 2935-9.
138. Marchica, V., et al., *Bone Marrow CX3CL1/Fractalkine is a New Player of the Pro-Angiogenic Microenvironment in Multiple Myeloma Patients.* Cancers, 2019. **11**(3): p. 321.
139. Xu, Y., et al., *Expression and function of toll-like receptors in multiple myeloma patients: toll-like receptor ligands promote multiple myeloma cell growth and survival via activation of nuclear factor-kappaB.* Br J Haematol, 2010. **150**(5): p. 543-53.
140. Bohnhorst, J., et al., *Toll-like receptors mediate proliferation and survival of multiple myeloma cells.* Leukemia, 2006. **20**(6): p. 1138-44.
141. Chiron, D., G. Jego, and C. Pellat-Deuceunynck, *Toll-like receptors: expression and involvement in multiple myeloma.* Leuk Res, 2010. **34**(12): p. 1545-50.
142. Mateo, G., et al., *Prognostic value of immunophenotyping in multiple myeloma: a study by the PETHEMA/GEM cooperative study groups on patients uniformly treated with high-dose therapy.* J Clin Oncol, 2008. **26**(16): p. 2737-44.
143. Nair, J.R., et al., *CD28 expressed on malignant plasma cells induces a prosurvival and immunosuppressive microenvironment.* Journal of immunology (Baltimore, Md. : 1950), 2011. **187**(3): p. 1243-1253.
144. Jin, Y., et al., *Active enhancer and chromatin accessibility landscapes chart the regulatory network of primary multiple myeloma.* Blood, 2018. **131**(19): p. 2138-2150.
145. Calo, E. and J. Wysocka, *Modification of Enhancer Chromatin: What, How, and Why?* Molecular Cell, 2013. **49**(5): p. 825-837.
146. Flavahan, W.A., E. Gaskell, and B.E. Bernstein, *Epigenetic plasticity and the hallmarks of cancer.* Science, 2017. **357**(6348).
147. Corces, M.R., et al., *The chromatin accessibility landscape of primary human cancers.* Science, 2018. **362**(6413).
148. Ott, C.J., et al., *Enhancer Architecture and Essential Core Regulatory Circuitry of Chronic Lymphocytic Leukemia.* Cancer Cell, 2018. **34**(6): p. 982-995.e7.
149. Beekman, R., et al., *The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia.* Nat Med, 2018. **24**(6): p. 868-880.
150. Bisig, B., et al., *c-Maf expression in angioimmunoblastic T-cell lymphoma reflects follicular helper T-cell derivation rather than oncogenesis.* Histopathology, 2012. **60**(2): p. 371-6.
151. Gustafsson, M., et al., *A validated gene regulatory network and GWAS identifies early regulators of T cell-associated diseases.* Sci Transl Med, 2015. **7**(313): p. 313ra178.
152. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nature Protocols, 2013. **8**(11): p. 2281-2308.
153. Gilbert, Luke A., et al., *CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes.* Cell, 2013. **154**(2): p. 442-451.
154. Gilbert, L.A., et al., *Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation.* Cell, 2014. **159**(3): p. 647-61.
155. Xue, H.Y., et al., *CRISPR-Cas9 for medical genetic screens: applications and future perspectives.* J Med Genet, 2016. **53**(2): p. 91-7.
156. Tiedemann, R.E., et al., *Identification of kinetin riboside as a repressor of CCND1 and CCND2 with preclinical antimyeloma activity.* The Journal of clinical investigation, 2008. **118**(5): p. 1750-1764.

157. Bergsagel, P.L. and W.M. Kuehl, *Molecular Pathogenesis and a Consequent Classification of Multiple Myeloma.* Journal of Clinical Oncology, 2005. **23**(26): p. 6333-6338.
158. Kang, K., et al., *Interferon-gamma Represses M2 Gene Expression in Human Macrophages by Disassembling Enhancers Bound by the Transcription Factor MAF.* Immunity, 2017. **47**(2): p. 235-250.e4.
159. Hughes, J.R., et al., *Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment.* Nat Genet, 2014. **46**(2): p. 205-12.
160. Sur, I. and J. Taipale, *The role of enhancers in cancer.* Nat Rev Cancer, 2016. **16**(8): p. 483-93.
161. Zaret, K.S. and J.S. Carroll, *Pioneer transcription factors: establishing competence for gene expression.* Genes Dev, 2011. **25**(21): p. 2227-41.
162. Iwafuchi-Doi, M. and K.S. Zaret, *Pioneer transcription factors in cell reprogramming.* Genes Dev, 2014. **28**(24): p. 2679-92.
163. Morgunova, E. and J. Taipale, *Structural perspective of cooperative transcription factor binding.* Current Opinion in Structural Biology, 2017. **47**: p. 1-8.
164. Jolma, A., et al., *DNA-dependent formation of transcription factor pairs alters their binding specificity.* Nature, 2015. **527**(7578): p. 384-388.
165. Kataoka, K., M. Nishizawa, and S. Kawai, *Structure-function analysis of the maf oncogene product, a member of the b-Zip protein family.* J Virol, 1993. **67**(4): p. 2133-41.
166. Low, M.S.Y., et al., *IRF4 Activity Is Required in Established Plasma Cells to Regulate Gene Transcription and Mitochondrial Homeostasis.* Cell Reports, 2019. **29**(9): p. 2634-2645.e5.
167. Shaffer, A.L., et al., *IRF4 addiction in multiple myeloma.* Nature, 2008. **454**(7201): p. 226-31.
168. Marecki, S. and M.J. Fenton, *The role of IRF-4 in transcriptional regulation.* J Interferon Cytokine Res, 2002. **22**(1): p. 121-33.
169. Li, P., et al., *BATF-JUN is critical for IRF4-mediated transcription in T cells.* Nature, 2012. **490**(7421): p. 543-6.
170. Fulciniti, M., et al., *Non-overlapping Control of Transcriptome by Promoter- and Super-Enhancer-Associated Dependencies in Multiple Myeloma.* Cell Rep, 2018. **25**(13): p. 3693-3705.e6.
171. Loven, J., et al., *Selective inhibition of tumor oncogenes by disruption of super-enhancers.* Cell, 2013. **153**(2): p. 320-34.
172. Mohammed, H., et al., *Endogenous Purification Reveals GREB1 as a Key Estrogen Receptor Regulatory Factor.* Cell Reports, 2013. **3**(2): p. 342-349.
173. Mohammed, H., et al., *Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes.* Nature Protocols, 2016. **11**(2): p. 316-326.
174. McCauley, M.J., et al., *HMGB Binding to DNA: Single and Double Box Motifs.* Journal of Molecular Biology, 2007. **374**(4): p. 993-1004.
175. Bustin, M., *Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins.* Molecular and cellular biology, 1999. **19**(8): p. 5237-5246.
176. Boonyaratanakornkit, V., et al., *High-mobility group chromatin proteins 1 and 2 functionally interact with steroid hormone receptors to enhance their DNA binding in vitro and transcriptional activity in mammalian cells.* Molecular and cellular biology, 1998. **18**(8): p. 4471-4487.
177. Ugrinova, I., I.G. Pashev, and E.A. Pasheva, *Nucleosome binding properties and Co-remodeling activities of native and in vivo acetylated HMGB-1 and HMGB-2 proteins.* Biochemistry, 2009. **48**(27): p. 6502-7.
178. Stros, M., *HMGB proteins: interactions with DNA and chromatin.* Biochim Biophys Acta, 2010. **1799**(1-2): p. 101-13.
179. Cheung, N., et al., *Targeting Aberrant Epigenetic Networks Mediated by PRMT1 and KDM4C in Acute Myeloid Leukemia.* Cancer cell, 2016. **29**(1): p. 32-48.
180. Rengarajan, J., et al., *Interferon regulatory factor 4 (IRF4) interacts with NFATc2 to modulate interleukin 4 gene expression.* J Exp Med, 2002. **195**(8): p. 1003-12.
181. Murakami, K., et al., *Structure of an RNA polymerase II preinitiation complex.* Proceedings of the National Academy of Sciences, 2015. **112**(44): p. 13543.
182. Rhee, H.S. and B.F. Pugh, *Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution.* Cell, 2011. **147**(6): p. 1408-1419.
183. Söderberg, O., et al., *Direct observation of individual endogenous protein complexes in situ by proximity ligation.* Nat Methods, 2006. **3**(12): p. 995-1000.
184. Spitz, F. and E.E.M. Furlong, *Transcription factors: from enhancer binding to developmental control.* Nature Reviews Genetics, 2012. **13**(9): p. 613-626.

185. Trasanidis, N., et al., *PBX1 Co-Operates with FOXM1 to Regulate Myeloma Cell Proliferation and to Define an Ultra High-Risk chr1q Gain Myeloma Patient Subgroup.* Blood, 2019. **134**(Supplement_1): p. 3760-3760.

Appendices

Appendix A

Reagents and kits used

Product	Company	Catalogue Number
2-Propanol	Sigma-Aldrich	I9516
Agarose gel extraction kit	Jena Bioscience	PP-202
Agencourt AMPure XP Beads	Beckman Coulter	A63881
Agilent Bioanalyser High Sensitivity DNA kit	Agilent	5067-4626
Agilent RNA 6000 Pico Kit	Agilent	5067-1513
Alfa Aesar™ HEPES-buffered saline, pH 7.0 (2X for transfection)	Thermofisher Scientific	15450257
AnnexinV apoptosis detection kit	eBioscience	88-8005
Anti-human CD138 FITC	BD Biosciences	552723
Anti-human CD14 Alexa647	Biolegend	325611
Anti-human CD16 Alexa647	Biolegend	302020
Anti-human CD19 APC eFluor780	eBioscience	17-0199
Anti-human CD2 Alexa647	Bio-Rad	MCA1194A647T
Anti-human CD27 PECy7	Biolegend	302837
Anti-human CD3 Alexa647	Biolegend	300321
Anti-human CD319 PE	eBioscience	12-2229-42
Anti-human CD38 BV421	BD Biosciences	562445
Anti-human CD38 eFluor450	eBioscience	48-0389-42
Anti-human CD45 Alexa700	Biolegend	368513
Anti-human CD56 PECy7	eBioscience	25-0567
Anti-human cMAF eFluor660	eBioscience	50-9855-80
Anti-human GPA Alexa647	Bio-Techne	FAB1228R
BD Pharm Lyse	BD Biosciences	555899
β-mercaptoethanol	Sigma-Aldrich	M6250
CD138 MicroBeads	Miltenyi Biotec	130-051-301
DMEM	Sigma-Aldrich	5796
DMSO	Sigma-Aldrich	D2650
Doxycycline hyclate	Sigma-Aldrich	D9891
Dynabeads Protein A	Thermofisher Scientific	10002D
Dynabeads Protein G	Thermofisher Scientific	10004D
EasySep CD14 positive selection kit	Stem cell Technologies	17858
EasySep CD3 positive selection kit	Stem cell Technologies	17851
eBioscience Foxp3 / Transcription Factor Staining Buffer Set	Thermofisher Scientific	00-5523-00
FastAP Thermosensitive Alkaline Phosphatase	Thermofisher Scientific	EF0651
FastDigest BshTI (AgeI)	Thermofisher Scientific	FD1464
FastDigest EcoRI	Thermofisher Scientific	FD0274
FastDigest Esp3I (BsmBI)	Thermofisher Scientific	FD0454
FastDigest Green Buffer (10X)	Thermofisher Scientific	B72

FastDigest KspAI	Thermofisher Scientific	FD1034
FBS	Thermofisher Scientific	10500-064
FcR blocking reagent	Miltenyi Biotec	130-059-901
Formaldehyde solution for molecular biology	Sigma-Aldrich	F8775
Formaldehyde, 16% Methanol Free	Thermofisher Scientific	28908
GeneJET Plasmid Maxiprep Kit	Thermofisher Scientific	K0491
GeneJET Plasmid Miniprep Kit	Thermofisher Scientific	K0502
Glycerol	Honeywell	G7757
hexadimethrine bromide (polybrene)	Sigma-Aldrich	H9268
Histopaque	Sigma-Aldrich	10771
Hoechst 33342 dye	Abcam	ab228551
IGEPAL	Sigma-Aldrich	I8896
IMDM	Sigma-Aldrich	I3390
L-Glutamine	Sigma-Aldrich	G7513
MACS columns	Miltenyi Biotec	130-042-401
MinElute Reaction Cleanup Kit	Qiagen	28204
Mouse IgG2bk isotype control eFluor660	eBioscience	50-4732-80
NEB® 10-beta Competent E. coli	New England Biolabs	C3019I
NEBNext High Fidelity 2x PCR Master Mix	New England Biolabs	M0541
NEBNext Multiplex Oligos for Illumina, Index Primers Set 1 and 2	New England Biolabs	E7335, E7500
NEBNext Poly(A) mRNA Magnetic Isolation Module	New England Biolabs	E7490
NEBNext Ultra II Directional RNA Library Prep Kit for Illumina	New England Biolabs	E7770
NEBNext Ultra II DNA Library Prep kit for Illumina	New England Biolabs	E7645S
NEBNext® Library Quant Kit for Illumina	New England Biolabs	E7630
Non-essential amino acids	Sigma-Aldrich	M7145
Nucleospin RNA II kit	Macherey Nagel	740955
PBS	Sigma-Aldrich	D8537
Penicillin / Streptomycin	Sigma-Aldrich	P0781
Phenol Chloroform	Sigma-Aldrich	77617
Phusion High-Fidelity PCR Master Mix with GC Buffer	New England Biolabs	F532S
PMSF	Sigma-Aldrich	93482
Propidium Iodide	Sigma-Aldrich	P4864
Proteinase K	Thermofisher Scientific	26160
Puromycin	VWR life science	VWRCJ593
Qubit High Sensitivity DNA kit	Life Technologies	Q32851
Qubit RNA assay kit	Life Technologies	Q32852
ReliaPrep RNA Cell Miniprep System kit	Promega	Z6012
RevertAid RT kit	Thermofisher Scientific	K1691
RNAse A	Thermofisher Scientific	EN0531
RoboSep	Stem Cell Technologies	20104
RPMI1640	Sigma-Aldrich	R8758
SIGMAFAST Protease inhibitor	Sigma-Aldrich	S8820
Sodium pyruvate	Sigma-Aldrich	S8636
SYBR Select Master Mix	Applied Biosystems	4472908
T4 DNA Ligase	New England Biolabs	M0202S

T4 DNA Ligase Reaction Buffer	New England Biolabs	B0202S
T4 Polynucleotide Kinase	New England Biolabs	M0201
Taqman Universal PCR Master Mix	Applied Biosystems	4304437
TD Tagment DNA Buffer (Nextera DNA Library Preparation Kit)	Illumina	FC-121-1030
TDE1 Tagment DNA enzyme ((Nextera DNA Library Preparation Kit)	Illumina	FC-121-1030
Tris EDTA (TE)	Fisher BioReagents	165404
Tris-HCl 1M	Thermofisher Scientific	15568025

Appendix B

Table S1: Patient and normal donor details

Primary Samples Used					
Sample Code	**Age**	**Gender**	**Disease / PMH**	**Disease state**	**Selection purity**
MAF s1	48	F	IgG k MM	Diagnosis	99%
MAF s2	65	M	LLC MM	Relapse	98%
ND1	64	M	HbAS, hypercholesterolaemia	N/A	sorted
ND2	34	M	Nil	N/A	sorted
ND3	36	M	HbAS, DM, hypercholesterolaemia	N/A	Sorted
Additional Samples Used					
HY s1	71	F	IgA l MM	Stable	96%
HY s2	77	M	IgG l MM	Diagnosis	97%
HY s3	76	M	IgG k MM	Diagnosis	93%
HY s4	77	F	IgA k MM	Diagnosis	99%
HY s5	83	M	IgA k MM	Relapse	98%
HY s6	53	M	KLC MM	Diagnosis	87%
HY s7	62	M	IgG k MM	Diagnosis	98%
HY s8	59	M	IgG l MM	Diagnosis	99%
HY s9	48	M	IgA k MM	Relapse	100%
HY s10	49	M	IgA k MM	Relapse	100%
HY s11	76	M	IgG k MM	Diagnosis	91%
HY s12	64	F	IgG k MM	Relapse	96%
HY s13	84	F	IgA l MM	Relapse	85%
HY s14	83	M	IgG k MM	Diagnosis	99%
HY s15	48	M	IgA k MM	Relapse	83%
CD1	72	M	IgA l MM	Diagnosis	87%
CD2	59	M	Non secretory MM	Diagnosis	96%
CD3	54	M	IgA k MM	Relapse	96%
CD4	45	M	IgD l MM	Relapse	99%
MS1	49	F	IgA k smouldering	Diagnosis	80%
MS2	72	M	KLC MM	Diagnosis	96%
MS3	74	M	IgG k MM	Diagnosis	98%
MS4	46	M	IgG k M to PCL	Relapse	97%
NA s1	76	F	IgG k MM	Relapse - PCL	99%
NA s2	68	M	Smouldering IgG L MM	Diagnosis	97%
NA s3	86	F	IgG k MM	Diagnosis	93%
NA s4	72	F	IgA k MM	Stable	90%
NA s5	79	F	IgG k MM	Diagnosis	99%
NA s6	72	M	IgA k MM	Relapse	85%
NA s7	79	F	IgG k MM	Relapse	98%

NB: ND; Normal donor, HbAS; sickle cell trait, DM; Diabetes Mellitus, N/A; Not applicable, HY; hyperdiploid, CD; CCND1-translocated, MS; MMSET-translocated, NA; Normal or undetermined cytogenetics. Cytogenetic annotation, as indicated by sample code, defined by FISH.

Table S2: MAF cDNA sequence

MAF cDNA Sequence (5’-3’)
ATGGCATCAGAACTGGCAATGAGCAACTCCGACCTGCCCACCAGTCCCCTG GCCATGGAATACGTCAACGATTTTGACCTCATGAAGTTTGAAGTGAAAAAGGAACCGGTGGAGACCGACC GCATCATCAGCCAGTGCGGCCGTCTCATCGCCGGGGGCTCGCTGTCCTCCACCCCCATGAGCACGCCGTGC AGCTCGGTGCCCCCTTCCCCCAGCTTCTCGGCGCCCAGCCCGGGCTCGGGCAGCGAGCAGAAGGCGCACCT CGAGGATTACTATTGGATGACCGGCTACCCGCAGCAGCTGAACCCCGAGGCGCTGGGCTTCAGCCCCGAG GACGCGGTCGAGGCGCTCATCAGCAACAGCCACCAGCTCCAGGGCGGCTTCGATGGCTACGCGCGCGGGG CGCAGCAGCTGGCCGCGGCGGCCGGGGCCGGTGCCGGCGCCTCCTTGGGCGGCAGCGGCGAGGAGATG GGCCCCGCCGCCGCCGTGGTGTCCGCCGTGATCGCCGCGGCCGCCGCGCAGAGCGGCGCGGGCCCGCACT ACCACCACCACCACCACCACGCCGCCGGCCACCACCACCACCCGACGGCCGGCGCGCCCGGCGCCGCGGG CAGCGCGGCCGCCTCGGCCGGTGGCGCTGGGGGCGCGGGCGGCGGTGGCCCGGCCAGCGCTGGGGGCG GCGGCGGCGGCGGCGGCGGCGGAGGCGGCGGGGGCGCGGCGGGGGCGGGGGGCGCCCTGCACCCGCA CCACGCCGCCGGCGGCCTGCACTTCGACGACCGCTTCTCCGACGAGCAGCTGGTGACCATGTCTGTGCGCG AGCTGAACCGGCAGCTGCGCGGGGTCAGCAAGGAGGAGGTGATCCGGCTGAAGCAGAAGAGGCGGACC CTGAAAAACCGCGGCTATGCCCAGTCCTGCCGCTTCAAGAGGGTGCAGCAGAGACACGTCCTGGAGTCGG AGAAGAACCAGCTGCTGCAGCAAGTCGACCACCTCAAGCAGGAGATCTCCAGGCTGGTGCGCGAGAGGG ACGCGTACAAGGAGAAATACGAGAAGTTGGTGAGCAGCGGCTTCCGAGAAAACGGCTCGAGCAGCGACA ACCCGTCCTCTCCCGAGTTTTTCATAACTGAGCCCACTCGCAAGTTGGAGCCATCAGTGGGATACGCCACAT TTTGGAAGCCCCAGCATCGTGTACTTACCAGTGTGTTCACAAAATGA **Total: 1,212bp**

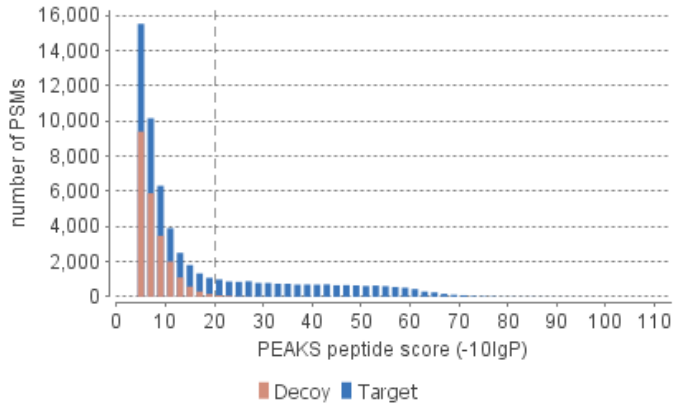


Figure S1: RIME peptide spectrum match distribution

Histogram displaying distribution of peptide spectrum matches (PSM) identified in the RIME reactions, ranked in order of matching score (PEAKS peptide score), indicating cut off at -10logP>20. False matches, using the decoy-fused method displayed in brown (decoy).

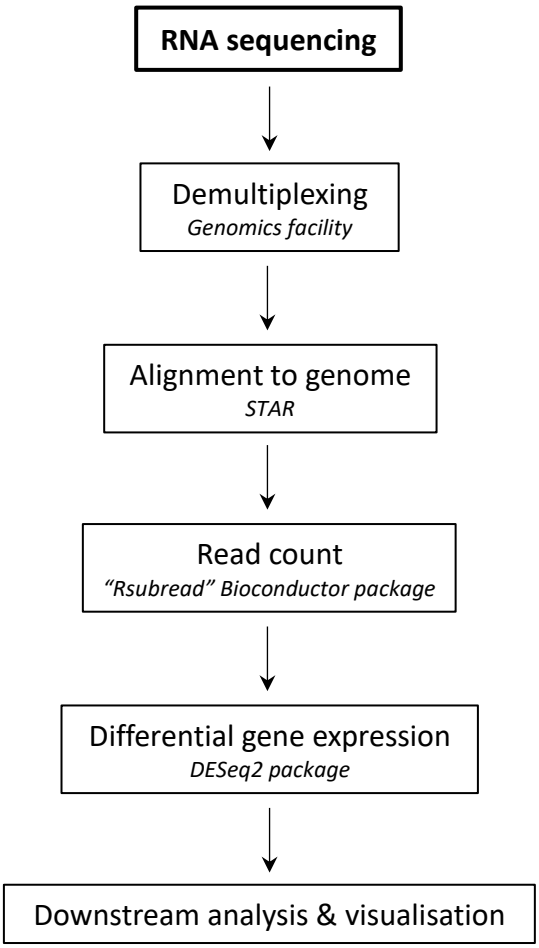


Figure S2: RNA-seq analysis workflow

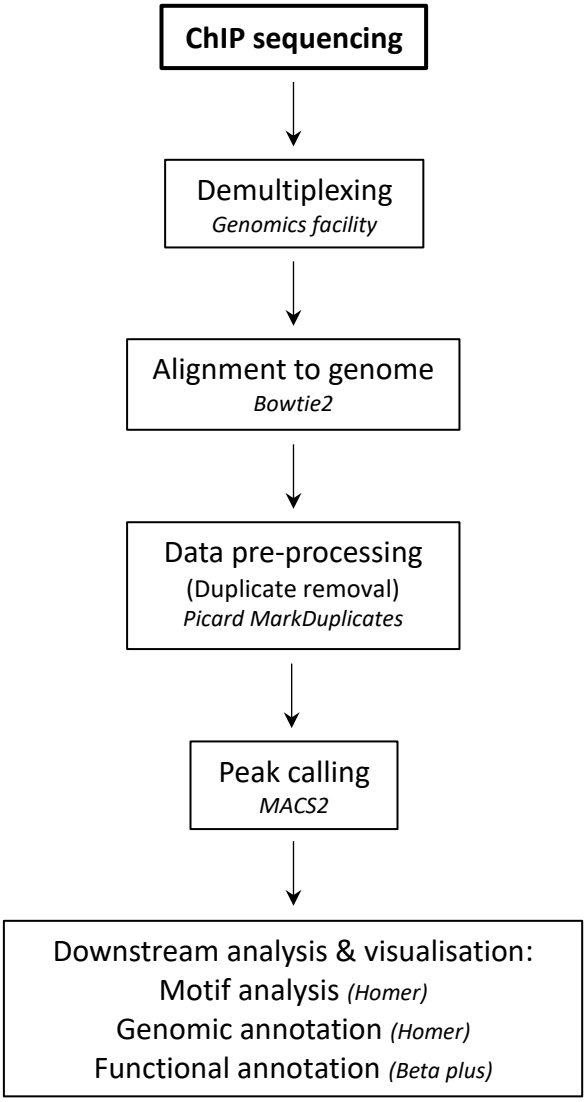


Figure S3: ChIP-seq analysis workflow

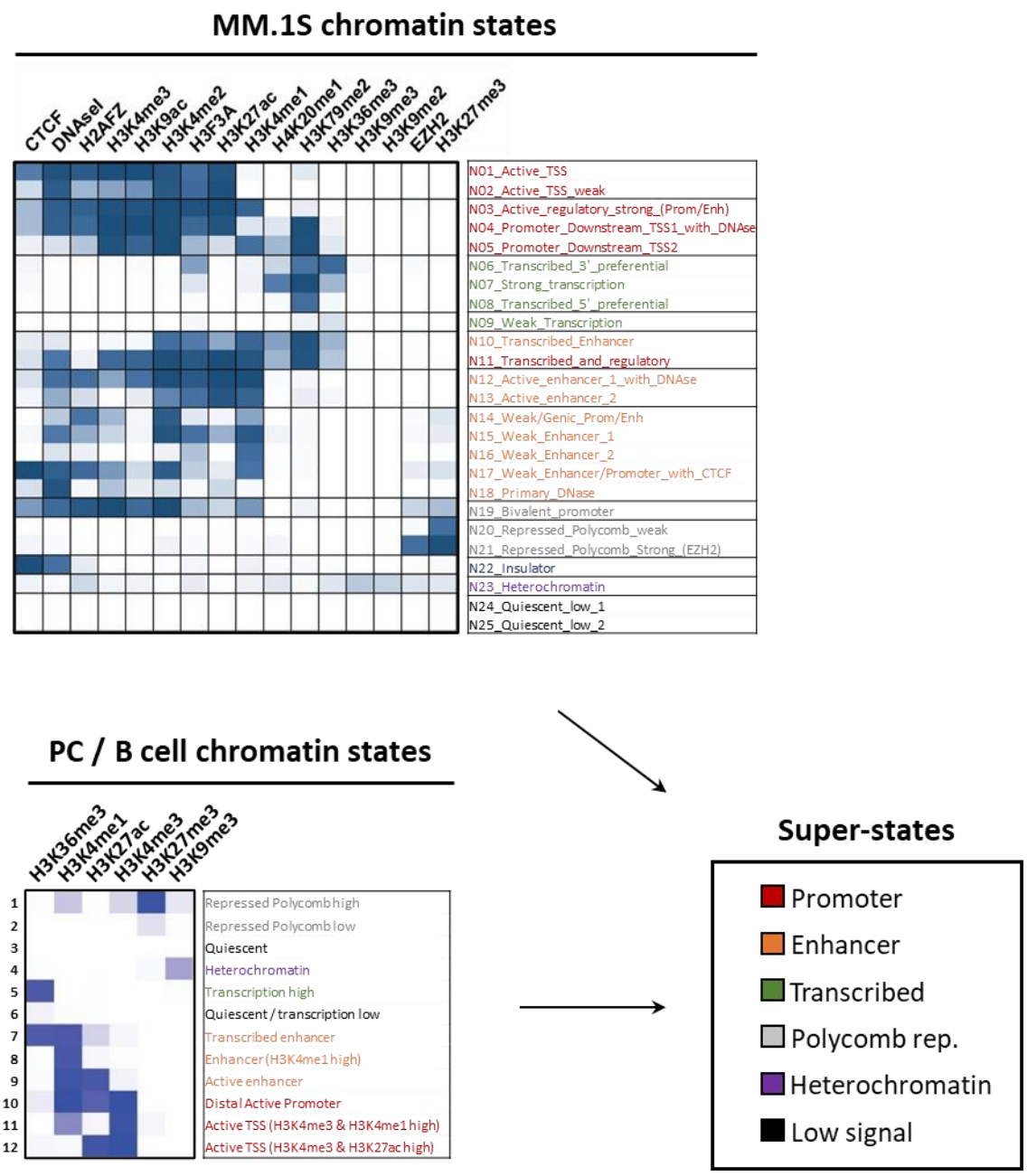


Figure S4: Epigenomic annotation of primary cells and cell lines using chromHMM chromatin states.

Epigenomic maps generated using chromHMM method identified distinct chromatin states for MM.1S (top; 25 states) and primary cells (bottom; 12 chromatin states[114, 115]). To enable easier and efficient interpretation, allowing comparison between myeloma cell lines and primary cells, states were clustered further into 6 distinct biological super-states[185]. Figure adapted from Nikolaos Trasanidis et al.[185].

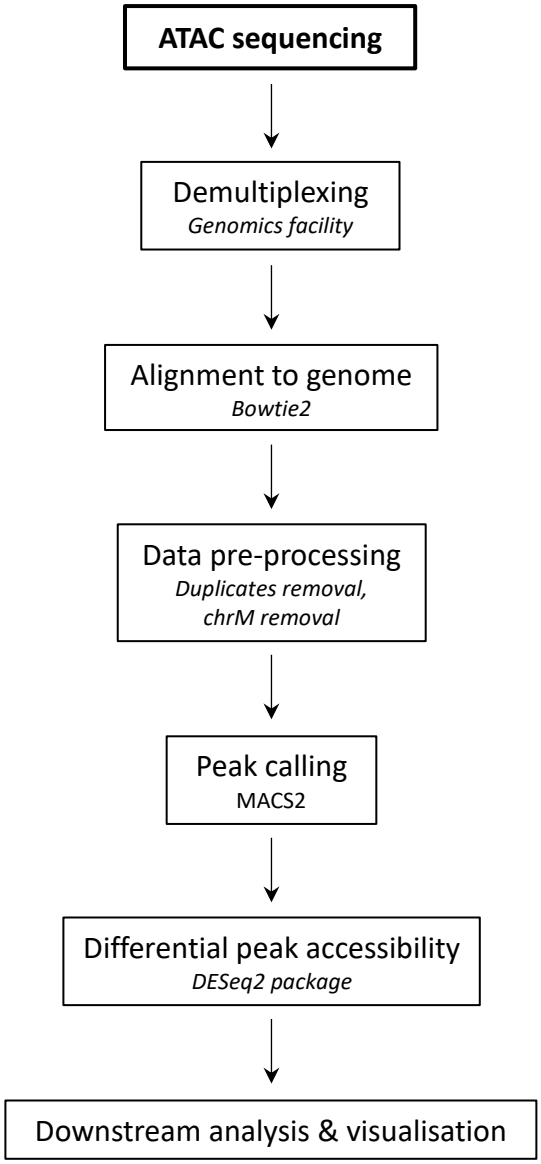


Figure S5: ATAC-seq analysis workflow

Appendix C

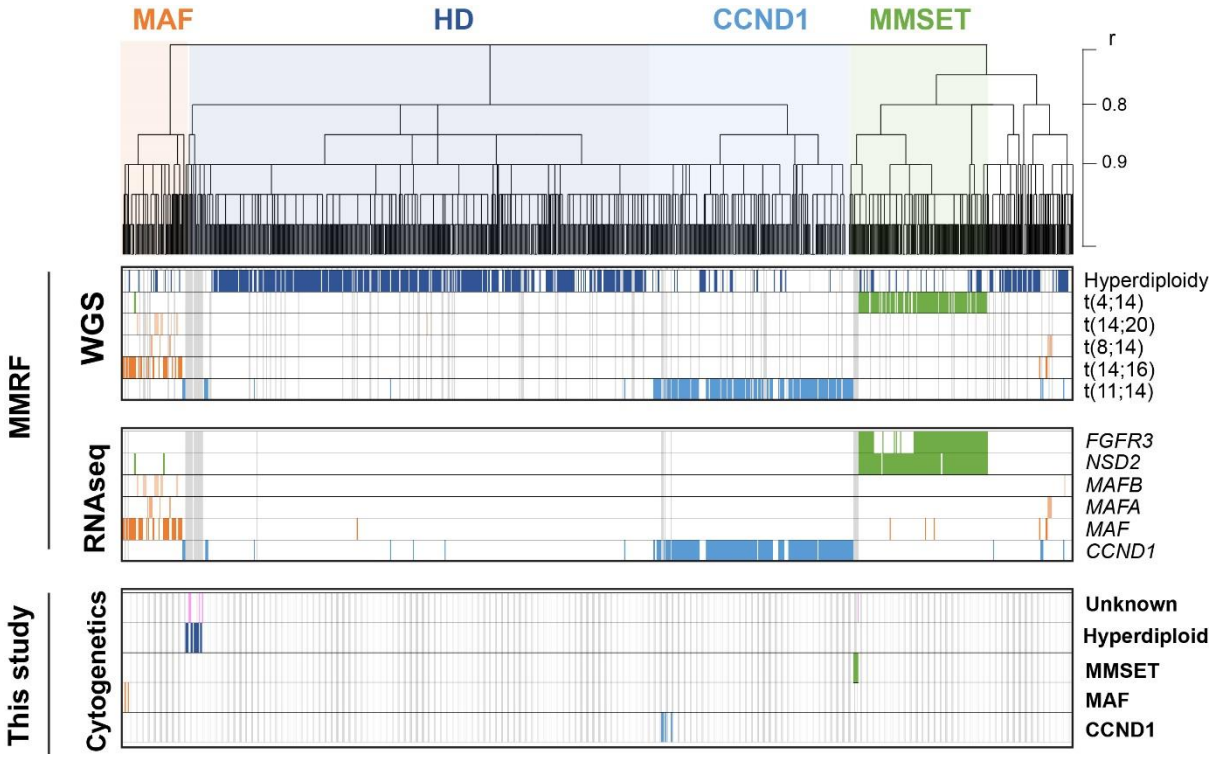


Figure S1: Hierarchical clustering of transcriptome profiles in MM samples

Hierarchical clustering of transcriptomes used in this analysis and as part of a larger collaborative project[100] and transcriptome profiles obtained from 892 MM samples available from the MMRF CoMMpass trial[121]. Clustering performed using the "group-defining" genes as identified in GEP studies[67]. In-house MM samples were assigned to different MM genetic subgroups by FISH. MMRF samples were assigned to genetic subgroups by the MMRF Consortium based on Whole-Genome sequencing (WGS) and RNA-seq. Clustering, using the Pearson correlation method, performed by Nikolaos Trasanidis as part of a larger collaborative work[100]. Figure from Trasanidis, Alvarez-Benayas, Katsarou et al.[100].

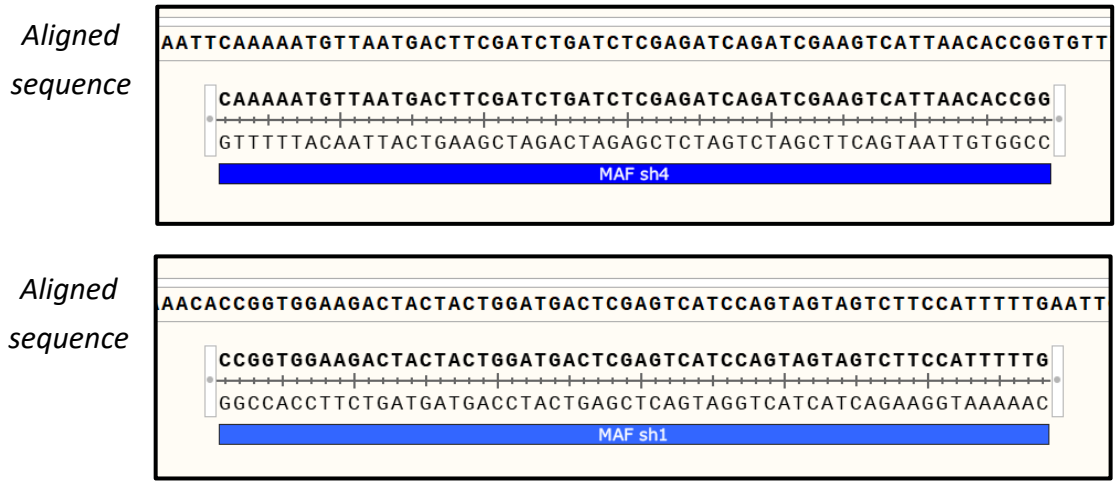


Figure S2: shRNA construct sequencing

shRNA constructs were sequenced with sanger sequencing confirming successful cloning.

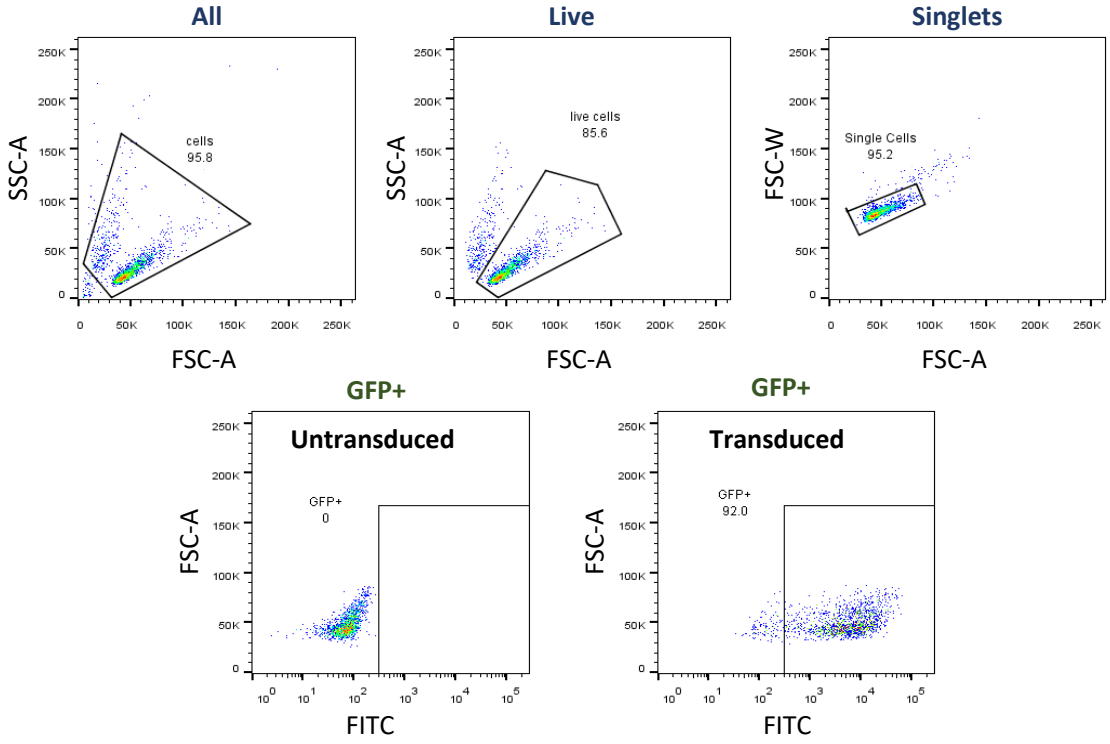


Figure S3: Gating strategy in pLKO.1 GFP+ transduced cells

Representative plots of MM.1S, untransduced and transduced with pLKO.1 lentivirus used for MAF shRNA knockdown experiments, depicting the gating strategy to identify transduced, GFP+ cells. Cells were transduced on day 0, and GFP positivity was assessed on day 3 of each experiment, as a percentage of live, single cells.

Table S1: Transduction efficiency obtained on day 3 of each knockdown experiment

Experiment	*Construct*	*Transduction efficiency*
MM.1S rep 1a	sh1	88.4%
	sh4	71.5%
	scrbl	49.7%
MM.1S rep 1b	sh1	85%
	sh4	69.3%
	scrbl	44.9%
MM.1S rep 2a	sh1	98.8%
	sh4	98.9%
	scrbl	85.5%
MM.1S rep 2b	sh1	99%
	sh4	98.3%
	scrbl	88.8%
MM.1S rep 3a	sh1	97.8%
	sh4	98.6%
	scrbl	88.3%
MM.1S rep 3b	sh1	98.5%
	sh4	99.4%
	scrbl	90.7%
JJN3 rep 1a	sh1	40.2%
	sh4	31.4%
	scrbl	13.5%
JJN3 rep 1b	sh1	37.4%
	sh4	29.8%
	scrbl	11.2%
JJN3 rep 2a	sh1	45.5%
	sh4	36.9%
	scrbl	14.1%
JJN3 rep 2b	sh1	41.6%
	sh4	42.4%
	scrbl	12.7%
JJN3 rep 3a	sh1	37.3%
	sh4	30.2%
	scrbl	18.3%
JJN3 rep 3b	sh1	35.5%
	sh4	32.6%
	scrbl	14.6%

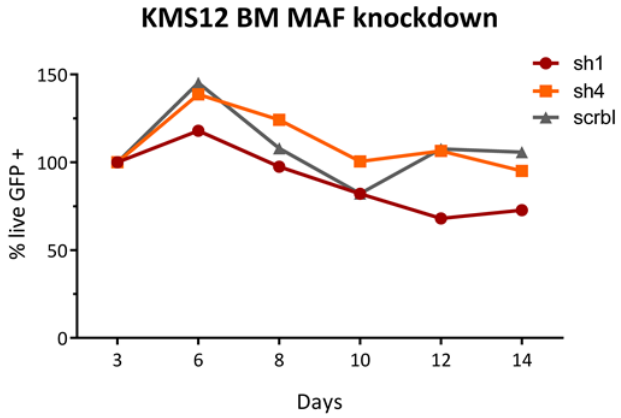


Figure S4: Viability of KMS12 BM MM cells transduced with MAF shRNA

Viability of transduced KMS12 BM cells (t(11;14) MM) over time, as assessed by percentage of live GFP+ cells, normalised to day 3. Percentages represented mean values of n=2 independent experiments.

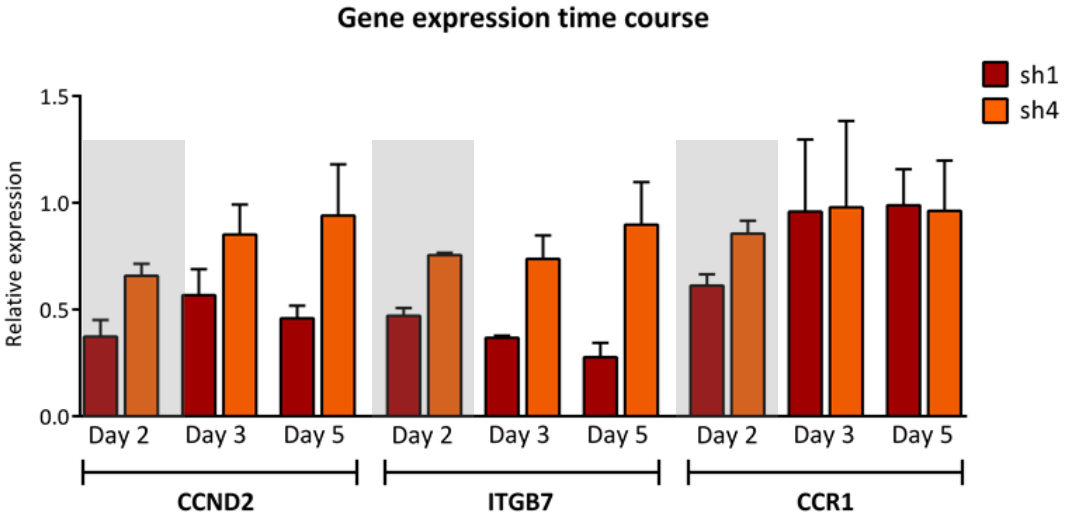


Figure S5: MAF gene target expression over time upon MAF knockdown

Expression of known MAF target genes at different time points following transduction of MM.1S cells with anti-MAF sh1 and sh4. Expression assessed by qPCR and normalised to *GAPDH* and scrbl control. Bars show mean and SD of n=3 experiments.

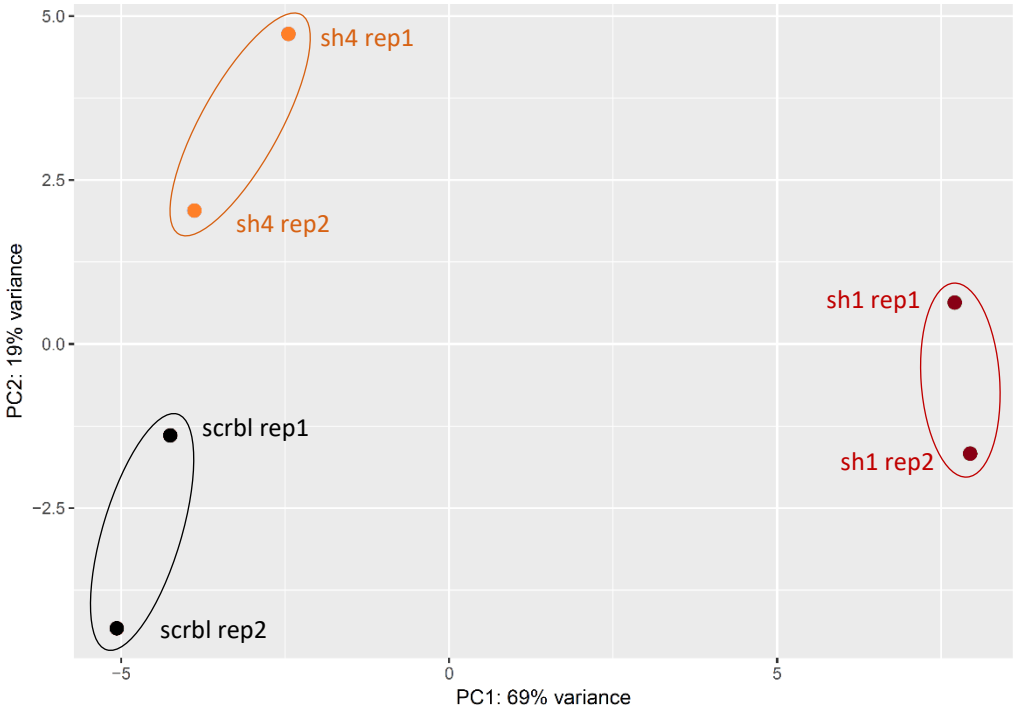


Figure S6: Correlation between MAF knockdown replicates in MM.1S cells

Principal component analysis plot depicting the variance between RNA sequencing samples following MAF knockdown with sh1 and sh4 or scrbl control in MM.1S cells, in two separate replicates.

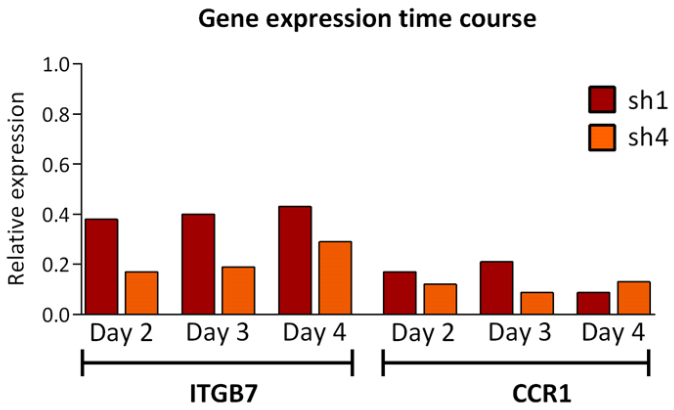


Figure S7: Time course of gene expression in JJN3 following MAF knockdown

Expression of MAF target genes across time following MAF knockdown with two shRNA constructs, sh1 and sh4, as assessed by qPCR. Expression normalised to *GAPDH* and scrbl control.

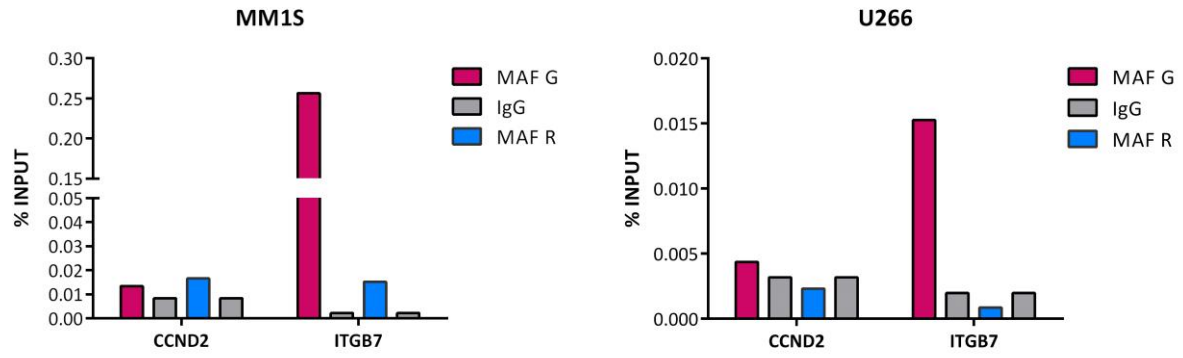


Figure S8: ChIP-qPCR against MAF in MM.1S and U266

ChIP-qPCR against MAF, showing enrichment over IgG control, expressed as % input, in two loci, *CCND2* (exon 5) and *ITGB7* (promoter). Assays were performed in MM.1S, a t(14;16) MMCL (n=2) and U266, a t(11;14) MMCL with no MAF expression (n=1) used as a negative control.

Table S2: Number of peaks and MAF motif analysis score in each of the MM.1S ChIp-seq reactions

ChIP-seq reaction	No of peaks	Motif analysis (MAFA, MAFB, MAFk) family score
MM.1S_MAF R_rep1	155,286	1E-185
MM.1S_MAF R_rep2	8,170	1E-1509
MM.1S_MAF R_low	22,157	1E-1855
MM.1S_MAF G_rep1	111,455	1E-36
MM.1S_MAF G_rep2	33,637	1E-252
MM.1S_MAF G_low	20,572	1E-170

NB: "low" refers to immunoprecipitation reactions performed on $1x10^6$ cells, as opposed to $10x10^6$ cells; rep, replicate

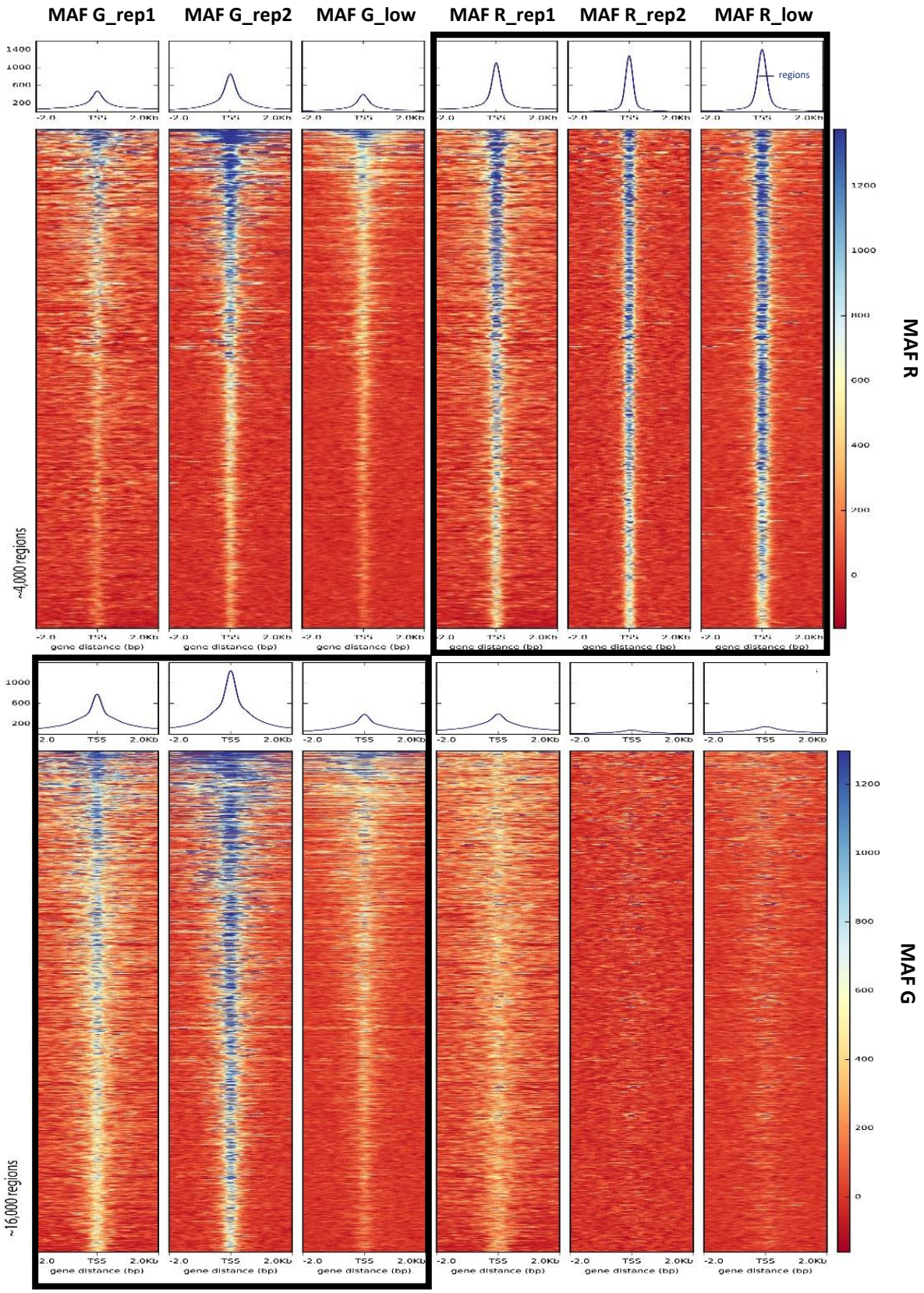


Figure S9: ChIP-seq heatmaps on reactions performed on MM.1S, focusing on peaks common to MAF R reactions (top) or MAF G reactions (bottom).

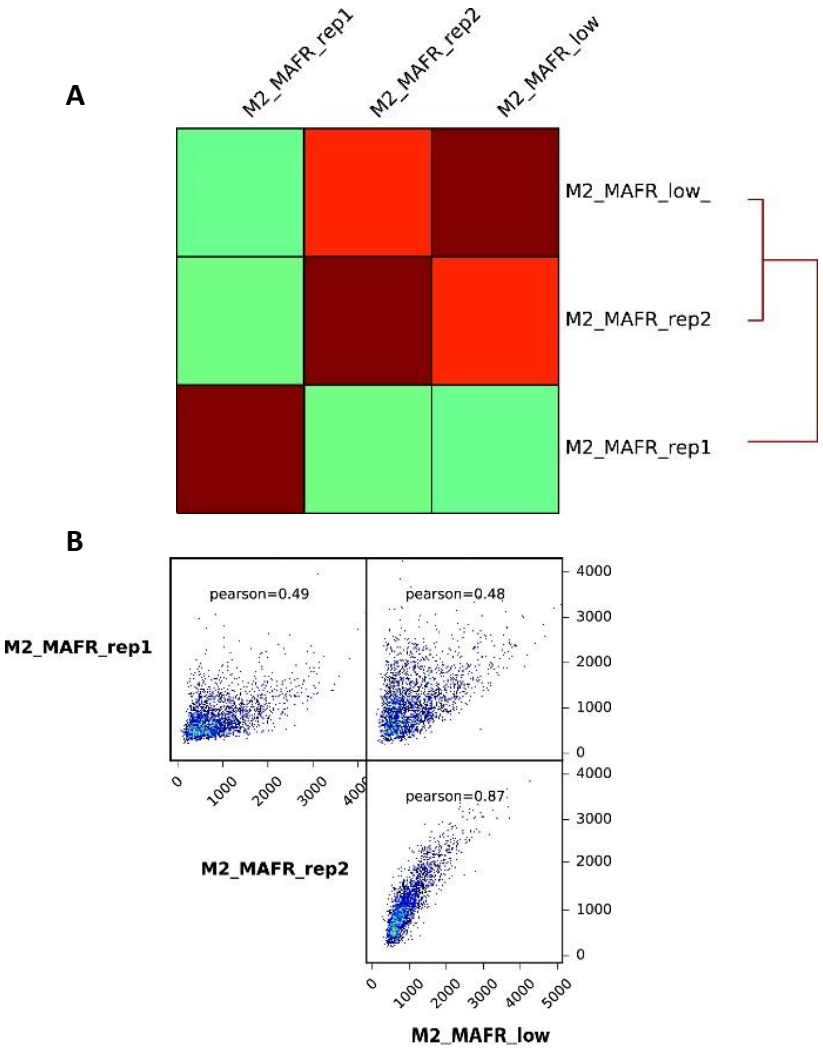


Figure S10: Comparison of ChIP-seq reactions using MAF R antibody in MM.1S cells

Pearson correlation matrix **(A)** and scatterplot **(B)** comparing three MAF R replicate ChIP-seq reactions, MAF R rep1, MAF R rep2 and MAF R low. rep; replicate, low assay performed on $1x10^6$ cells, whereas rep1 and rep2 performed on $10x10^6$ cells. Scatterplots depict enriched ChIP-seq regions identified in each reaction (RPM, normalised to input).

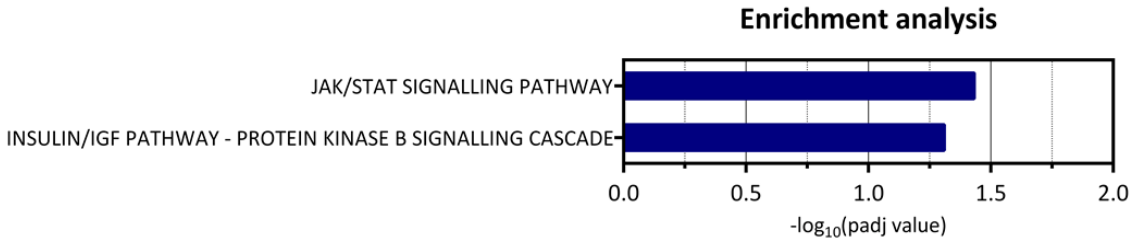


Figure S11: Pathway enrichment analysis of MAF repressed genes

Pathway enrichment analysis of MAF repressed genes in MM.1S cells using the EnrichR online tool (Panther pathways)[105]. Only pathways achieving a padj value<0.05 are shown.

Table S3: List of identified MAF core target genes

(directly upregulated in MMCL and differentially upregulated in MAF primary samples)

CRYBB1	TLR4	SYNDIG1	SMARCA2
ITGB7	ASPHD1	IQGAP2	TMTC1
CCND2	PBXIP1	DOCK4	GADD45A
NUAK1	PCYOX1L	ABLIM1	ANKRD33B
CCR1	RASSF5	GALNT10	RNF165
STMN1	SPECC1	C3AR1	NBPF14
RASSF4	E2F7	HAVCR2	GPRC5D
IGFBP3	STON2	AUTS2	TMPRSS15
ADRB2	LRRFIP1	SHMT1	KREMEN2
TRIM47	CGN	RND3	TMOD1
LCP2	TIMP2	SPTAN1	EEF2KMT
MYO5A	ANKRD39	ENC1	PRRG4
CCDC86	CD82	GPATCH11	MORC1
ABHD12	MGAT5	FMN1	CREG1
CX3CR1	FAM174B	LHFPL2	PARD3B
TOX2	KANK2	MCC	NR3C1
HMOX1	GPRIN1	COTL1	GNAI1
TMEM178B	CDK6	CD200R1	ETS1
INAFM2	SYCP3	SMOX	WSCD1
TYMS	LILRB4	ABHD3	FRMD4B
CAV2	CD28	APBA2	ABCG2
MYO7A	BHLHE22	ANKRD23	PHLPP1
LGMN	PYGB	SYNM	EMC3-AS1
PLA2G15	TUBB2B	PLK4	DGKG
RAP2B	TMEM255A	ATRNL1	
ODC1	FAM107B	RHBDL3	

Appendix D

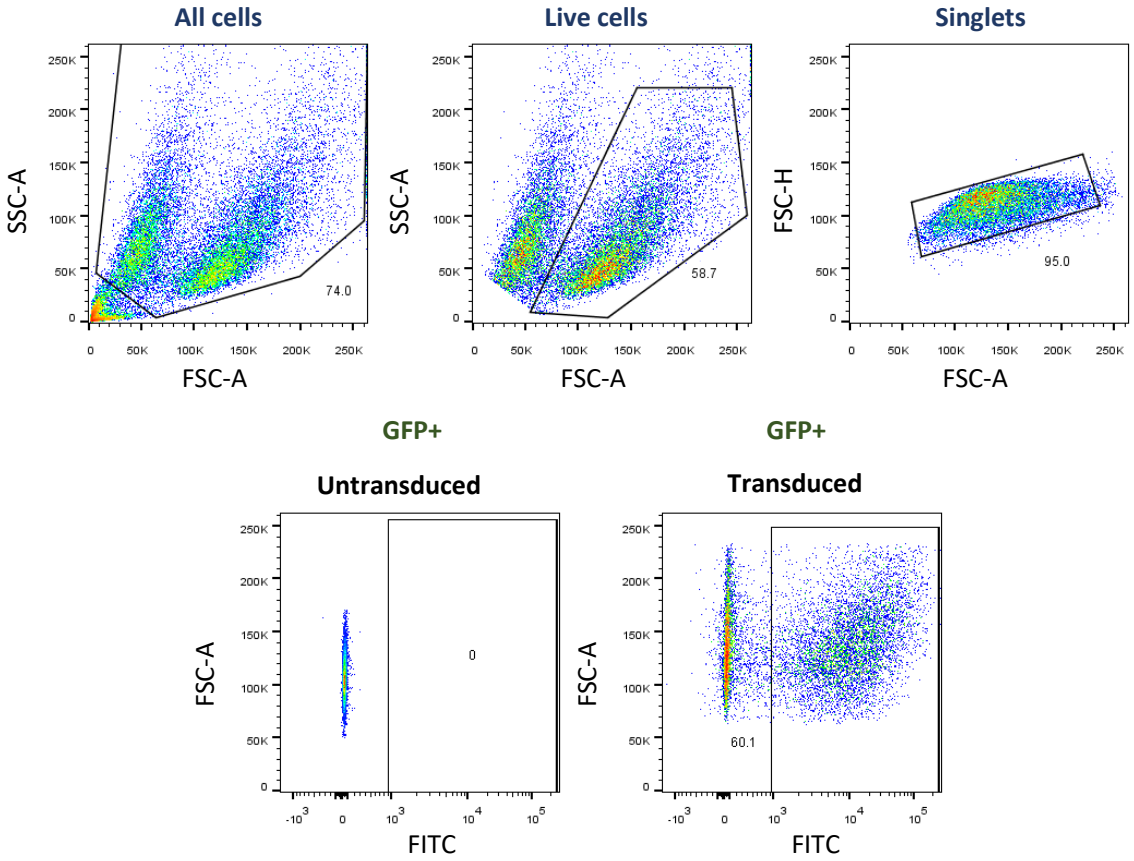


Figure S1: Gating strategy for FACS-sorting MIGR1 GFP+ transduced U266 cells

Representative plots of untransduced U266 and cells transduced with MIGR1 retrovirus used for MAF overexpression experiments, depicting the gating strategy to FACS-sort transduced, GFP+ cells. Cells were sorted on day 5 of transduction, with GFP positivity assessed, as a percentage of live, single cells.

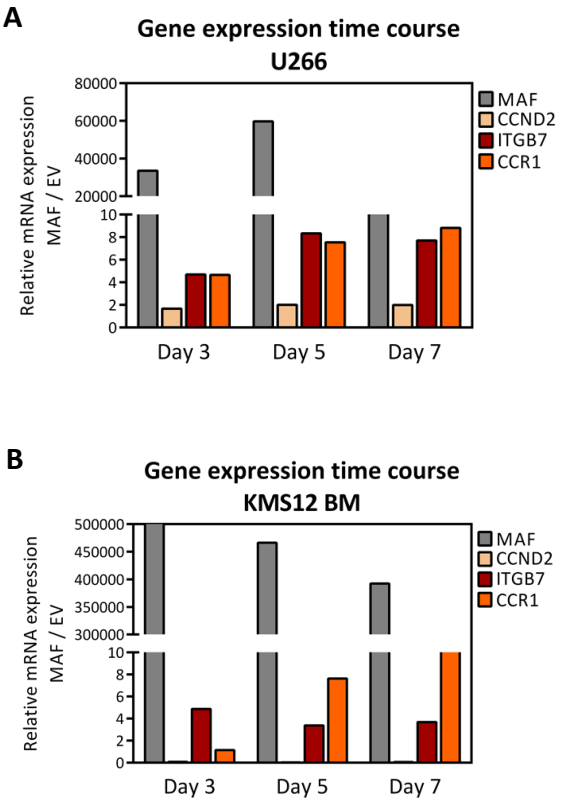


Figure S2: Time course of *MAF* and target gene expression following MAF overexpression

Expression of *MAF* and known target genes at different time points following transduction of U266 **(A)** and KMS12 BM **(B)** with MIGR1 MAF or MIGR1 EV control, as assessed by qPCR. Expression normalised to *GAPDH* and expressed as ratio between MAF and EV control for n=1. EV; empty vector.

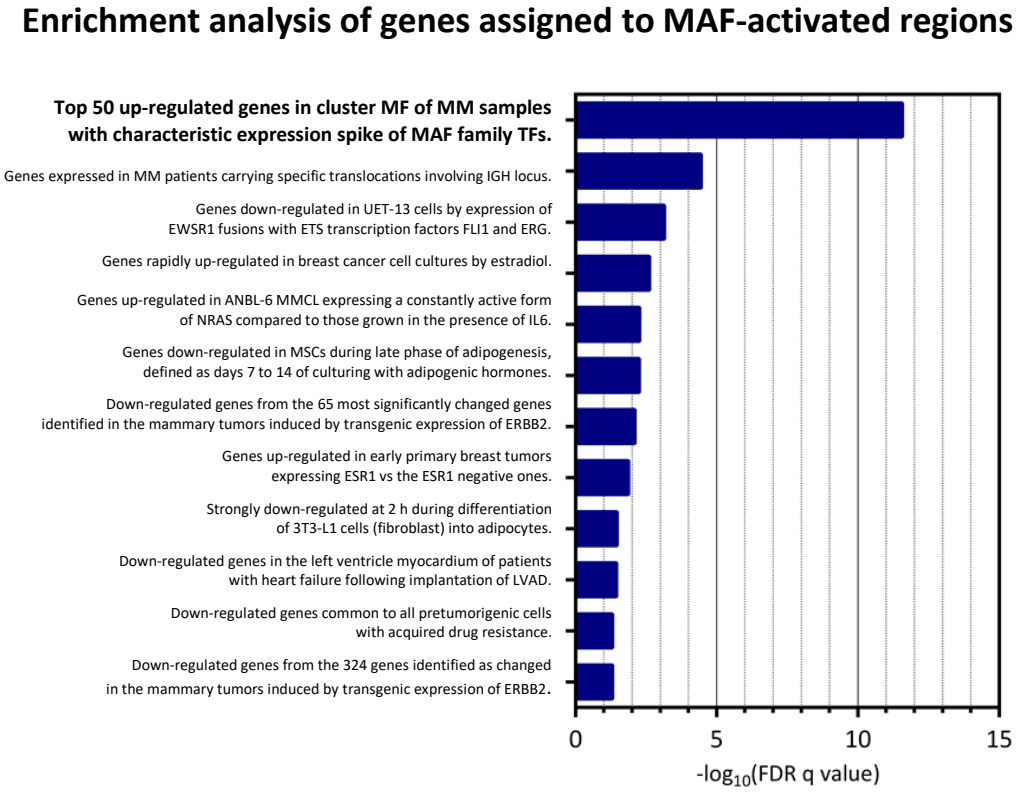


Figure S3: Enrichment analysis of genes assigned to MAF-activated regions with enhanced accessibility following MAF overexpression

Enrichment analysis of the genes assigned to the MAF-activated regions found to have enhanced accessibility following MAF overexpression in the U266 MMCL, using the online tool GREAT[112].

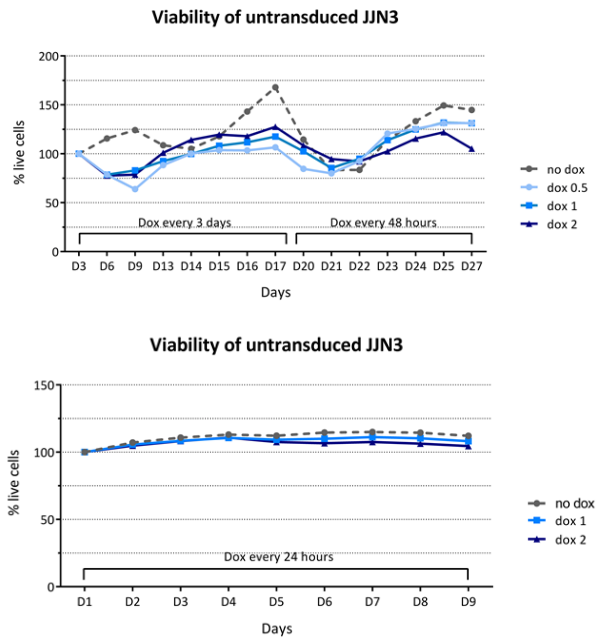


Figure S4: Viability of JJN3 following doxycycline

Percent live JJN3 cells, as assessed by measurements of FSC / SSC in flow cytometry, following application of doxycycline at 0.5µg/ml (dox 0.5), 1µg/ml (dox 1) or 2µg/ml (dox 2), replenished every 48 or 72 hours (top panel) or every 24 hours (bottom panel), across time. Viability of cells is compared to those without doxycycline (no dox) and expressed as a percentage normalised to day 3 (top) or day 1 (bottom).

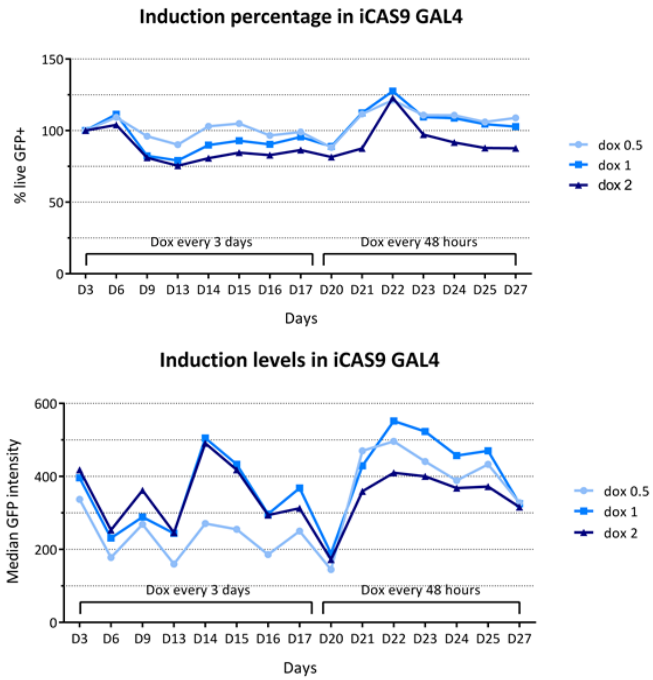


Figure S5: Induction of iCAS9 gal4 with different doxycycline regimens

Percentage of live GFP+ cells (normalised to day 3, top panel) and median GFP intensity (bottom panel) of JJN3 cells transduced with an inducible CRISPR-CAS9 vector with a non-targeting gal4 sgRNA, following induction with doxycycline at 0.5µg/ml (dox 0.5), 1µg/ml (dox 1) or 2µg/ml (dox 2), replenished every 48 or 72 hours.

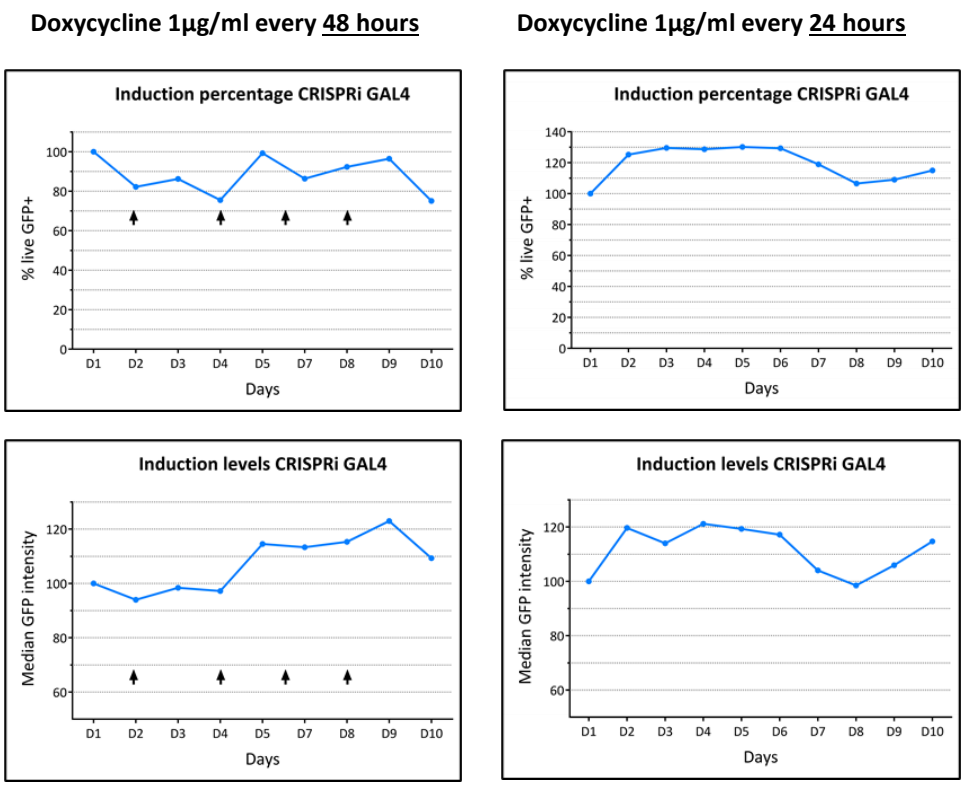


Figure S6: dCas9-KRAB induction at different doxycycline schedules

Percent of live GFP+ cells (top panels) and median GFP intensity (bottom panels) in dCas9-KRAB transduced JJN3 cells with a non-targeting gal4 sgRNA, following induction with 1µg/ml doxycycline applied every 48 hours (left) or daily (right). Arrows indicate application of doxycycline on the 48-hourly regimen. Measurements normalised to day 1.

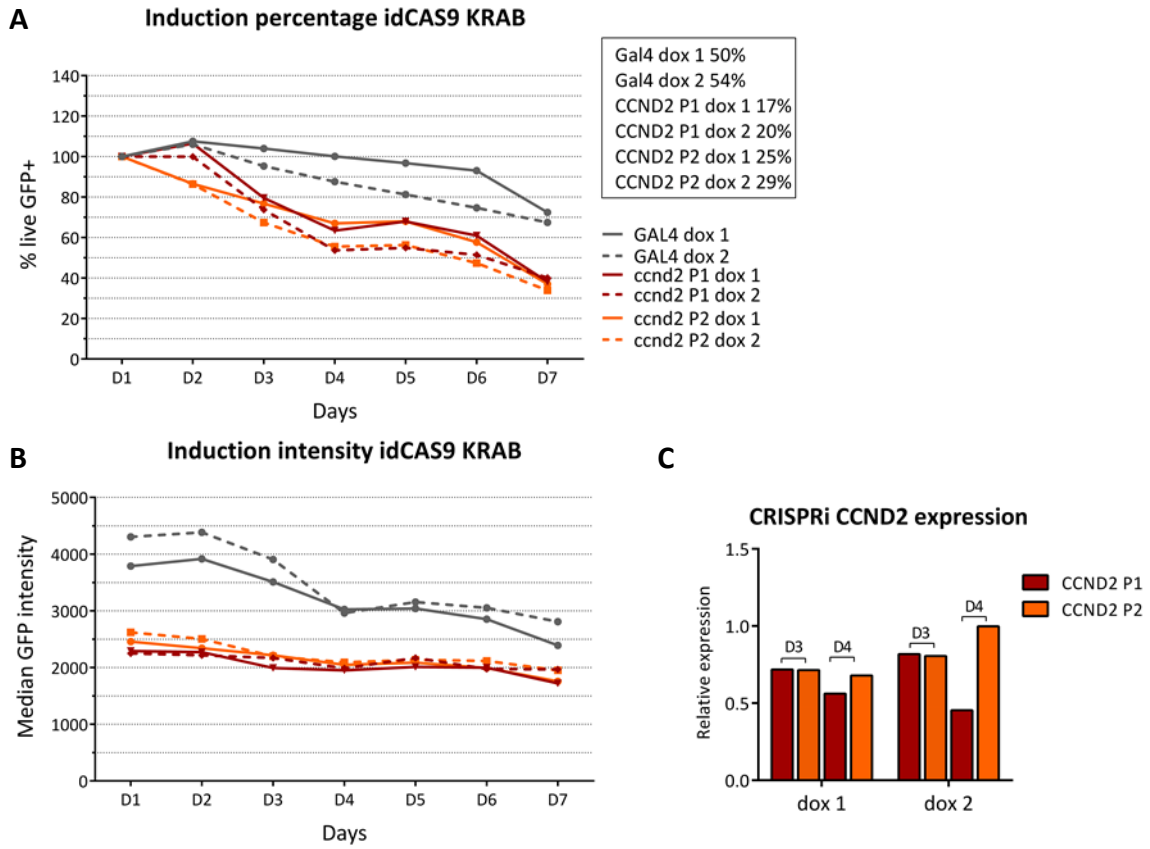


Figure S7: Induction and gene expression levels at different doxycycline concentrations

(A) Percent of live GFP+ cells, transduced with the dCas9-KRAB vector and a sgRNA against the *CCND2* promoter (CCND2 sgRNA P1 or P2) or a non-targeting gal4, following induction with doxycycline at 1µg/ml (dox 1) or 2µg/ml (dox 2). Percentages in the box reflect percent of GFP+ cells on day 1 after induction. **(B)** Median GFP intensity of the same cells following doxycycline induction. Legend for (B) as in (A). **(C)** *CCND2* expression, as assessed by qPCR, on day 3 and day 4 following induction of the above cells with the two different concentrations of doxycycline. Gene expression relative to *GAPDH* and normalised and gal4 control.

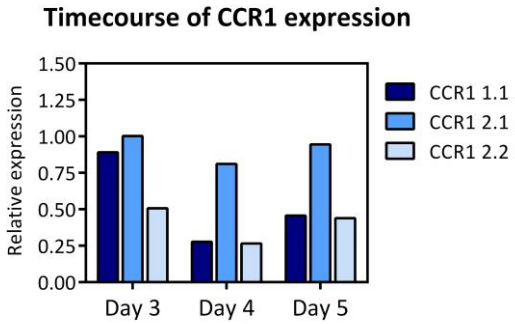


Figure S8: *CCR1* expression following CRISPRi across time

CCR1 expression in cells transduced with 3 different guides against putative *CCR1* enhancer peaks, on days 3, 4 and 5 post doxycycline induction. Expression assessed by qPCR and normalised to *GAPDH* and gal4 control guide.

Table S1:

Transduction (48 hours post puromycin) and induction efficiencies (24 hours post doxycycline) per sgRNA

Guide	*Transduction efficiency*	*Induction % (day 1)*
CCR1 1.1	39.1%	8.6%
		11.1%
		14.7%
CCR1 1.2	42.0%	20.8%
		27%
		19.6%
CCR1 2.1	39.6%	10.7%
		10.4%
		15.1%
CCR1 2.2	33.3%	10.8%
		12.5%
		15.7%
CCR1 3.1	45.9%	20.3%
		30.1%
		20.2%
CCR1 3.2	24.3%	7.1%
		10.4%
		14.9%
CCR1 P1	49.4%	11.4%
		12.4%
		15.5%
CCR1 P2	58.3%	18.1%
		19.8%
		23.9%
GAL4	61.5%	33.0%
		33.4%
		40.0%

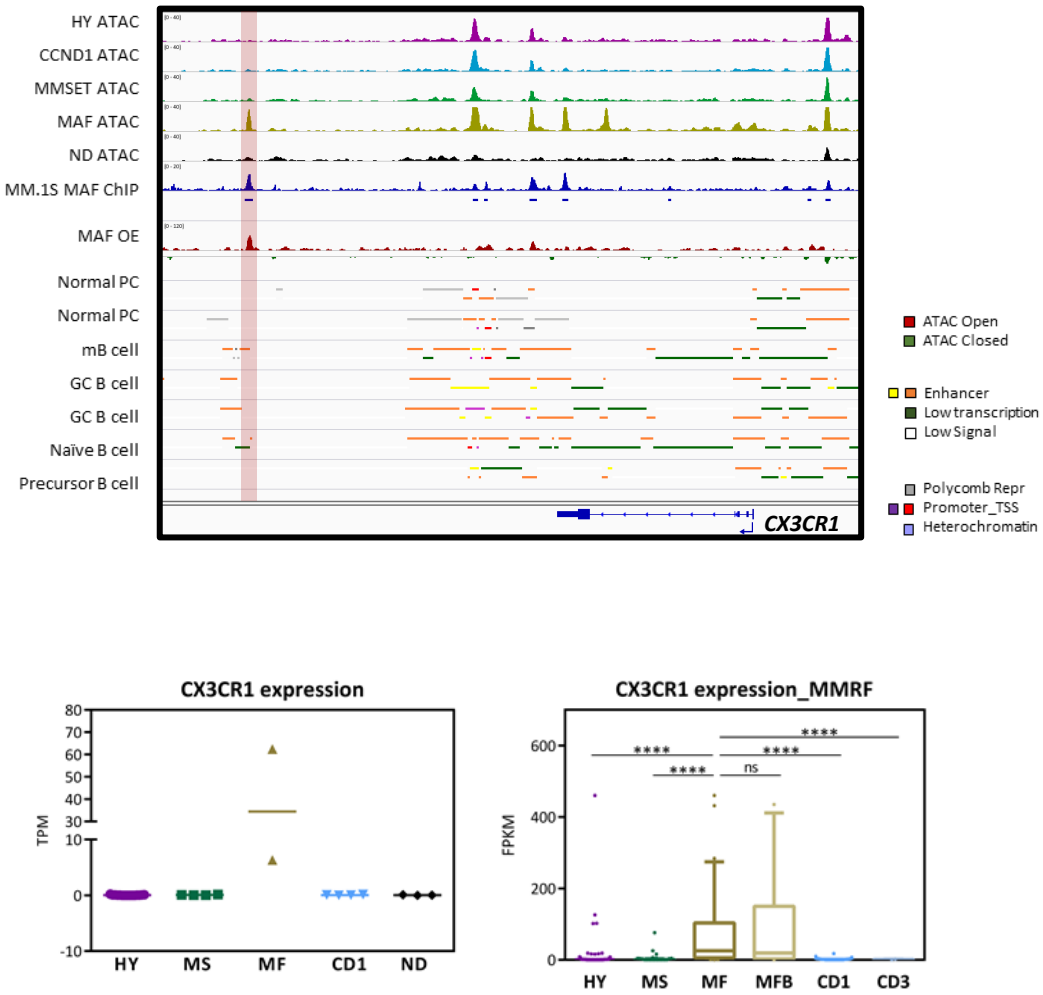


Figure S9: *CX3CR1* "MAF-activated" regulatory region and expression across MM subgroups

(A) IGV plot depicting *CX3CR1* accessibility peaks across MM subgroups and normal plasma cells (average normalised ATAC-seq signal), ChIP-seq against MAF in MM.1S (FC over input), differential accessibility in MAF-overexpression in U266 (MAF OE; MAF vs empty vector) and chromatin mark status across normal plasma cells and B cell development (Blueprint consortium data[115]). Highlighted is the "MAF-activated" peak. **(B, C)** *CX3CR1* expression across MM subgroups in our in-house data set of primary samples (**B**, in TPM) and MMRF Compass data[121] (**C**, FPKM). Box plots depict 10-90th percentile. Statistical analysis performed using a 1-way non-parametric ANOVA test, with Dunn's post-hoc correction, ***p < 0.001, ns; non-significant. HY; hyperdiploid, MS; MMSET, MF; MAF, MFB; MAFB, CD1; CCND1, CD3; CCND3, ND; normal donor plasma cells, mB; memory B cells, CG; germinal centre B cells.

Appendix E

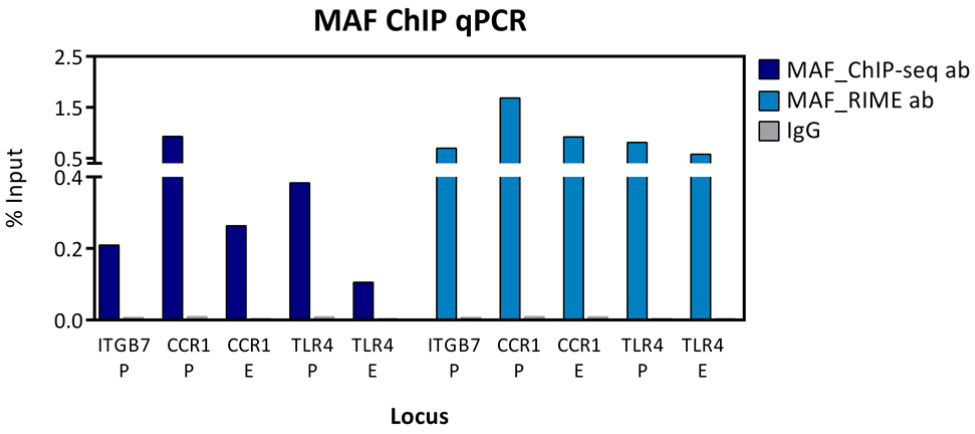


Figure S1: Validation of anti-MAF antibody used for RIME by ChIP-qPCR

ChIP-qPCR against MAF in JJN3 cells, using the MAF antibody used for RIME, in comparison to the antibody used for ChIP-seq. Enrichment in loci of interest expressed as % input and compared to IgG control, bars represent mean of n=2 independent experiments.