

Transcriptional and epigenetic basis for the treatment of inherited disorders of the glycolytic and pentose phosphate pathways using histone deacetylase inhibitor

A thesis submitted for the degree of Doctor of Philosophy

Dr Steven Chika Okoli

May 2022

Centre of Haematology, Faculty of Medicine
Imperial College London

Declaration 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.

Steven Okoli

May 2022

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.

Acknowledgements

I would like to, firstly, thank my supervisor Prof Anastasios Karadimitris for giving me the opportunity to join his lab; being part of this team has been the privilege of my career. Without his support, guidance and expertise, this project would not have been possible.

I would also like to extend my warmest gratitude to my other supervisors, Dr Valentina Caputo and Prof Mark Layton, who were both so patient and generous with their expertise. Both supported me in overcoming the challenges of this novel and complex project and encouraged me to do my very best possible work.

I will be forever grateful to Dr Harry Whitwell for his collaboration in this project. His hard work, advice and expertise were critical to the project's success.

Drs Alexia Katsarou and Nikolaos Trasanidis provided invaluable support to me, and I would like to thank both for their insights and advice at some of the most challenging stages of the work.

Finally, I want to thank my family – Rosie Okoli, Christiana Okoli, Grace Okoli, Christina Okoli and my beloved children Gabrielle and Nathaniel Okoli – for keeping my spirits high at those times of uncertainty that every PhD student will, I am sure, know well. I dedicate my thesis to them.

Thank you.

Funding

This work was funded by a start-up grant from the British Society for Haematology Early-Stage Research; the Blood fund; and Imperial Private Healthcare.

ABSTRACT

Glucose-6-phosphate dehydrogenase is a housekeeping enzyme responsible for the first and rate-limiting catalytic step in the pentose phosphate pathway. This pathway generates NADPH which, in recycling glutathione, neutralises reactive oxygen species a process essential to maintaining oxidative homeostasis.

G6PD deficiency affects approximately four hundred million people worldwide. While the majority are asymptomatic, some experience episodic haemolysis precipitated by physical stressors such as birth, infection and oxidants. A small but significant group with severe (Class I) mutations develop a chronic haemolytic anaemia. Histone acetylase (HDAC) inhibitors selectively enhance transcriptional output of the mutant *G6PD* allele that increase G6PD protein expression and enzyme activity to clinical significance. These findings provide proof-of-concept, but the underlying molecular mechanisms are unclear.

This project employed two complementary chromatin chemo-precipitation techniques to decipher these mechanisms. Click ChIP utilised suberoylanilide hydroxamic acid (SAHA) as a genetic probe to determine its interactors - protein complexes, histone marks, chromatin binding proteins, and other genomic regions - on a genome-wide level. The second technique (CAPTURE) examined the interactions of SAHA on a locus-specific level using modified K562 cells to define the molecular changes that result in the selective induction of *G6PD* activity following treatment with a HDAC inhibitor.

Data from these two techniques indicated that SAHA activates cellular stress response pathways which either directly or indirectly induce G6PD expression. Histone proteomics showed that the *G6PD* promoter exists in a poised state with both activating and repressive

post translational modifications (PTMs) occurring at H3K9 and H4K20 on the same nucleosome. The asymmetric configuration of the nucleosome associated with the *G6PD* locus is unique amongst enzymes of the glycolytic and pentose phosphate pathways, preparing the gene for rapid activation, and that SAHA increases activating PTMs which in turn selectively increases gene expression.

Identifying the critical roles of histones H3K9 acetylation, H3K9 dimethylation and H4K20 dimethylation are important developments in the understanding of the mechanism which underpins the selective induction of *G6PD* by SAHA. This work will potentially lead to the development of more targeted treatments for this and similar disorders.

TABLE OF CONTENTS

Abstract.....	5
List of Abbreviations	14
1 Introduction	17
1.1 Background	18
1.1.1 The glycolytic and pentose phosphate pathways	18
1.1.2 The glycolytic and pentose phosphate pathways in the red blood cell.....	20
1.1.3 Glucose-6-phosphate dehydrogenase	21
1.1.4 Glucose-6-Phosphate Dehydrogenase deficiency	24
1.2 Previous Work.....	31
1.3 Transcriptional regulation.....	34
1.3.1 Chromatin and transcriptional regulation	35
1.3.2 Histone modification.....	37
1.3.3 Acetylation	43
1.3.4 HATs and HDACs	45
1.4 HDAC inhibitors.....	56
1.4.1 Sodium butyrate	59
1.4.2 Suberoylanilide hydroxamic acid (SAHA)	60
1.5 ChIP techniques	64
1.6 Click ChIP.....	65
1.6.1 EnChIP/CAPTURE.....	68
1.7 Aims.....	71
2 Materials and Methods.....	73
2.1 Cell lines	73
2.1.1 Cell culture	74
2.1.2 Large-scale cell culture.....	75
2.1.3 Histone deacetylase inhibitor induction	76
2.1.4 HDAC inhibitor induction assays	77
2.2 Development of CAPTURE cell lines.....	77
2.2.1 Insert design.....	83
2.2.2 Cloning of pU6-sgX-NT1 vector	87
2.2.3 Transformation of competent cells	88
2.2.4 Amplification of pU6 vector	89
2.2.5 Electroporation of plasmids for CAPTURE cell lines	89

2.2.6	Lentiviral production and concentration	90
2.2.7	Cellular transduction.....	92
2.3	Protein expression assays	94
2.3.1	Western blot	94
2.3.2	Co-immunoprecipitation.....	96
2.3.3	Immunofluorescence	97
2.3.4	Antibiotic resistance assay	99
2.4	Gene expression assays	100
2.4.1	RT- qPCR assays.....	100
2.5	Chromatin immunoprecipitation assays.....	102
2.5.1	Click ChIP-seq	102
2.5.2	Click ChIP-MS	111
2.5.3	CAPTURE-seq.....	113
2.5.4	CAPTURE-MS.....	115
2.6	Statistical analysis	119
2.7	Bioinformatic analysis	119
2.7.1	RNA-seq.....	119
2.7.2	Proteomic analysis	119
2.8	Other software.....	120
2.8.1	Online software.....	121
3	Systems design and development	123
3.1	Introduction	123
3.2	Cellular model	124
3.2.1	Establishment of clones	127
3.3	Optimisation of target gene enrichment	130
3.3.1	Chromatin pull down	130
3.3.2	EnChIP to CAPTURE.....	131
3.3.3	Modification of pEF plasmids.....	133
3.3.4	Establishment of CAPTURE cell lines.....	145
3.4	dCas9 interference.....	149
3.4.1	Guide RNAs	149
3.4.2	Chromatin fragmentation	151
3.5	Optimal conditions for selective induction of G6PD expression by SAHA.....	153
3.6	Steric hindrance due to biotin moiety	155
3.6.1	Adoption of Clickable SAHA	157
3.7	Scaling up	159

3.7.1	Measures to limit cell numbers	160
3.7.2	Cell culture	161
3.7.3	Chromatin fragmentation	162
3.8	Summary of process.....	163
4	Click ChIP	166
4.1	Introduction	166
4.1.1	Background	166
4.1.2	Current understanding of the epigenetic effects of SAHA	167
4.2	Experimental design.....	167
4.3	Impact of SAHA on Transcriptomic Landscape	169
4.3.1	Introduction	169
4.3.2	Analysis of gene expression in SAHA treated cell lines.....	169
4.3.3	Differentially expressed genes in SAHA and SAHA BPyne samples	172
4.4	Enrichment analysis of HDAC inhibitor treated cell transcriptome.....	175
4.4.1	Transcriptional and histone modification gene set analysis.....	175
4.4.2	Pathway analysis	177
4.5	The impact of SAHA on cis regulatory elements.....	178
4.5.1	Introduction	178
4.6	Post translational modification of Non-histone protein following treatment with SAHA..	180
4.6.1	Introduction	180
4.6.2	Experimental design.....	180
4.6.3	Pathway analysis	182
4.7	Epigenetic changes associated with SAHA treatment	185
4.7.1	Introduction	185
4.7.2	Experimental design.....	186
4.7.3	H3K9 dimethylation	187
5	CAPTURE	190
5.1	Introduction	190
5.1.1	Background	190
5.1.2	Current understanding of the epigenetic control of G6PD induction	191
5.2	Change in the transcriptomic landscape following SAHA treatment	192
5.2.1	Experimental design.....	192
5.2.2	Effect of SAHA on K562 transcriptome	196
5.2.3	Pathway analysis	198
5.3	The contribution of cis regulatory elements to selective G6PD induction	199
5.3.1	Introduction	199

5.3.2	Experimental design.....	199
5.4	General proteomic changes at G6PD locus following SAHA treatment	201
5.4.1	Introduction	201
5.4.2	Experimental design.....	201
5.4.3	Pathway analysis	204
5.5	Histone post translational changes at G6PD locus following SAHA treatment	206
5.5.1	Introduction	206
5.5.2	Experimental design.....	206
5.5.3	Relative abundance of histone modifications.....	207
6	Discussion.....	214
6.1	Conclusions	227

List of Figures

Table 1.1 WHO Classification of G6PD variants	20
Figure 1.2 G6PD gene head-to-head orientation with NEMO gene	23
Figure 1.3 Examples of common mutations in G6PD coding region	26
Figure 1.4 The structural features of G6PD dimer	26
Figure 1.5 Histone acetylation on GPPP gene promoters of native B cells	33
Figure 1.6 Features of G6PD gene and regulatory control	54
Figure 1.7 Structure of SAHA	61
Figure 1.8 Histone and non-histone targets of SAHA (Vorinostat®) from STRING	62
Figure 1.9A Copper-catalysed Azide-alkyne Click reaction	67
Figure 1.9B Clickable SAHA	67
Figure 1.9C Formation of SAHA-Streptavidin conjugate	68
Figure 1.10 Schematic of enChIP system	69
Figure 2.1 map of pLVhUbc-dCas9T2aGFP lentiviral vector	79
Figure 2.2 pU6-sgx-NT1	80
Figure 2.3 CAPTURE vectors	81
Figure 2.4 Schematic of dCas9-mediated CAPTURE system	83
Figure 2.5 Genetic map of G6PD promoter	85
Figure 2.6 Schematic of pU6-sgG6PD-NT1 vector cloning strategy	88
Figure 2.7 Schematic illustrating transfection of Lenti-X 293T cells with pU6-sgG6PD-NT1	91
Figure 2.8 24-well plate illustrating antibiotic dilutions for resistance assay	100
Figure 2.9 (A-E) Primer optimisation testing using serial dilutions of random DNA	105
Figure 2.10 CAPTURE-MS workflow	106
Figure 3.1 Western blot showing expression profile of 293T, 293T dCas9 and 293T dCas9 g4	126
Figure 3.2 Fold change in G6PD DNA using different ChIP pull down approaches	131
Figure 3.3 pEF1a-BirA-V5-neo → pEF1a-BirA-V5-BFP-virus v7 for viral transduction	134
Figure 3.4 Modification of pEF1a-BirA-V5-neo plasmid to pEF1a-BirA-V5-BFP-virus v7	136
Figure 3.5 Step 1 of pEF1a-FB-dCas9-puro → pEF1a-FB-dCas9-GFP-virus v4 for viral transduction	137
Figure 3.6 Step 2 of pEF1a-FB-dCas9-puro → pEF1a-FB-dCas9-GFP-virus v4 for viral transduction	139
Figure 3.7 Modified pEF1a-FB-dCas9-GFP-virus v4	142
Figure 3.8 Expression profile of modified CAPTURE plasmids by flow-cytometry and western blot	143
Figure 3.9 Genetecin (A) and puromycin (B) toxicity curve for K562 cell line	147
Figure 3.10 In vitro confirmation of expression profile for CAPTURE K562 cell lines	148
Figure 3.11 Comparative induction of inducible and non-inducible GPPP enzymes	150
Figure 3.12 Sonication control confirming optimisation of 2Kb fragments for CAPTURE protocol	152
Figure 3.13 Biotinylated SAHA dose response experiment at 8 hours, normalised to control	154
Figure 3.14 SAHA and biotinylated modifications	155

Figure 3.15 SAHA and tagged modifications	157
Figure 3.16 Optimisation of reaction conditions for selective HDAC inhibitor induced G6PD upregulation	158
Figure 3.17 Sonication control confirming effective and uniform fragmentation of chromatin	162
Figure 4.1 Clustering of differentially expressed genes in SAHA and SAHA BPyne treated K562 cells	170
Figure 4.2 Click ChIP PCA plot	171
Figure 4.3 Illustrates the proportion of shared genes within the SAHA and SAHA BPyne datasets	172
Figure 4.4 Volcano plot highlighting the expression profile of SAHA BPyne treated samples	173
Figure 4.5 Top 50 overrepresented genes in the SAHA BPyne treated samples	174
Figure 4.6 Enrichment analysis of the differentially expressed genes in SAHA BPyne dataset using Enrichr ...	175
Figure 4.7 RT-PCR to confirm enrichment of G6PD expression following Click ChIP pull down	179
Figure 4.8 Selectively expressed proteins in Click ChIP general proteomics - molecular function	182
Figure 4.9 Selectively expressed proteins in Click ChIP general proteomics – protein class	183
Figure 4.10 Click ChIP general proteomics data comparing SAHA BPyne with vehicle	184
Figure 4.11 Pathway analysis performed on Click ChIP general proteomics dataset	185
Figure 5.1 Selective induction of G6PD expression following treatment with SAHA	192
Figure 5.2 Correlation between CAPTURE RNA seq in K562 cells treated with DMSO/SAHA	193
Figure 5.3. Differentially expressed genes of SAHA treated cells when compared to vehicle.....	194
Figure 5.4 Heat map of gene expression in SAHA and vehicle treated K562 cells	195
Figure 5.5. Enrichment analysis of differentially expressed genes in SAHA and vehicle treated K562 cells	196
Figure 5.6 Pathway analysis of differentially expressed genes in SAHA and vehicle treated K562 cells	198
Figure 5.7 RT-PCR performed on purified DNA using CAPTURE seq protocol	200
Figure 5.8 Heat map illustrating differential expression of proteins in CAPTURE proteomics	202
Figure 5.9 Correlation between CAPTURE proteomics conditions in K562 cells treated with DMSO/SAHA	203
Figure 5.10 Selected pathway analysis of the differentially abundant proteins	204
Figure 5.11 String network illustrating the protein interactions in CAPTURE	205
Figure 5.12 Mass spectra of most frequent histones modified affecting histones H3 and H4	207
Figure 5.13 Relative abundance of histone modifications in CAPTURE	209
Figure 5.14 Correlation between CAPTURE histone proteomics conditions in K562 cells	210
Figure 5.15 Proportion of lysine residues modified in different CAPTURE cell lines	211
Figure 6.1 Relationship between various cellular stress response pathways	219
Figure 6.2 Density graphs for key features that influence G6PD induction by SAHA	224

List of Tables

Table 1.1 WHO Classification of G6PD variants	27
Table 1.2 Drugs and chemicals associated with haemolysis in G6PD deficient subjects	29
Table 1.3 Chromatin-binding domains in known HAT complexes	48
Table 1.4 Classification and activity of HDAC inhibitors	57
Table 1.5 Non-histone targets for HDAC inhibitors	63
Table 2.1 Cell lines	74
Table 2.2 sgRNA specific pU6-sgX-NT1 primers for inserts	86
Table 2.3 Optimised electroporation parameters used to transduce cell lines with pEF1a-BirA-V5-neo	90
Table 2.4 Calcium phosphate methods for virus production	92
Table 2.5 Antibody dilutions for western blotting	96
Table 2.6 Antibody dilutions for Immunofluorescence	98
Table 2.7 List of TaqMan probes	101
Table 2.8 List of ChIP probes	109
Table 3.1 Components of pEF1a-BirA-V5-BFP and associated primers	135
Table 3.2 Primers and vectors used in first step of pEF1a-FB-dCas9-GFP-virus v4 construction	138
Table 3.3 Primers and vectors used in second step of pEF1a-FB-dCas9-GFP-virus v4	141
Table 3.4 Electroporation parameters	146

LIST OF ABBREVIATIONS

Name	Abbreviation
AKT Serine/Threonine Kinase 1	Akt
Blue fluorescent protein	BFP
Base pairs	bp
Bovine serum albumin	BSA
CRISPR affinity purification <i>in situ</i> of regulatory elements	CAPTURE
Chromatin immunoprecipitation	ChIP
ChIP followed by real time polymerase chain reaction	ChIP-qPCR
ChIP followed by DNA sequencing	ChIP-seq
Chronic non-spherocytic haemolytic anaemia	CNSHA
Chronic myeloid leukaemia	CML
Copper ion	Cu (I)
Clustered regularly interspaced palindromic repeats	CRISPR
Deactivated Cas9	dCas9
Dimethyl sulfoxide	DMSO
Dithiothreitol	DTT
Dulbecco's Modified Eagle Medium	DMEM
Ethylenediaminetetraacetic acid	EDTA
Encyclopaedia of DNA Elements Consortium Project	ENCODE
Fetal Bovine Serum	FBS
Fold change	FC
False discovery rate	FDR
Fetal Bovine Syndrome	FBS
Glucose-6-phosphate dehydrogenase	G6PD
Glutathione	GSH
Glycolytic and pentose phosphate pathways	GPPP
Glycosylphosphatidylinositol	GPI
Green fluorescent protein	GFP
Acetylation of histone H3 at 9 th lysine residue	H3K9ac
Dimethylation of histone H3 at the 9 th lysine residue	H3K9me2
Trimethylation of histone H3 at 9 th lysine residue	H3K9me3
Trimethylation of histone H3 at 27 th lysine residue	H3K27me3
Dimethylation of histone H4 at 20 th lysine residue	H4K20me2
Heat shock protein	HSP
Histone acetyltransferases	HAT
Histone deacetylases	HDAC
Kilobases	Kb

Liquid chromatography-tandem mass spectrometry	LC-MS/MS
Neonatal jaundice	NNJ
Nicotinamide adenine dinucleotide phosphate	NADP ⁺
Phenylmethanesulfonyl fluoride	PMSF
Phosphate buffered saline	PBS
Phosphoglycerate kinase	PGK
Polycomb repressive complex 2	PRC2
Post translational modifications	PTM
Quantitative real time PCR	qPCR
Reactive Oxygen Species	ROS
Red blood cell	RBC
Reduced Nicotinamide Adenine Dinucleotide Phosphate	NADPH
RNA integrity number	RIN
RNA sequencing	RNA-seq
Roswell Park Memorial Institute Medium	RPMI
Room temperature	RT
Standard deviation	SD
Sodium dodecyl sulphate	SDS
Standard Error of the Mean	SEM
Single guide RNA	sgRNA
Specificity protein 1	Sp1
Specificity protein 3	Sp3
Spt-Ada-Gcn5-acetyltransferase	SAGA
Suberoylanilide hydroxamic acid	SAHA
Transcription factor	TF
Transcription start site	TSS
Woodchuck Hepatitis Virus Post-transcriptional Response Element	WPRE

Chapter One

Introduction

1 INTRODUCTION

Glucose-6-phosphate dehydrogenase (G6PD) catalyses the committed step of the pentose phosphate pathway, a pathway critical to neutralising reactive oxygen species (ROS) to maintain oxidative homeostasis (1). Deficiency in this enzyme limits available glutathione, destabilising the equilibrium, leading to an accumulation of ROS.

The mature red cell is uniquely susceptible to ROS accumulation as an enucleated cell is unable to generate additional G6PD. This build-up of toxic metabolites (bilirubin and free haem) causes haemolysis, and results in anaemia. Although G6PD deficiency is extremely common, for the majority it remains asymptomatic, unless exposed to oxidative stress. However, a small cohort of those with the severe form of the disease constantly haemolyse and suffer significant sequelae as a result.

Currently, treatment of haemolytic crises due to G6PD deficiency are limited to supportive therapies (2). Previous work in the Karadimitris lab identified HDAC inhibitors as a potential therapeutic intervention for this disorder. This project aims to identify the transcriptional and epigenetic mechanisms responsible for the selective induction of *G6PD* by suberoylanilide hydroxamic acid (SAHA) with a view to developing targeted pharmacological agents for safe and effective restoration of G6PD activity.

1.1 BACKGROUND

1.1.1 The glycolytic and pentose phosphate pathways

Most cellular pathways require multiple enzymatic steps that usually occur in sequence, i.e. the product of one reaction becomes the substrate of the next enzyme in the process. Some have evolved to form networks where such that an intermediate from one pathway becomes the substrate for another dependant pathway (3). An example of this is the pentose phosphate pathway: this pathway is reliant on an intermediate from the glycolytic pathway, glucose-6-phosphate, as substrate for the initiating step of the shunt (4).

The glycolytic and pentose phosphate pathways (GPPP) are ubiquitously active. The former responsible for anaerobic generation of ATP and the latter is the main cellular source of NADPH which is utilised in biosynthetic reactions and oxidative defence (5, 6). These pathways are intrinsically linked, and rate limited by the catalytic activity of G6PD, **Figure 1.1** (7).

The glycolytic pathway (Embden–Meyerhof–Parnas pathway) is one of the most conserved metabolic pathways (8). Discovered by Gustav Embden, Otto Meyerhof, and Jakub Karol Parnas in 1940 (9), this pathway is responsible for the conversion of glucose to pyruvate releasing energy which is stored as ATP for later use. This cytosolic cascade comprises ten sequential enzyme-catalysed reactions, **Figure 1.1**, generating intermediates that may be utilised by other pathways as well as being essential for the overall reaction. This is well illustrated by glucose-6-phosphate generated following the first step of the pathway and the substrate of G6PD. Glycolysis is an oxygen-independent metabolic pathway. This fact is

crucial to the mature red blood cell as they lack most organelles (most notably mitochondria) essential for all oxygen dependant ATP genesis (10, 11). In the absence of a mechanism to remove oxygen metabolites the red cell haemolyses.

The pentose phosphate pathway performs the other critical function necessary to maintain red cell integrity: neutralising oxygen free radicals. This parallel pathway to glycolysis generates NADPH, pentoses and ribose-5-phosphate, a precursor for nucleotide synthesis (3). It has two distinct phases: the first or oxidative phase, occurs in the cytosol and is responsible for generating NADPH. This is followed by the non-oxidative synthesis of pentose sugars. An essential by-product of the reaction is the reduction of NADP^+ which regenerates glutathione providing the cell's reducing power critical for neutralising ROS (7, 12). Under physiological conditions, the glycolytic pathway metabolises about 90% of glucose and only 10% enters the pentose shunt. However, in cases of oxidative stress, the contribution of the pentose phosphate pathway is significantly increased.

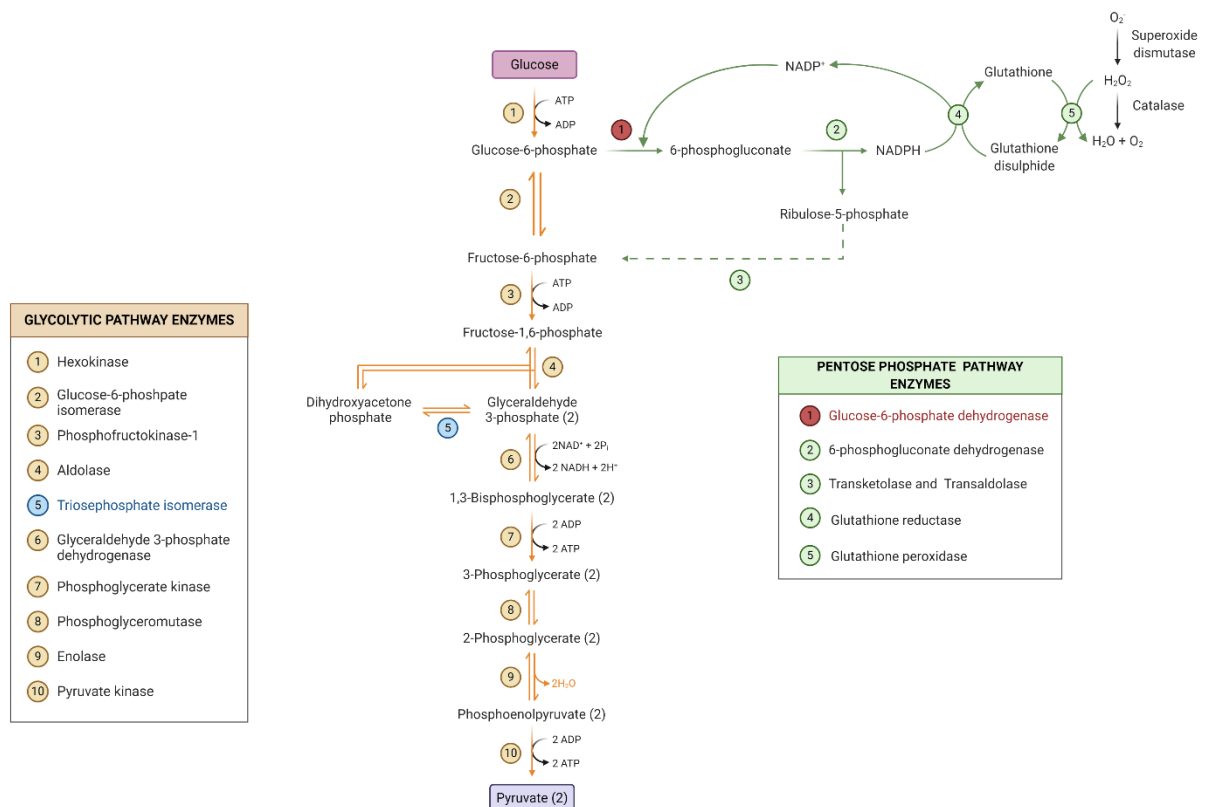


Figure 1.1 Glycolytic and Pentose phosphate pathways

Schematic depiction of the ten enzymes of the Embden–Meyerhof–Parnas pathway of glycolysis (glycolytic pathway) and five enzymes of the pentose phosphate pathway. Glucose-6-phosphate dehydrogenase, in red, and Triosephosphate isomerase, in blue. Figure created in BioRender by author.

(Adapted from Vives Corrons, Red blood cell enzyme defect, European School Haematology, 2009. 17: p.436-453 (13) and Luzzatto L & Aress P, Favism and Glucose-6-dehydrogenase deficiency, N Engl J Med, 2018. 378 (11): p. 1068-1069 (14).

1.1.2 The glycolytic and pentose phosphate pathways in the red blood cell

The glycolytic and pentose phosphate pathways have critical roles in maintaining the integrity and function of the mature red blood cell (RBC) (15). These cells have lost their nuclei and mitochondria so are incapable of protein synthesis, and are therefore reliant on organelle independent metabolic processes. To maintain cellular redox balance, a source of NADPH is essential. Thus, the RBC, devoid of essential organelles, relies solely on G6PD

which is the only NADPH-producing enzyme that can be activated in response to oxidative stress. As such, G6PD acts as the guardian of the red cell redox potential (4, 5). If NADPH concentrations cannot be maintained – as in G6PD deficiency – the GSH levels fall, and oxidative damage occurs ultimately resulting in haemolysis. In normal red cells, more than 90% of the total glutathione pool is in the reduced form. Following oxidative stress, free radicals accumulate, these are scavenged by glutathione generating glutathione disulphide which protects the sulphhydryl groups in proteins such as haemoglobin and in the red cell membrane from oxidation, **Figure 1.1**.

Glutathione is restored by the action of glutathione reductase, in the presence of NADPH. In most cells other NADP-dependent reactions provide a substantial proportion of the NADPH required under normal conditions.

1.1.3 Glucose-6-phosphate dehydrogenase

Cytosolic G6PD catalyses the first and rate-limiting step of the oxidative branch within the pentose phosphate pathway. This is an alternative route to glycolysis for the dissimilation of carbohydrates, a major source of reducing power, and metabolic intermediates for fatty acid and nucleic acid biosynthetic processes.

1.1.3.1 Biochemistry

G6PD is a housekeeping enzyme, present in all forms of life (16, 17). It performs a key regulatory role in the pentose phosphate pathway (PPP), promoting oxidation of glucose-6-

phosphate to 6-phosphogluconate which initiates a cascade leading to the reduction of NADP⁺ to NADPH (18).

The G6PD monomer is 515 amino acids in length; the active form of the enzyme is a homodimer or tetramer (2), depending on cellular conditions (19). Dimer structures of the two subunits are symmetrically located across a complex interface of β -sheets with each subunit bound to a NADP⁺ molecule for stability (20). Each monomer in the complex has two molecules of NADP⁺ and has a substrate binding site for Glucose-6-phosphate between them (21). NADP⁺ first binds in the N-terminal domain acting as a co-substrate for the metabolic reaction, while the second is bound to the C-terminal domain and functions as a structural element necessary to maintain the long-term stability of the enzyme (22).

1.1.3.2 Genetics

The *G6PD* gene is located near the telomeric region of the long arm of the X chromosome, chrX:154,531,390-154,547,572. The human *G6PD* gene forms a cluster with other genes that are associated with disorders - fragile X (23), colour blindness (24), haemophilia A (25), and dyskeratosis congenita (26). The gene was cloned in 1986 independently by Persico *et al.* and Takizawa *et al.* (27-29). It spans 18.5Kb, 13 exons and 12 introns, two alternative transcription start sites (TSS) and encodes two variants: a short variant that consists of exons 2-13 and a longer inactive variant consisting of all 13 exons (29, 30). In the short variant, the first 600bp of the 5' end of the gene, which correspond to exon 1 and part of exon 2, are not translated (31).

The *G6PD* gene locus is proximally arranged in a head-to-head orientation with *NEMO* gene separated by no more than 800bp (32), **Figure 1.2**. The *G6PD* gene is transcribed in the telomeric direction, partly overlapping *NEMO*, which is transcribed in the centromeric direction. The genes share a conserved promoter region that has bidirectional housekeeping activity. In addition, intron 2 of the *G6PD* gene contains an alternate promoter for *NEMO* only (33, 34). *NEMO*, NF kappa B essential modulator, alternatively named *IKK gamma* gene encodes the regulatory subunit of the inhibitor of kappa B kinase (IKK) complex, which activates NF-kappa B resulting in activation of genes involved in inflammation, immunity, cell survival, and other pathways (35, 36).

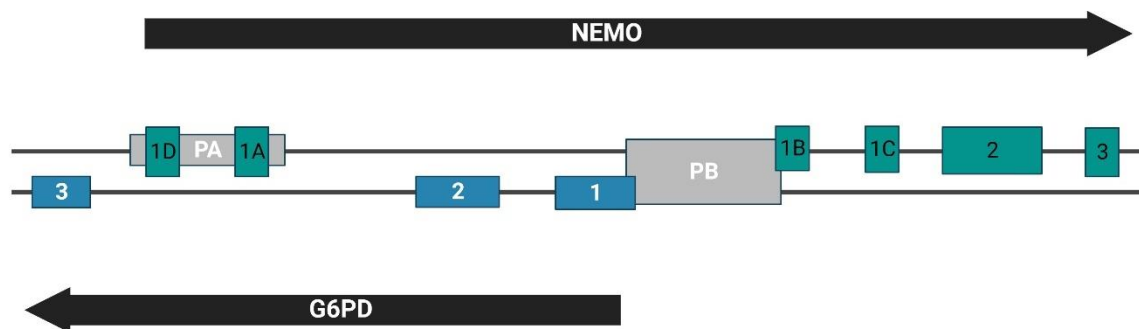


Figure 1.2 G6PD gene head-to-head orientation with NEMO gene

Schematic representation of genomic structure of *G6PD* gene (blue), *NEMO* gene (green), and promoters (grey). Promoter A (PA) is unidirectional *NEMO* promoter and found in *G6PD* intron 2, whereas promoter B (PB) is bidirectional, impacting both DNA strands, and located between the *G6PD* and *NEMO* genes.

Schematic created using the UCSC (University of California, Santa Cruz Genomics Institute) genome browser, GRCh38/hg38 genome. Figure created in BioRender by author.

(Adapted from Fusco F, et al., Genomic architecture at the Incontinentia Pigmenti locus favours de novo pathological alleles through different mechanisms, *Human Molecular Genetics*, 2012. 21 61: p. 1260-1271 and Zinatizadeh M, et al., A Review of NEMO Protein and its Relationship with Genetic Diseases, *Journal of Genetics and genome Research*, 2019. 5:033 p. 1-6).

1.1.4 Glucose-6-Phosphate Dehydrogenase deficiency

1.1.4.1 History

The earliest reports of G6PD deficiency date back to ancient Greece when philosopher Pythagoras forbid his students from consuming fava beans (37). He had deduced that these beans had lethal potential, although at the time he did not understand exactly why.

Almost two millennia later, in 1926, G6PD deficiency was first identified when investigating a group of patients who developed a haemolytic anaemia while being treated for malaria with 6-methoxy-8-aminoquinoline drugs (38). However, it was to be several decades and the convergence of numerous events before the mechanism for this drug-induced haemolysis was identified.

Favism is characterised by life-threatening haemolysis triggered by vicine and convicine when a G6PD-deficient individual ingests raw or partially cooked fava beans. When hydrolysed, these β -glucosides release highly reactive oxidising compounds, divicine and isouramil (39, 40).

1.1.4.2 Epidemiology

G6PD deficiency is the second most common human enzyme defect, affecting over 400 million people worldwide (41). Fortunately, the majority of G6PD deficient subjects have no clinical manifestations and remain asymptomatic unless exposed to an oxidative trigger. Over 4000 deaths occurred in 2013, and the incidence is rising (42). As an X-linked recessive

disorder, there is a male preponderance, although women can be affected. The highest frequencies of the deficiency are detected in Africa, Middle East and South Asia (43). The 'malaria protection theory' postulates that G6PD deficiency confers protection against malaria, in particular the deadliest form of the infection caused by *Plasmodium falciparum* (44). So, the prevalence of G6PD deficiency is remarkably high in parts of the world where this powerful selection pressure is endemic. Unsurprisingly, the most common G6PD variants are found in people who originated from these malarious regions: the Mediterranean countries (the Mediterranean type; G6PD Med), Sub-Saharan Africa (the African type; G6PD A-) and parts of India and South-East Asia (45-47), **Figure 1.4**.

1.1.4.3 G6PD deficiency and red blood cells

G6PD deficiency has a particularly deleterious effect on RBCs for two reasons. First, most known mutations cause a decreased enzyme stability so, as cells age, enzyme levels fall prematurely. As the cell approaches 120 days, the end of its lifespan, G6PD enzymatic activity will have decreased by more than eight-fold (48). Second, red cells are uniquely susceptible to oxidative stress from exogenous oxidizing agents in the blood. In addition, oxygen radicals are continuously generated as haemoglobin cycles between its deoxygenated and oxygenated forms. Due to these factors the deficient red cell has a diminished capacity to withstand oxidative stress, becomes overwhelmed and haemolyses.

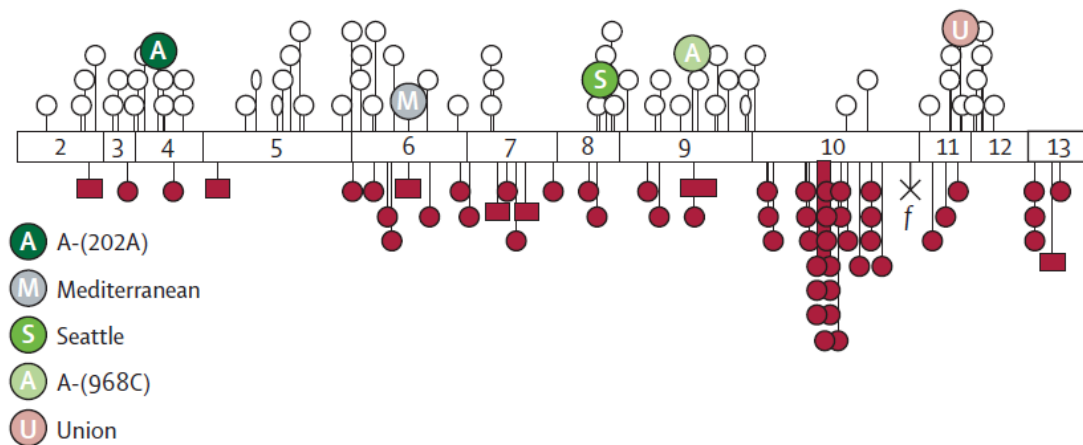


Figure 1.3 Examples of common mutations in G6PD coding region

Exons 2-13 are represented in numbered boxes. Open circles are class II and III variants. Coloured circles represent class I/severe mutations. New variants continue to emerge.

Note: G6PD-A variants has two sites of base substitution at 202A and 968C.

(Adapted from Cappellini M D & Fiorelli G, Glucose-6-dehydrogenase deficiency, 2008. *Lancet*, 371: 64-67 (49) and Gómez-Manzo, et al. Functional and biochemical analysis of glucose-6-phosphate dehydrogenase (G6PD) variants: elucidating the molecular basis of G6PD deficiency, 2017. *Journal of Catalysts*, 7: 135 (50)).

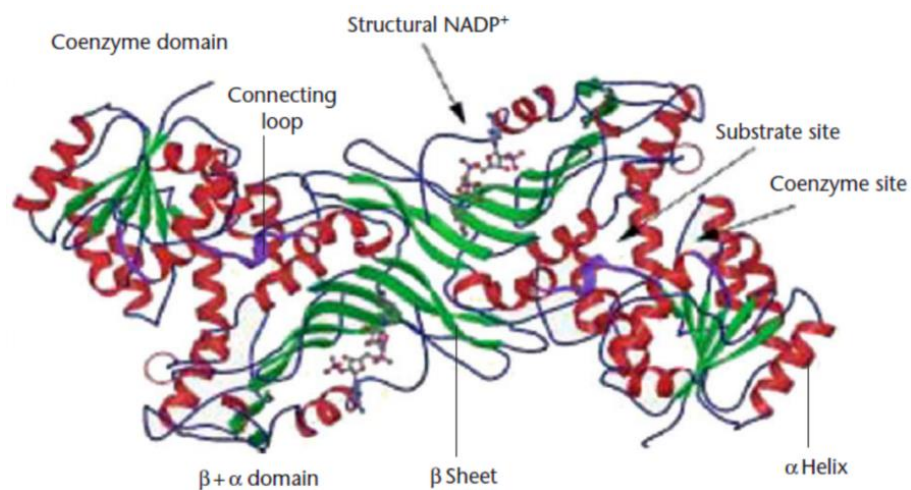


Figure 1.4 shows the structural features of G6PD dimer

(Mason P J & Vulliamy T J, Glucose-6-dehydrogenase (G6PD) deficiency: genetics. *Encyclopedia of Life Sciences*, John Wiley & Sons, Ltd., 2010 (51)).

1.1.4.4 Classification and clinical manifestations

Over 450 protein variants have been characterised (41, 52), the majority of which affect the stability of the protein rather than the catalytic activity of the enzyme (41, 53, 54). They are categorised into five classes according to enzyme activity and clinical manifestations in the affected individuals, **Table 1** (55). These variants were subsequently categorised on a molecular level, and many were found to describe identical genetic mutations (56).

Class	Residual G6PD activity (% of normal)	Clinical manifestation
I	<10	Chronic non-spherocytic haemolytic anaemia (neonatal jaundice, acute exacerbation)
II	<10	None at steady state
III	10-60	None at steady state
IV	100	None
V	>100	None

Table 1.1 WHO Classification of G6PD variants

(Adapted from Roper D, et al., Laboratory diagnosis of G6PD deficiency. British Society for Haematology Guidelines, British Journal of Haematology, 2020. 189: p24-38 (57)).

The *G6PD* gene resides in a highly variable region of the genome giving rise to marked genetic, biochemical and clinical heterogeneity. Also noteworthy, is that almost half of the mutations are associated with the most severe phenotypes but are very rare (51).

Therefore, of the 217 mutations described (31, 58) to date, only a few account for the majority of cases (59). These mainly affect the coding region while none are found in the promoter region. 182 (83.9%) are single nucleotide substitutions, 19 (8.7%) are multiple

mutations (two or more substitutions), 11 (5.1%) are deletions, and 5 (2.3%) are intronic mutations (31). Most of these mutations result in a decrease in enzyme stability or reduced catalytic efficiency (54). The degrees of enzymatic dysfunction detected in variants correlates well with the severity of the disease (50, 60).

The clinical manifestations of G6PD deficiency have a broad spectrum ranging from asymptomatic individuals to those who suffer acute haemolytic episodes, severe neonatal jaundice and chronic non-spherocytic haemolytic anaemia (CNSHA) (6, 52, 61). The most common clinical manifestations are neonatal jaundice (NNJ) and acute haemolytic anaemia triggered by ill health or exogenous oxidants (6, 44). G6PD deficient neonates are seven times more likely to develop kernicterus (62), a severe consequence of hyperbilirubinaemia associated with significant mortality and morbidity (6, 63). The severity and duration of a haemolytic episode, in most cases, depends on the G6PD variant, nature of the oxidative trigger and the duration of exposure (57). Some cases will first present with neonatal jaundice which peaks at 2–3 days after birth (64). Although highly variable, without effective treatment NNJ is associated with bilirubin encephalopathy and permanent neurological damage or death. Others will present following exposure to significant oxidative stress,

Table 1.2.

Class	Definite association	Possible association
Anti-malarial drugs	Primaquine	Chloroquine
	Dapsone	Quinine
	Methylene blue	
	Tafenoquine	
Antibiotics	Sulfamethoxazole	Sulfasalazine
	Nitrofurantoin	Chloramphenicol

	Co-trimoxazole	Isoniazid
	Quinolones	
	Sulfadiazine	
Other drugs	Rasburicase	Aspirin (high dose)
	Phenazopyridine	Glibenclamide
		Mesalazine
		Vitamin K analogues
		Ascorbic acid
Other chemicals	Toluidine blue	
	Naphthalene	

Table 1.2 Drugs and chemicals associated with haemolysis in G6PD deficient subjects

Adapted from Luzzatto L, et al., Glucose 6-phosphate dehydrogenase deficiency. The metabolic and molecular bases of inherited disease, 8th edn. Columbus: McGraw-Hill, 2001: p.4517–53 (5) and from Roper D, et al., Laboratory diagnosis of G6PD deficiency. British Society for Haematology Guidelines, British Journal of Haematology, 2020. 189: p24-38 (57).

The most common manifestations of disease occur following exposure of the deficient red cells to oxidative stress. These include foods, drugs and infections (such as viral hepatitis, bacterial infections, cytomegalovirus and typhoid fever), **Table 1.2**, although there have been reports of myocardial infarction and even vigorous exercise triggering an attack (30, 49). The resulting intravascular haemolysis may lead to acute renal failure, anaemia and sometimes death (65, 66).

Eradication of malaria in individuals chronically infected by *Plasmodium vivax*, the most prevalent form of malaria in Asia and South America (of which there are an estimated 14 million cases a year with major health and economic impact), represents a global health

priority. Until recently the only effective treatment was primaquine which generates intermediate oxidative metabolites causing haemolysis in G6PD-deficient individuals, generally those with activity <30% of normal (67). Current WHO guidelines advise avoidance of primaquine in severe G6PD deficiency and an attenuated eradication regimen in individuals with mild-moderate G6PD deficiency (68). Despite this, a considerable cohort will suffer significant primaquine-induced haemolysis. Even if G6PD testing were readily available, distinguishing the degree of severity is not possible with most diagnostic tests. The clinical risks of exposing individuals with “moderate” G6PD deficiency to an oxidative substance such as primaquine are not entirely known, but the consequences of drug-induced haemolysis, anaemia, renal impairment and death, are substantial. Newer agents such as tafenoquine, developed to eradicate the dormant, liver stage of *P. vivax*, carries a similar risk of haemolysis in deficient individuals. In the absence of treatment, infected individuals continue to act as a reservoir for transmission of malaria. G6PD deficiency thus poses a major obstacle to successful eradication programmes. The development of a targeted inducer of G6PD activity to mitigate haemolysis and enable safe administration of anti-malarial therapy to G6PD deficient individuals would represent a significant breakthrough and go some way in solving this global dilemma.

Most cases of G6PD deficiency present with single nucleotide polymorphisms (class II and III) following exposure to oxidative stress. A small group of deficient patients experience constant haemolysis (68). This group accounts for less than 0.1% of all G6PD deficient individuals. These class I patients are severely affected (<10% normal activity) causing CNSHA. These occur as sporadic variants at extremely low frequency. More than 95 variants are associated with this phenotype (44, 58, 69), the majority of the causative mutations are

single amino acid changes, although several in-frame deletions have been observed (52).

The absence of frame shift and nonsense mutations is most likely because these genotypes result in embryonic lethality. The majority of these CNSHA cases are due to a limited number of amino acid changes, implying that there are only a few changes that give rise to this phenomenon (60, 70). Major clusters of class I mutations occur in exons 6, 8, 10, and 13 encoding the regions that bind the enzyme substrate, dimer interface, and NADP⁺ structural site, respectively (71, 72).

1.2 PREVIOUS WORK

G6PD deficiency, like most Mendelian disorders, has no specific treatment (18). This is especially pertinent for those with class I deficiency which is associated with significant morbidity and mortality.

The Karadimitris group proposed that transcriptional amplification of genes with hypomorphic mutations that give rise to these disorders would ameliorate disease, particularly in those caused by enzyme deficiencies. It has been noted that enzymatic activity of 20-50% of normal is sufficient for GPPP enzymes to maintain normal or near-normal red cell function and homeostasis (72, 73). Therefore, a therapeutic intervention that would increase enzyme activity by at least 10% would have significant therapeutic benefit.

Early work by Makarona and colleagues in this group provided a new insight into the mechanism underlying transcriptional upregulation of G6PD activity in response to HDAC

inhibition. The group identified HDAC inhibitors as a potential therapeutic intervention for G6PD deficient patients (74).

To address the 'Karadimitris hypothesis', normal B cell lines were treated with sodium butyrate (NaBu), a HDAC inhibitor, and the impact on expression of GPPP genes was assessed by measuring change in mRNA compared to baseline. This showed selective increase in *G6PD* transcriptional activity in response to NaBu. The same effect was observed in a B cell line derived from a patient with G6PD Brighton, a class I G6PD deficiency. This effect was observed across various lineages including T cells and epithelial cells. However, given the unique vulnerability of red blood cells to the impact of G6PD deficiency, it was essential to demonstrate the effect on this lineage. To assess this peripheral blood mononuclear cells (PBMCs) were harvested from normal donors and patients with polymorphic and sporadic G6PD variants. These cells were differentiated to produce proerythroblasts which were treated with NaBu. Once again, the selective increase of *G6PD* expression and increase in enzymatic activity was observed. Importantly, these effects were replicated with different classes of HDAC inhibitors - SAHA and trichostatin (73, 74).

Further work examined the epigenetic regulation of *G6PD* mRNA induction. This revealed an increase in acetylation of histone H3 and H4, **Figure 1.5**, as well as Sp1 and Polymerase II occupancy in the G6PD promoter, features not seen in any of the other GPPP gene promoters. This work also confirmed that HDAC inhibitor-mediated transcriptional upregulation is Sp1-dependent and that a previously described 562bp region, associated with the promoter, is responsible for increased transcriptional activity in response to NaBu (74).

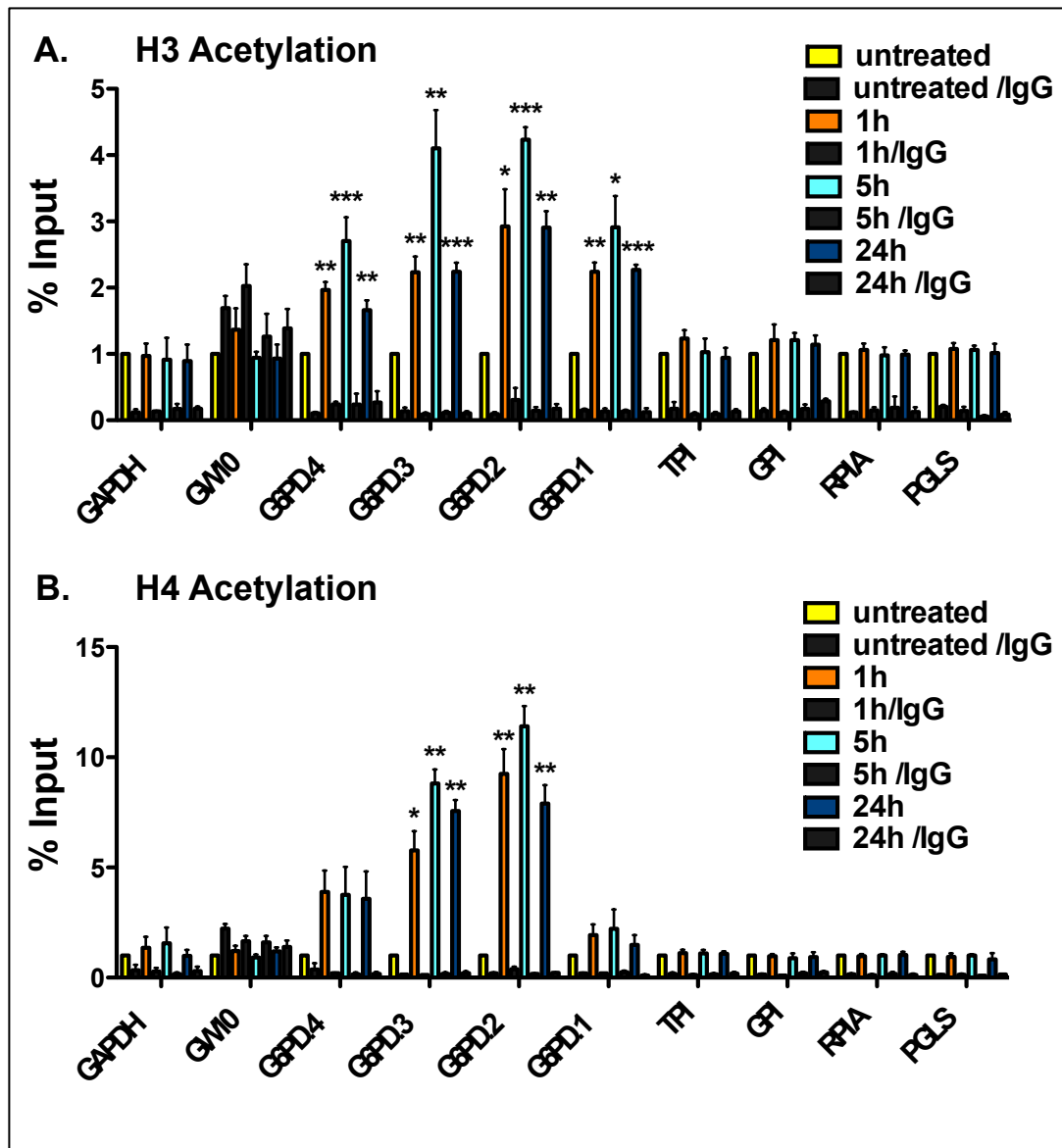


Figure 1.5 Histone acetylation on GPPP gene promoters of native B cells. (A) Histone 3 and (B) histone 4 acetylation assessed by ChIP qPCR after NaBu treatment at 1h, 5h and 24h. Enrichment on the promoters is calculated as % of input relative to the untreated for each primer set. Positive control (GAPDH) and negative control for acetylation. Immunoprecipitation against IgG is employed to set the enrichment background. Mean and S.E.M. are shown for n=3. Student's t-test compare the untreated to the treated conditions (74).

These findings provided evidence to support the hypothesis that enhanced transcription of genes with hypomorphic mutations can restore functional activity beyond a threshold that has therapeutic significance. However, several unanswered questions remain, the most important of which is to understand the mechanism that underpins the selective induction

of *G6PD* expression. This likely holds the key to deciphering the mechanism by which HDAC inhibitors exert their pharmacological effect.

1.3 TRANSCRIPTIONAL REGULATION

To understand why HDAC inhibitors selectively induce *G6PD* transcription we must first appreciate the complexity of translational regulation. HDAC inhibition stimulates a global increase in histone acetylation. Despite this, changes in expression are only seen in 3% of genes, nearly half of which are downregulated (74-76).

A genome-wide study by Wang *et al.* determined only active genes exhibited dynamic acetylation and deacetylation, and that it is only these genes that could be hyperacetylated by HDAC inhibitors. Repressed and primed genes could not be influenced in this manner (77). It was speculated that at the end of each transcription cycle HDACs are required to deacetylate active chromatin to prevent hyperacetylation of the gene. This hypothesis explains the high level of HAT and HDAC binding to the promoters and gene bodies of active genes. The promoters of primed genes demonstrate low levels of HATs and HDACs in association with H3K4 trimethylation, maintaining these genes in a primed state for activation. While repressed genes have no detectable HAT or HDAC binding, maintaining a repressed transcriptional state primarily due to Polycomb complex and the associated H3K27 trimethylation repressive post-translational modification (PTM). While this study explains why only a small proportion of genes are affected by HDAC inhibitors, it does not explain why other housekeeping/active genes within the GPP pathway are not influenced by HDAC inhibition in the same way as *G6PD*.

The answer is also not found when examining the baseline transcriptional regulation of GPPP genes. Makarona's work demonstrated that they are characterised by similar levels of HDAC, HAT and Sp1 promoter binding, similar levels of histone H3 and H4 acetylation, RNA polymerase II binding and transcriptional output (74). Despite these similarities, why is only *G6PD* transcription enhanced following treatment with HDAC inhibitors?

Further clues to the epigenetic mechanism of *G6PD* responsiveness to HDAC inhibitors is the speed of response. Within 4 hours of treatment, there is a 2-3 times increase in transcription, and in 24 hours, a 3-4 times increase in protein synthesis and enzymatic activity (73, 74). This rapid onset is indicative of a direct effect on *G6PD* transcription.

These observations highlight the limitations of current understanding of the epigenetic events that underpin the selective responsiveness of *G6PD* to HDAC inhibition. Yet, this understanding is crucial to the development of targeted therapies in the treatment of the disorder.

1.3.1 Chromatin and transcriptional regulation

The concept of the epigenotype was introduced by Waddington in 1942 (78), who further added to the story by coining the phrase 'epigenetic landscape' in reference to the factors that influence an organism's development (79). But it was another 30 years before the interaction between the histone and chromatin, i.e. the nucleosome as the basic unit of chromatin was conceived (80). The compartmentalisation of DNA within the nucleus of eukaryotic cells requires extreme compaction of this heavily charged polymer. This is

achieved by an association with a highly basic protein unit, histone, which DNA winds, as a spool, to form chromatin (81).

Histones are abundant with lysine and arginine residues. Histone PTMs contribute to the epigenetic regulation of gene expression. Distinct patterns of histone modifications are observed at specific chromosomal regions and affect chromosomal interaction, namely: transcription, replication, repair and recombination. Since their discovery in 1964, histone modifications had long been proposed to have a profound effect on eukaryotic gene expression. This concept was later verified with the identification of enzymes responsible for these modifications, histone acetyltransferases (HATs), which acetylate lysine residues in histone amino-terminal tail regions. HATs are involved in the regulation of both promoter-specific transcription and long-range, chromosome-wide, transcription. Analyses of HATs and other chromatin modifying enzymes have revealed mechanisms of epigenetic regulation that are mediated by a range of post-translational modifications of histones.

The nucleosome is the fundamental unit of the eukaryotic chromosome, and consists of a histone octamer wrapped with 146 bp of DNA. Each histone has a mass comparable to that of the DNA it is associated with. The core histones, H2A, H2B, H3 and H4, all interact with DNA in an analogous way, creating 'histone folds' – a long central α -helix flanked by shorter helices and loops that interact with DNA (82). These core histones have 15-30 unstructured residues at their N-terminus, commonly referred to as the 'histone tail'. This tail can be subjected to post-translational modifications, which are key to modulating chromatin structure and function. These functions include the acetylation of lysine residues, lysine and arginine methylation, serine and threonine phosphorylation, lysine ubiquitination and lysine

SUMOylation (83). This thesis will focus on the transcriptional control of gene expression due to the action of histone acetylation.

1.3.2 Histone modification

As the understanding of the interaction of different PTMs became clear, Strahl and Allis postulated that histone modifications, on one or multiple histone tails, act in a coordinated manner to direct specific and distinct downstream functions, the histone code hypothesis (83). Soon afterward, the terms 'writer', 'eraser' and 'reader' were coined (84). Other components of the epigenetic tool kit include DNA methylation, histone variants, nucleosome positioning, noncoding RNAs and histone chaperones - all of which are necessary for the fine-tuning of chromatin structure and function. There are two mechanisms by which histone PTMs affect chromatin structure and dynamics, either by directly influencing histone-DNA and histone-histone interactions, or by influencing protein effectors also known as histone-binding domains, readers (85). Following recognition of a specific PTM, the reader recruits the nuclear signalling network which mediates fundamental processes such as gene transcription. Chromatin-associating complexes show specificity for different PTMs by virtue of multiple readers within one or several subunits. Numerous histone effectors have been identified and characterised as readers of various PTMs such as forkhead domains which recognise phosphorylated threonines and serines and tudor domains that bind methylated lysine and arginine residues (86). Subsets of readers include effectors (which are domains that bind specific PTMs recruiting other activities contained within the same polypeptide or complex to which it belongs), and presenters (which are able to bind a particular sequence discriminating for a pattern of

PTMs and then present a side chain of this bound epitope for further modification) (87).

Writers also play a key role in epigenetic regulation. These include a wide range of enzymes capable of introducing reversible PTMs to nucleotide bases and specific amino acid residues on histones, such as HATs. Erasers are enzymes that remove the aforementioned PTMs introduced to the nucleotide bases or amino acid residues on histones, examples of which are HDACs (88).

1.3.2.1 Acetylation

Acetylation of the ϵ -amino group of the lysine residue is probably the most critical PTM involved in transcriptional regulation (89). The addition of an acetyl moiety, neutralises the lysine amino acid, altering the charge of the protein. This PTM competes for the same residue with other modifications, such as methylation, SUMOylation and ubiquitination, while phosphorylation can stimulate acetylation supporting transcriptional activity (90). In general, acetylation results in a more open chromatin structure (as the PTM weakens the interaction with negatively charged DNA), facilitating access of transcription factors to the gene promoter thus activating gene transcription (91).

1.3.2.2 Methylation

Methylation is a highly versatile PTM with a key epigenetic role. It is associated with transcriptional activation and repression based on the location of the modified lysine residue. Histone methyltransferases (HMTs) catalyse the transfer of methyl groups using S-Adenosyl Methionine (SAM) as a cofactor, with three possible methylation states. Lysine methylation may occur on six lysine residues of histone H3, K20 of histone H4 and K26 of

histone H1. Transcriptional activation is predominantly associated with methylation of H3 lysine 4, 36 and 79 (92), while repressive epigenetic signatures are linked to methylation of H3 lysine 9 and 27. H3K4 trimethylation accumulates predominantly at the 5'-end of active genes and promotes transcription by interaction with RNA polymerase II. Similarly, H3K36 trimethylation clusters at the open reading frame (ORF) and 3'-end of genes to promote transcription (93), while methylation of H3K9 and of H4K27 are linked to heterochromatin formation and gene repression. Polycomb repressive complex 2 (PRC2) has a SET domain with histone methyltransferase activity and is essential to maintaining chromatin in its repressive state by introducing dimethyl and trimethyl groups to H3K27. This highly conserved protein complex can bind H3K27me3 and neighbouring nucleosomes, thus spreading the repression (94). These modifications can be erased by two families of demethylases that are defined by their enzymatic mechanisms.

The versatility of this PTM is further illustrated by its ability to not only modify lysine residues of histones, but also arginine residues with arginine methyltransferases (PRMTs). This family of enzymes transfer methyl groups from SAM to the terminal nitrogen of the guanidine moiety of arginine amino acids. And, similar to lysine residues, arginine has multiple methylation sites. Methylation of arginine residues is also a dynamic process involved in inflammation and early development (95, 96).

1.3.2.3 Phosphorylation

The addition of a bulky, negatively charged, phosphoryl group to a serine tyrosine or threonine alters both electrostatic and topographic properties of histones. These highly dynamic PTMs can occur on residues of all four core histones and are added by specific kinases and

removed by phosphatases (85). These modifications are critical in the DNA damage response pathways, mitosis and transcriptional regulation (97).

1.3.2.4 Ubiquitination

Ubiquitination involves the formation of an isopeptide bond between the C-terminal glycine of one or more of this small (76 amino acids in length) protein and the ϵ -amino group of a lysine residue (98). Unlike other post-translational modifications, the process of ubiquitination requires the action of three enzymes – E1-activating, E2-conjugating and E3-ligating – acting sequentially, and its removal is achieved by a group of enzymes known as isopeptidases (99). All the core histones may undergo ubiquitination, although it has been estimated that 5-15% of H2A and 1-2% of H2B are ubiquitinated. Although the role of ubiquitination remains controversial, different studies indicate that its function is dependent on genomic location and the context of histone modifications (100, 101). Some suggest H2B ubiquitination may participate in transcriptional activation by facilitating H3K4 methylation and transcriptional elongation (98). While others illustrate its key role in transcriptional repression through H2A ubiquitination at lysine 119 and subsequent binding of the polycomb repressive complex (102).

1.3.2.5 SUMOylation

In addition to ubiquitin, there are several ubiquitin-like proteins capable of modifying histone tails. One such protein is a small ubiquitin-related modifier (SUMO) polypeptide. SUMOylation has a similar enzymatic cascade to ubiquitination - activating, conjugating and ligating – but initially requiring the action of a sentrin-specific peptidase to remove the C-

terminal of lysine residue. SUMOylation has been detected on all four core histones and seems to function by antagonising acetylation and ubiquitination that might otherwise occur on the same lysine residue, suggesting this PTM has a mainly repressive effect on transcription (103, 104). This was further illustrated by work performed by Shiio and Eisenman who first described histone SUMOylation in 2003, suggesting its repressive effect on transcription occurs due to recruitment of HDAC1 and heterochromatin protein 1 (HP1) (103).

1.3.2.6 Nucleosome positioning

The nucleosome limits access of the transcriptional machinery to DNA through specific positioning of the nucleosome core particle. This dynamic process can be influenced by DNA sequence, histone variants and modifications, but also by ATP-dependant nucleosomal remodelling to alter accessibility to gene promoters and regulatory regions (105). There are four families of chromatin remodelling complexes: SWI/SNF, ISW, CHD and INO80 (106). Histone composition and nucleosome positioning effect DNA methylation, decreasing DNA flexibility which shortens the linker regions, facilitating internucleosomal interactions (107). This process increases the influence one nucleosome has on another, thereby affecting the accessibility of the associated chromatin.

1.3.2.7 Interaction of histone PTMs

Cross-talk between different modifications can occur via diverse, and sometimes complex mechanisms including:

1. Communication at the level of a single histone tail, the *cis* effect. An example of this is the methylation of H3K9 that inhibits acetylation of the H3 tail and methylation of H3K4 (108). This illustrates that specific lysine residues act as acceptor sites for an array of different PTMs that are in direct competition for the amino acid residue.
2. The *trans* effect, or interactions at the nucleosome level where modifications to different histones interact, such as H3K9 trimethylation which is required for the trimethylation of H4K20 to occur (109).
3. The effect of DNA methylation and histone modification pathways. The relationship between DNA methylation and H3 methylation has been well established and most likely is mediated by SET domain HMTs and DNA methyltransferases (DNMTs) (110).

1.3.2.8 Non-coding RNA

Genome-wide studies have revealed that mammalian genomes are pervasively transcribed. High-throughput transcriptomic analyses have shown that eukaryotic genomes transcribe up to 90% of the genomic DNA (111), but only 1-2% of these transcripts encode for proteins (112). This class of non-coding RNAs (ncRNAs) influence a variety of cellular functions and can be divided into two groups, regulatory and structural ncRNA. Structural or housekeeping ncRNAs are constitutively expressed and are required for normal cellular functions. This group includes transfer RNA, ribosomal RNA, small nuclear, small nucleolar RNA, RNase P RNAs and telomerase RNA (113). In contrast, regulatory ncRNAs can affect the expression of other genes at the level of transcription or translation and play a role in chromatin regulation via interaction with chromatin-modifying enzymes and transcription factors. They regulate gene expression at the post-transcriptional level, directly or indirectly, by either

cleaving and repressing translation of target mRNA, or directly inhibiting enzymes such as DNMTs, histone modifications and chromatin remodelling (114, 115).

1.3.3 Acetylation

Histone lysine acetylation is the post-translational modification of histones enabling transcription and is the main focus of this project. The reversible addition of an acetyl moiety to the ϵ -amino group of the lysine residue was first described in two publications in the early 1960s (116, 117). Allfrey and colleagues proposed that histone acetylation had a role in regulating gene expression (118). This was supported by the observation that hyperacetylation of histones correlates with transcriptional activation (119). This modification results in charge neutralisation of conserved, often invariant, lysine residues located in amino-terminal domains of the histones. This stabilises the electrostatic contacts between the histones and the negatively charged DNA, loosening of the histone-DNA interactions, allowing DNA proximal to this region to adopt a transcription competent confirmation (120).

Interest in this field surged with the identification of the first HAT and HDAC associated with acetylation, and their contribution to gene regulation (121-124). Since then, their diverse role and dynamic equilibrium has been unravelled, including lysine-like acetylation (such as propionylation and butyrylation). Histone acetylation is now a robust therapeutic target with agents directed to various elements of the mechanism. It has been shown to be an essential regulatory mechanism that controls transcription of approximately 2-10% of all genes (125). Deacetylation of histone tails cause chromatin condensation, while hyperacetylation has the reverse effect (126). These changes may result in an increase or

decrease of transcription depending on the gene. It is worth noting that HDACs also deacetylate non-histone proteins altering their activity. These include, chromatin remodelling proteins, DNA-binding nuclear receptors, DNA repair enzymes, signalling mediators, structural proteins, transcription coregulators and DNA-binding transcription factors (127). Acetylation can also affect signalling pathways, mRNA splicing, transport, integrity, translation, protein activity, protein localisation, protein stability and interactions. Therefore, acetylation can modify every element of the regulatory processes from signalling to transcription to protein degradation (128). This provides a plausible explanation of why gene expression determined by HDAC inhibition, is not always increased even if the chromatin structure is open.

1.3.4 HATs and HDACs

1.3.4.1 HATs

HATs are a diverse group of enzymes that acetylate conserved lysine of histones in a highly dynamic process using acetyl CoA modifying the nucleosomal conformation enabling transcription (129). Structurally, HATs are characterised by a well-conserved core region that consists of a three-stranded β -pleated sheet with an associated α -helix parallel to it (130). They are divided into two different classes based on their subcellular location (81). Type A HATs are found in the nucleus and are involved in gene regulation by acetylation of nucleosomal histones. These HATs contain a bromodomain, a conserved protein motif that modulates binding of acetylated lysine residues, a so called 'reader' of lysine acetylation, and they cooperate with activators to enhance transcription (131, 132). Most HATs, such as General control non-derepressible 5 (Gcn5) and p300/CBP, are type A. Type B HATs are located within the cytoplasm. Here, they catalyse acetylation events linked to the transport of newly synthesised histones from the cytoplasm to the nucleus prior to integration with newly replicated DNA. Type B HATs are structurally distinct from type A HATs as they lack a bromodomain.

Type A HAT catalytic activity is regulated by either interaction with regulatory subunits or by autoacetylation. They are subclassified into three broad groups, or families, based on sequence homology, shared catalytic domains and functions (133). These families include Gcn5-related N-acetyltransferases (GNATs), p300/CBP (CREB-binding protein) and MYST proteins.

1.3.4.2 GNATs

Gcn5 is the founding member of this family of HATs, which also includes PCAF, Elp3, Hat1, Hpa2 and Nut1 (133). These HATs are characterised by the presence of a bromodomain, up to four conserved motifs, and classically acetylate lysine residues on H2B, H3 and H4 histones (133). Gcn5 was the first nuclear HAT, identified in protozoa and named p55 until the homology with a transcriptional coactivator Gcn5 was appreciated (121). Soon after Gcn5 was found to possess HAT activity, histone acetylation was directly linked to activation of gene expression (134). Gcn5 also works in conjunction with multiple subunits to enable nucleosome acetylation. This complex named SAGA (Spt-Ada-Gcn5-Acetyltransferase) preferentially acetylates the histone tails of histones H3 and H2B (135). SAGA has served as a paradigm for the coupling of histone modifications with transcriptional co-activation in eukaryotes.

1.3.4.2.1 MYST family

This family is highly conserved in eukaryotes and perform a significant proportion of nuclear acetylation (136). The MYST family derives its name from its four founding members: MOZ (monocytic leukaemia zinc finger protein), Ybf2 (Sas3), Sas2 and Tip60 (133). These HATs are typically characterised by the presence of a small structural motif - MYST domain or zinc fingers, surrounding the acetyl CoA binding domain that contains at least one zinc ion which maintains the secondary structure (137). Other domains - chromodomains (chromatin organisation modifier) have a role in chromatin manipulation and remodelling. The MYST family is commonly associated with acetylation of lysine residues on histones H2A, H3 and

H4. The circadian rhythm protein CLOCK has also been described as having a MYST-like HAT domain, which is a principal component of the circadian pacemaker (138).

1.3.4.2.2 Others

This minor group of HATs possesses intrinsic HAT activity, but do not contain true consensus domains (120). These include p300/CBP (CREB-binding protein), Taf1 and several nuclear receptor co-activators (133). The two most widely investigated acetyl transferases in this group are p300 and CBP. These two homologous proteins often act interchangeably and have several discrete domains – three zinc finger domains, a bromodomain, a HAT domain; and at least two independent regions that interact with multiple transcription factors, a feature shared with Gcn5 (139).

1.3.4.2.3 HAT Complexes

HATs are a group of diverse enzymes and the complexes from which they derive are equally varied. Different HAT complexes, even within the same family, are composed of various unique subunits that enable them to carry out a distinct function on a common substrate. For example, some subunits have domains, or ‘readers’, which recognise modified histone tails that cooperate to recruit the HAT to the appropriate location in the genome; these include bromodomains, chromodomains, WD40 repeats, Tudor domains and PHD fingers. Table 3 illustrates how these, and other chromatin-binding domains confer HAT specificity.

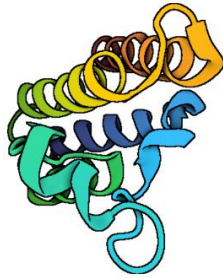
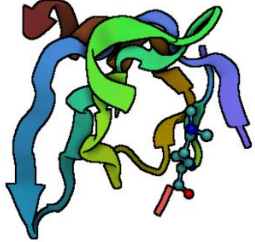
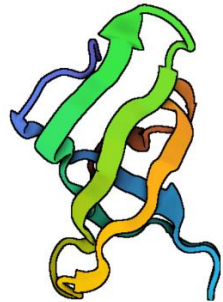
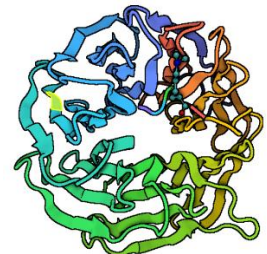
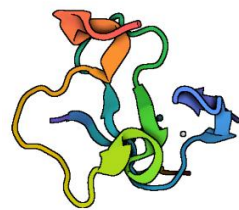
Domain	Structure	Role	Complexes	Proteins
Bromo		Binds acetylated lysine	SAGA	Gcn5, Spt7
			NuA3	Yta7
Chromo		Binds methylated lysine	SAGA	Chd1
			NuA4	Esa1, Eaf3
Tudor		Binds methylated lysine and arginine	SAGA	Sgf29
WD40		Binds methylated lysine	SAGA	Spt8
			HatB	Hat2
			Elongator	Elp1, Elp2
PHD		Binds methylated lysine	NuA4	Yng2
			NuA3	Yng1, Nto1

Table 1.3 Chromatin-binding domains in known HAT complexes

(Adapted from Lee K & Workman J, Histone acetyltransferase complexes: one size doesn't fit all, Nature Reviews Molecular Cell Biology, 2007. 8 (4): p. 284-295 (133)).

Table 1.3 illustrates how HATs derive similarities and specificity from their component parts.

This is demonstrated by comparing three HAT complexes that have both overlapping substrates and the specialised functions. SAGA is a 15-subunit complex that preferentially acetylates lysine residues of the N-terminal tails of H3K9 and, to a lesser extent, H3K14 of the same histones (140). In contrast, NuA3, a Gcn5-independent HAT complex which shares several subunits, preferentially modifies H3K14 (141). Meanwhile, the Elongator complex has overlapping substrate specificity with the SAGA complex but it is thought to function in gene-coding regions, rather than at promoters (as is the case with SAGA) to acetylate nucleosomes during transcription (142).

1.3.4.3 HDACs

HDACs are highly conserved and classified based on their homology with yeast orthologs. Eighteen isoforms have been identified and classified based on sequence similarity and cofactor dependency. HDACs are grouped into six classes and two families (classical and silent information regulatory 2 (Sir2)-related protein (sirtuin) families). The members of the classical family include class I (HDAC1, HDAC2, HDAC3 and HDAC8), class II (HDAC4, HDAC5, HDAC6, HDAC7, HDAC9 and HDAC10) (143) and class IV (HDAC11) (144). They share sequence similarity and require a zinc moiety that facilitates the hydrolysis of the acetyl-lysine bond necessary for deacetylation.

Class I relates to yeast RPD3 gene and phylogenetic analyses suggest this class can be further subdivided: class Ia (HDAC1, HDAC2), class Ib (HDAC3) and class Ic (HDAC8) (145). They are primarily localised in the nucleus and ubiquitously expressed. Class II are related to yeast histone deacetylase 1 and again subdivided into class IIa (HDAC4, HDAC5, HDAC6 and HDAC9) and Class IIb (HDAC6 and HDAC10), which contains two catalytic sites (146, 147). Phylogenetic analysis of these HDACs suggest their evolution preceded that of histones (145), as these HDACs do not have a primary histone target (148). Class IIa HDACs are primarily cytoplasmic with conserved binding sites for transcription factors and chaperone proteins which shuttle these enzymes between the nucleus and the cytoplasm (148). Class IIb HDACs are cytoplasmic proteins.

HDAC11 is nucleocytoplasmic and unique. It is the only member of Class IV as it has conserved residues in the catalytic core region shared by both class I and II enzymes and a ubiquitin binding site (149). It is expressed by a restricted number of tissues and is

principally cytoplasmic, although it is capable of migration between the cytoplasm and nucleus.

Class III (sirtuins) are distinct from the other classes of HDACs. Instead of using a zinc ion to hydrolyse the amide bond, sirtuin-mediated deacetylation is coupled to NAD⁺ hydrolysis (150). The seven human SIRT paralogs are ubiquitously expressed (151). SIRT1, 6 and 7 are nuclear proteins with differential subnuclear distribution. SIRT6 is linked with heterochromatic regions and SIRT7 is found in nucleoli. Although SIRT1 is mainly localised to the nucleoplasm, it also carries out some cytoplasmic functions (152). SIRT2 is found in the cytoplasm and SIRT3, 4 and 5 are mitochondrial proteins (153). This project focuses on the classical HDACs as work by our group indicated that inhibitors targeting classical HDACs selectively induce *G6PD* transcription and activity (74).

HDACs do not bind directly to DNA, they are recruited to target genes by associated transcriptional activators and repressors and incorporated into large multiprotein transcription complexes (148). Their regulation occurs at multiple levels including protein–protein interactions, post-translational modifications and by metabolic cofactors (154). The classical HDACs form the catalytic core of large complexes involved in chromatin modification and gene repression. So far, four such molecular machines have been characterised (155). These include, the CoREST, NuRD and Sin3 complexes which contain a HDAC1-HDAC2 dimer at their core and the NCoR complex which is formed around HDAC3 (156). The roles of these complexes are diverse and often tissue specific, although we have yet to develop a full understanding of their function in cellular homeostasis. The CoREST complex couples histone deacetylation to demethylation to repress neuronal genes (157), the NuRD complex links deacetylation to a chromatin-remodelling ATPase and promotes

gene silencing (158). The Sin3 complex regulates genes involved in a variety of developmental pathways (159), while the NCoR complex is the major corepressor for nuclear receptors (160, 161).

1.3.4.4 HAT/HDAC interaction

To fully understand the regulation of HAT and HDACs it is important to consider them in a cellular context, and particularly from the point of view of how they interact with each other and other regulators. This interplay occurs via several mechanisms.

First, the ϵ -amino group of a lysine residue is extremely versatile and subject to numerous modifications including acetylation, methylation, ubiquitination, SUMOylation, propionylation and butyrylation. However, once modified this precludes other histone modifications. Therefore, HATs and HDACs control the availability of certain lysine residues for modification. There is some evidence that hypoacetylation is a prerequisite to heterochromatin formation and gene silencing in mammals (162). Second, acetylation of a lysine residue may affect modification of a neighbouring amino acid residue. For example, acetylation of H3K9 prevents phosphorylation of the adjacent serine 10 (163). Third, some HATs and HDACs have intrinsic catalytic activity related to other PTMs. Histone acetyltransferases - PCAF and p300 possess intrinsic ubiquitin ligase activities (164-166). Fourth, some HATs and HDACs possess features that enable sequential actions with different modifications. HDAC6 possesses a zinc-finger domain for ubiquitin binding, allowing the recognition and transport of ubiquitinated proteins and controlling polyubiquitin-chain turnover (167). Fifth, HATs and HDACs interact with other epigenetic players such as ncRNA. Like transcription factors, ncRNA can recruit HATs and HDACs in a sequence-specific manner. RNA interference machinery can induce a repressive chromatin state characterized by cytosine methylation which is catalysed by HDAC6 (168). Finally,

different HATs and HDACs can target a single substrate. PCAF, p300 and two members of the MYST family acetylate p53 (169, 170).

These six factors illustrate the complex interaction of HAT and HDACs, and how their actions are inextricably linked. So, to understand the action of HDAC inhibitors, one must consider not only the impact on the target complex, HDAC, but also its interplay with other key factors.

1.3.4.5 G6PD and regulatory control

The bidirectional *G6PD* promoter is 868 bp in length with a core promoter of 192bp (171). It has a high GC content, greater than 70%, and is embedded within a CpG island. This island extends to 1245 bp in length which includes the *NEMO* gene (33). Lacking a TATA box, RNA polymerase binding was thought to occur at an ATTAAAT sequence located at –202bp from the ATG start codon. However, subsequent work has refuted this hypothesis (172-174).

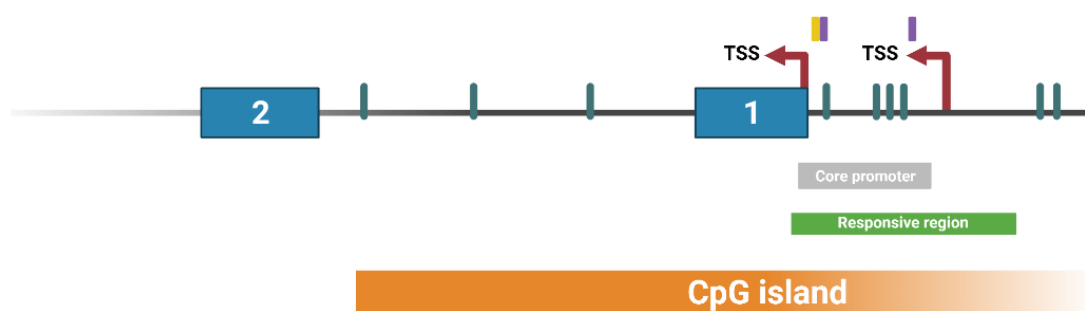


Figure 1.6 Features of G6PD gene and regulatory control

Schematic representation of G6PD gene. First two exons numbered 1 and 2, two transcription start sites (TSS), CpG island extends beyond the G6PD gene. Bidirectional promoter shared with NEMO gene. Responsive region. TATA box - green, GC boxes - purple, and Sp1 sites – cyan.

UCSC reference (Transcript): hg38 chrX: 154,531,390-154,547,572.

Figure created in BioRender by author.

(Adapted from Makarona K, et al., Transcriptional and epigenetic basis for restoration of G6PD enzymatic activity in human G6PD-deficient cells. *Blood*, 2014. 124: p.134-141 (74); Franzè A, et al., Molecular anatomy of the human glucose 6-phosphate dehydrogenase core promoter. *FEBS Letters*, 1998. 437: p313-318 (175); Philippe M, et al., Promoter function of the human glucose-6-phosphate dehydrogenase gene depends on two GC boxes that are cell specifically controlled. *European Journal of Biochemistry*, 1994. 226: p.377-384 (171) and Galgóczy P, et al., Human-mouse comparative sequence analysis of the NEMO gene reveals an alternative promoter within the neighbouring G6PD gene. *Gene*, 2001. 271: p93-98 (174)).

G6PD expression is Sp1 dependant (174). Sp1 is a member of the Sp/KLF family of transcription factors which contains a zinc finger motif that binds GC-rich elements of many promoters.

Work by our group demonstrated that baseline transcriptional regulation such as Sp1 binding is similar for all enzymes of the GPP pathways (74). Bioinformatic analysis predicts twelve Sp1 binding sites surrounding exon 1 of the *G6PD* gene (33). Sp1 plays an important regulatory role in basal G6PD expression (171) and in response to HDAC inhibition (74). The core promoter has five Sp1 binding motifs that are crucial to the HDAC inhibitor interaction. And certain motifs seem to be especially important in this process as it is highly enriched in non-GPPP genes that were upregulated in erythroid precursor cells treated with NaBu (74).

1.4 HDAC INHIBITORS

Interestingly, the discovery of histone deacetylase inhibitors (HDACi) preceded the discovery of HDACs. Sodium butyrate was the first compound found to induce histone acetylation (176). Currently, numerous agents are used for a diverse range of clinical indications – anti-epileptics, mood stabilisers, anti-fungal agents, autoimmune disease and cancer therapy.

HDACi are a large group of diverse compounds that prevent the deacetylation of ϵ -amino group of the lysine of the histone tail, a key event in chromatin remodelling (177). HDAC inhibitors can be classified in many ways, based on the scope of inhibition (i.e. they inhibit only subgroups of HDACs (isoform-selective inhibitors) or act against all HDACs (pan-inhibitors)). Other classifications are based on chemical structure: hydroxamic acids (hydroxamates); short chain fatty (aliphatic) acids; benzamides; cyclic tetrapeptides; and sirtuin inhibitors. And finally, classical HDAC inhibitors inhibit classical HDACs while sirtuin inhibitors inhibit sirtuins, **Table 1.5**.

Class	HDAC inhibitor	Target HDAC group
Hydroxamic acids	Suberoylanilide hydroxamic Acid (SAHA)	Pan-inhibitor
	Trichostatin A	Pan-inhibitor
	Panabioestat	Pan-inhibitor
	Belinostat	Pan-inhibitor
	Resminostat	Pan-inhibitor
	Abexinostat	Pan-inhibitor
	Quisinostat	Pan-inhibitor
	Givinostat	Pan-inhibitor
	Rocilinostat	II
	Practinostat	I, II and IV
	Nanatinostat	I
Short chain fatty acids	Valproic acid	I, IIa
	Butyric acid	I, II
	Phenylbutyric acid	I, II
Benzamides	Entinostat	I
	Tacedinaline	I
	Domatinostat	I
	Mocetinostat	I, IV
Cyclic tetrapeptides	Romidepsin	I
Sirtuin inhibitors	Nicotinamide	All class III
	Sirtinol	SIRT 1 and 2
	Cambinol	SIRT 1 and 2
	EX-527	SIRT 1 and 2

Table 1.4 Classification and activity of HDAC inhibitors

(Adapted from Eckschlager T, et al. Histone Deacetylase Inhibitors as Anticancer Drugs, International journal of molecular sciences, 2017. 18 (7): p.1414 (127).

This project focuses on deciphering the mechanism by which classical HDAC inhibitors, specifically SAHA, exert their pharmacological effect. Currently, we understand little of the mechanism by which SAHA affects gene expression. We know that HDAC inhibitors induce global histone acetylation. *In vitro*, only 2-10% of genes show any meaningful change in expression with almost equal numbers upregulated and downregulated (75, 178). However, the concentration, duration of incubation and choice of HDACi are critically important to the number of genes affected and magnitude of change (179, 180). This variability may be, in part, due to the suppressive effect HDAC has on the dynamic equilibrium between HAT and HDAC. The exact mechanism by which HDACi variably exert their effect on gene expression is unclear, but we also know that the scope of HDAC inhibitors extends beyond histone acetylation, and this would certainly contribute to the impact of HDAC inhibition. Promoters are both induced and repressed following HDACi treatment (181, 182).

These data suggest that histone hyperacetylation alone is not a reliable predictor of gene activity. Additionally, the activity of HDAC inhibitors extends beyond histone modification to other proteins which likely contributes to the pharmacological effect but also gives rise to unwanted side effects (128).

HDACi, particularly hydroxamic acids, have common structural features. This comprises of a hydrophobic capping group which binds the rim of the substrate channel blocking it, a spacer comprised of an alkyl, vinyl or aryl group that spans the channel, and a zinc chelating moiety at the other end. This is illustrated by SAHA, **Figure 1.7 A and B**. Interestingly, as the amino acid sequence at the rim of the HDAC has greater diversity than other regions of the protein (183), this region can be manipulated to develop HDACi specificity. This project aims

to decipher the mechanism by which these drugs work to develop more targeted agents, possibly using this approach.

1.4.1 Sodium butyrate

Butyrate is produced by microbial fermentation of dietary fibre in the large intestine. It plays a vital role in the maintenance of colonic tissue homeostasis by regulating the expression of genes associated with proliferation, differentiation and apoptosis (178).

Sodium butyrate is the salt of the short chain fatty acid butyric acid. It was the first chemical identified as having HDAC inhibitory properties. Short chain fatty acids (**Table 1.4**), including butyric acid, are the weakest inhibitors of HDAC due to their low zinc binding affinity (184).

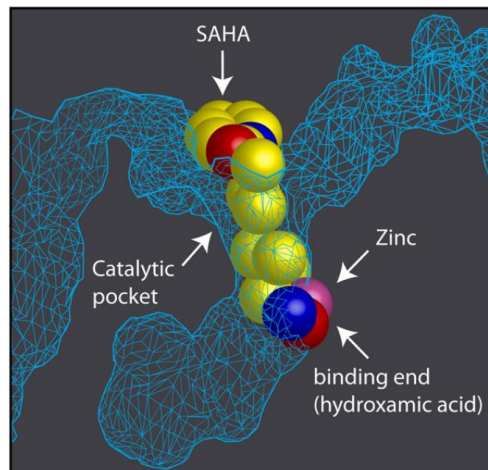
Microarray data of over 19,000 genes in response to sodium butyrate, identified 221 potentially butyrate-responsive genes, of which 59 are upregulated and 162 downregulated (178). Within the promoter region of these genes is a butyrate responsive element which may be structured in one of two ways (185). The promoter has a common AGCCACCTCCA sequence, which is indicative of a common transcription factor binding site (185). These butyrate-responsive genes can be induced or repressed by butyrate. Alternatively, a responsive gene shares a Sp1/Sp3 binding site within the butyrate responsive element. These transcription factors are ubiquitously expressed, acting as both activators and repressors of gene expression and interacts with both HATs and HDACs. Our group mapped the HDACi induced transcriptional upregulation of *G6PD* to a 562bp promoter region (74).

1.4.2 Suberoylanilide hydroxamic acid (SAHA)

Small-molecule HDAC inhibitors have been developed and are licensed for treatment of malignant and non-malignant disorders (186). On a cellular level, the active site of Zn^{2+} -dependent HDAC consists of a tubular region, located within the base of the catalytic pocket (187). The hydroxamic acid group of SAHA reversibly binds the Zn^{2+} ion preventing hydrolysis of the acetyl-lysine bond required for deacetylation. An alkyl chain linker spanning the length of the tubular pocket, and a phenylamine cap blocks the active site by interacting with the external surface inhibiting the deacetylase activity, **Figure 1.6** (187, 188). Despite the importance of protein–protein interactions for HDAC activity, very little is known about how this regulatory effect is exerted at a molecular level (189).

The structural details of the HDAC-HDACi interaction have been elucidated in studies using histone deacetylase-like protein with SAHA (187, 190). These studies provide insight into the mechanism of deacetylation, but unfortunately our understanding of this drug's impact on transcription is limited. What we do know is that this is an extremely complex, multi-level interaction, not only involving histones but also non-histone proteins. The Click ChIP project aims to better understand all of these interactions, while CAPTURE will discover specifically how SAHA exerts its effect on G6PD activity.

A



B

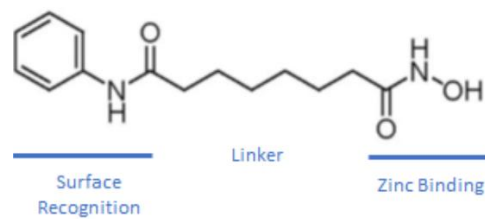


Figure 1.7 Structure of SAHA

(A) Schematic depiction of the crystal structure of the histone deacetylase like protein with suberoylanilide hydroxamic acid (SAHA) inserted into the pocket-like catalytic site of the enzyme.

At the base of the catalytic pocket is a zinc molecule with which the hydroxamic moiety of SAHA binds.

(B) Structural features of SAHA.

(Adapted from Marks P A, Histone deacetylase inhibitors: A chemical genetics approach to understanding cellular functions, Biochem Biophys Acta, 2010. 1799: p. 717-25(183).

SAHA also targets non-histone proteins (**figure 1.8 and table 1.5**) altering their acetylation status. These include transcription actors, chaperone proteins and cell signalling proteins. This change in acetylation status alters protein stability, protein-protein and protein-DNA interactions (191).

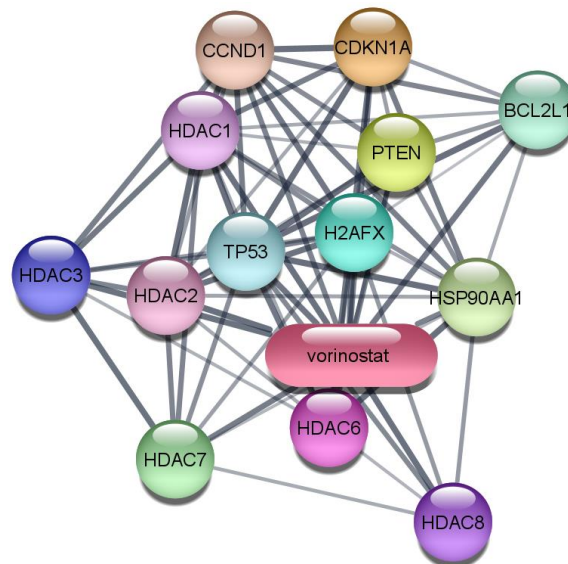


Figure 1.8 Histone and non-histone targets of SAHA (Vorinostat®) from STRING text mining using Cytoscape. String network shows the interactions documented in literature with HDACs, cell cycle proteins, tumour suppressor proteins and heat shock proteins

Function	Proteins
DNA binding transcriptional factors	p53, c-Myc, AML1, BCL-6, E2F1, E2F2, E2F3, GATA-1, GATA-2, GATA-3, GATA-4, Ying Yang 1 (YY1), NF- κ B (RaIA/p65), MEF2, CREB, HIF-1a, BETA2, POP-1, IRF-2, IRF-7, SRY, EKLF
Steroid receptors	Androgen receptor, oestrogen receptor α , glucocorticoid receptor
Transcription coregulators	Rb, DEK, MSL-3, HMGI(Y)/HMGA1, CtBP2, PGC-1 α
Signalling mediators	STAT3, Smad7, b-catenin, IRS-1
DNA repair enzymes	Ku70, WRN, TDG, NEIL2, FEN1
Nuclear import	Rch1, importin- α 7
Chaperone protein	HSP90
Structural protein	α -Tubulin
Inflammation mediator	HMGB1
Viral proteins	E1A, L-HDAg, S-HDAg, T antigen, HIV Tat

Table 1.5 Non-histone targets for HDAC inhibitors (Xu WS, et al; Histone deacetylase inhibitors: molecular mechanisms of action, Oncogene, 2007. 26: p. 5541-5552)

Surprisingly, inhibition with SAHA alters only about 10% of the potential 3,600 acetylated lysines in 1,750 proteins identified (192). These sites were found on proteins responsible for gene expression, RNA signalling, DNA damage repair, cell cycle progression, nuclear transport, cytoskeletal function, protein chaperone, and ribosome formation and function (148). Clearly, the mechanism by which SAHA selectively induces G6PD activity is complex. My project aims to decipher this mechanism with two novel approaches.

1.5 CHIP TECHNIQUES

Pioneering work in this lab has provided proof of the concept that increasing transcriptional output from a defective gene with HDAC inhibitors can ameliorate clinical phenotype.

However, the molecular mechanisms that determine this phenomenon and explain the differential transcriptional response of *G6PD* in contrast to other GPPP genes to HDAC inhibitors remains elusive. To decipher these mechanisms, I adopted a bipartite approach using SAHA, a reversible pan-HDAC inhibitor, and two novel ChIP techniques to monitor the *de novo* functional state of HDAC complexes on chromatin *in vivo*.

Traditional approaches such as the incorporation of affinity tags and use of targeted antibodies, have gone some way in enhancing our understanding of epigenetic pathways and how drugs interact with them. However, to gain a deeper understanding of how HDAC inhibitors work and more specifically how they impact the GPP pathways would require more novel approaches. Two strategies were adopted to manipulate established bioorthogonal reactions, these reactions selectively modified chemical complexes *in vivo*. These complexes enabled isolation of target chromatin in an unbiased manner.

Click ChIP is a technique that facilitates affinity enrichment of HDAC associated chromatin using 'tagged' SAHA followed by targeted mass spectrometry and DNA sequencing. This allows genome-wide mapping of SAHA binding which is compared with relevant epigenomic data obtained from the ENCODE project (193). This replaced Chem ChIP, a similar technique which used a biotinylated SAHA as an affinity tag.

Streptavidin is a 52 kDa tetramer that forms a high affinity, non-covalent interaction with biotin. This is one of the most avid interactions in nature, with a dissociation constant on the order of 10^{-14} mol/L (194). The biotin-streptavidin complex is employed to enable pull down of molecular associations with SAHA for downstream analysis. The protocols developed allowed purification of the tethered molecules for analyses of DNA and proteomic interactions.

CAPTURE adapts the CRISPR-Cas9 systems to purify specific genomic regions while retaining molecular interactions. The technique allows locus-specific targeting of *G6PD* and other GPPP gene regulatory areas to enable comparative analysis of histone PTMs and chromatin binding proteins in an agnostic manner (195).

1.6 CLICK CHIP

Click Chemistry describes a chemical reaction between two functional groups that occurs under physiological conditions (196). The reaction has the advantage of being rapid, highly selective and non-toxic. There are three main types of reactions which fulfil reactivity, selectivity and biocompatibility criteria:

1. **Cu(I)- catalysed Azide - Alkyne Click Chemistry reaction (CuAAC)**
2. **Strain-promoted Azide - Alkyne Click Chemistry reaction (SPAAC)**
3. **Tetrazine – Alkene Ligation**

The CuAAC reaction is the most commonly used methodology. A terminal alkyne reacts with an azide group to form a stable conjugate via a triazole moiety, **Figure 1.7A**. This azide-

alkyne bioconjugate is characterised by high affinity and specificity. The two interacting species are inert, so they are dependent on Cu (I) ions to catalyse the reaction (197).

Tyler *et al.* described the use of Click-ChIP, a chromatin immunoprecipitation technique that facilitates affinity enrichment of targeted chromatin. They used a modified BET inhibitor which included a terminal alkyne that was amenable to the CuAAC reaction but remained functionally conserved (198). This modified drug was used as a molecular probe to examine the gene regulatory function of the drug's main epigenetic target and explore the transcriptional changes induced by BET inhibitors.

I adapted the approach using a 'clickable' SAHA, SAHA-BPyne, which includes a benzophenone linker and a terminal alkyne tag, **Figure 1.7B**, yet retains HDAC inhibitory properties (199).

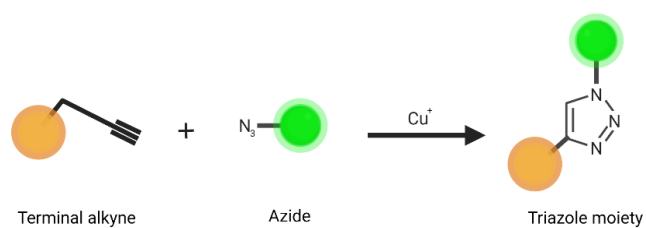


Figure 1.9A Copper-catalysed Azide-alkyne Click reaction

Figure created in BioRender by author.

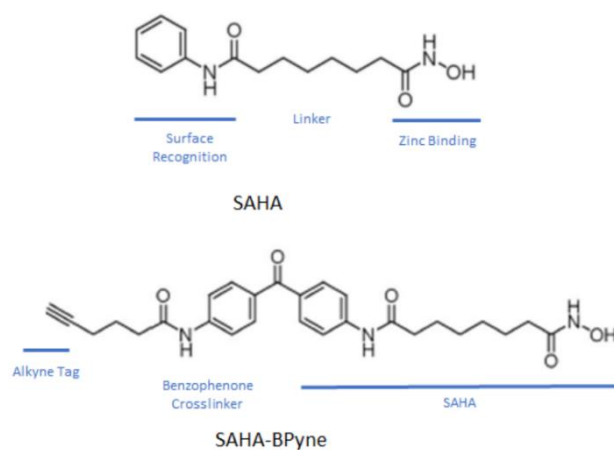


Figure 1.9B Clickable SAHA

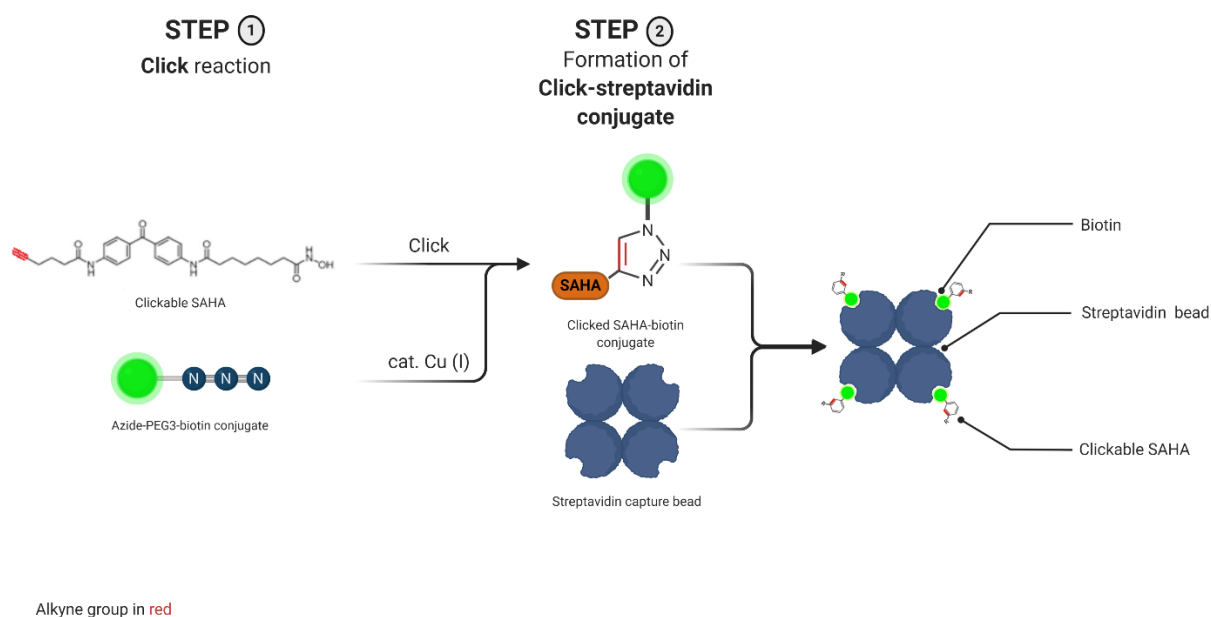


Figure 1.9C Formation of SAHA-Streptavidin conjugate

Figure created in BioRender by author.

The strategy adopted is outlined in **Figure 1.9C**. The introduction of a terminal alkyne group to SAHA allows for a linking reaction with an azide-biotin conjugate moiety which, in turn, binds streptavidin beads and facilitates affinity purification. These bead complexes were then precipitated with tethered target chromatin in a process similar to ChIP-seq/proteomics. Data was analysed as per ChIP-seq workflow established protocols in the Karadimitris Lab. These data were compared with ENCODE datasets which include histone PTMs, CBPs and transcription factor binding maps in K562 cell line.

1.6.1 EnChIP/CAPTURE

In contrast to Click ChIP, the engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) technique aims to isolate a single genomic locus and associated proteins to analyse these interactions by DNA sequencing and mass

spectrometry. Fujita *et al.* manipulated the CRISPR-Cas9 system utilising a nuclease deficient Cas9 (dCas) with a 3xFLAG tag (200). The inactive Cas9 and sequence-specific gRNA enabled tethering of target chromatin while 3x FLAG tag allowed immunopurification.

Further development of the technique by Liu *et al.* established a dCas9 with a biotin acceptor which facilitates *in vivo* biotinylation by a biotin ligase allowing for more efficient chromatin purification, CAPTURE (195). *G6PD* and *TPI* gRNAs recruit the biotinylated dCas9 to the targeted sequence, the locus associated macromolecules are isolated by high affinity streptavidin precipitation for downstream analysis, **Figure 1.8**.

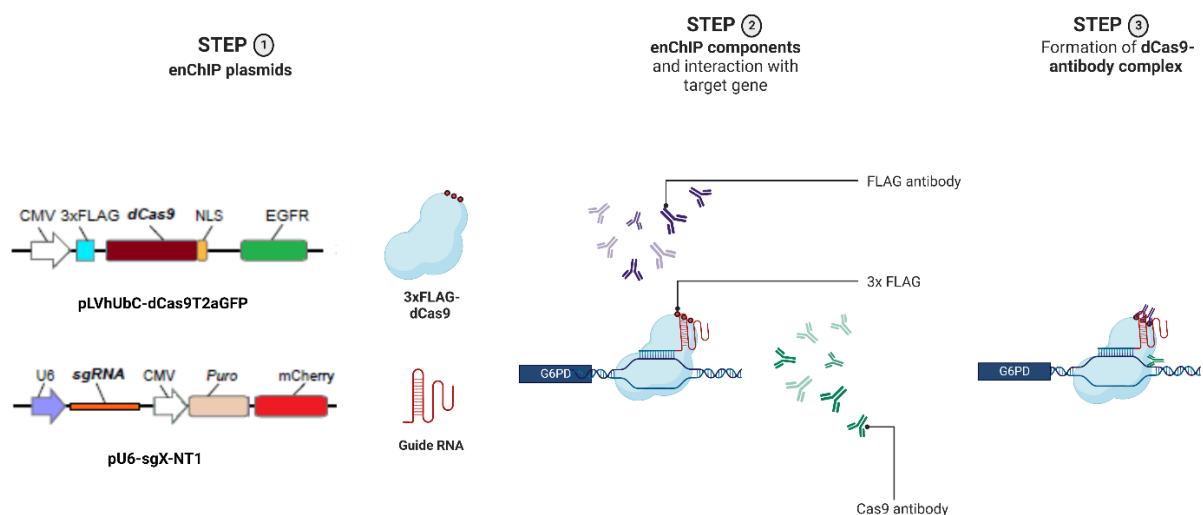


Figure 1.10 Schematic of enChIP system

The enChIP protocol described by Fujita *et al.* uses the CRISPR system, dCas9 and gRNA, to purify locus specific chromatin. The system consists of a fusion protein – deactivated Cas9, 3x FLAG tag, and the nuclear localization signal - encoded by pLVhUbC-dCas9T2aGFP plasmid. The second plasmid, pU6-sgX-NT1 plasmid, encodes for a locus specific guide RNA. EnChIP cell lines were treated with SAHA/vehicle, crosslinked, chromatin extracted, and sonicated to 2Kb fragments. The chromatin of interest was then purified by immunoprecipitation with either anti-Cas9 (green) or anti-FLAG antibody (purple). Molecules (DNA, proteins, and others) associated with the target genomic region were isolated and characterised.

Adapted from *In Situ* Capture of Chromatin Interactions by Biotinylated dCas9, Liu X; *et al.* Cell. 2017.

Image created in BioRender by author.

Following chemical cross-linking, the chromatin was exposed and washed with a series of buffers to remove non-DNA bound materials. The naked chromatin is sonicated generating fragments of approximately 2Kb in length (**Appendix F Table S1**). The sonicate is then subjected to stringent washes to ensure only dCas9-tethered chromatin segments are isolated. The purified chromatin-associated protein complexes are recovered under denaturing conditions, digested to peptides for identification by liquid chromatography-mass spectrometry. These advanced modalities should provide unprecedented insight into the mechanisms of transcriptional regulation by HDAC inhibitors that control *G6PD* expression.

1.7 AIMS

The project focuses on the classical HDACs as previous work by our group established that inhibitors targeting these HDACs selectively induce *G6PD* transcription and increase enzyme activity. By utilising two novel ChIP techniques: Click ChIP (to understand all SAHA interactions) and, CAPTURE (to discover specifically how SAHA exerts its effect on *G6PD* activity). I aim to decipher the mechanism by which SAHA, a pan HDAC inhibitor, exerts its epigenetic effect with the aim of developing more targeted agents.

The specific objectives were:

1. to determine whether the addition of a molecular tag to SAHA affects the transcriptomic output of the drug,
2. to employ locus specific chromatin analysis of responsive and non-responsive GPPP genes to identify histone modifications and chromatin binding proteins that determine HDACi induction of GPPP genes,
3. to optimise technique for genome-wide mapping of 'tagged' HDAC inhibitor binding to chromatin, and
4. to use these approaches to predict specific targets for *G6PD* transcriptional regulation induced by HDAC inhibition to be validated by functional characterisation.

Chapter Two

Material and Methods

2 MATERIALS AND METHODS

2.1 CELL LINES

Human embryonic kidney (HEK) 293T is a cell line derived from HEK 293 that expresses SV40 large T antigen. This confers aminoglycoside resistance so that the transfected plasmid that carry the SV40 origin of replication will maintain a high plasmid copy number. This adherent cell line was used for the early optimisation of G6PD induction by HDAC inhibitors, and a subclone – purchased from Takara – was used for plasmid transfection. Lenti-X 293T is a subclone of HEK 293T which is highly transfectable and supports high levels of viral protein expression. It is able to produce up to six times more virus than HEK 293T (201).

GM12878 is a lymphoblastoid cell line, obtained from the Coriell Cell Repositories. This suspension cell line is derived from peripheral B cells of a healthy Caucasian female with normal karyotype, which were immortalised by Epstein-Barr Virus transformation.

Purchased from ATCC, K562 is an immortalised erythroleukaemia cell line established from the pleural effusion of a 53-year-old female with chronic myeloid leukaemia in terminal blast crisis. This suspension cell line is widely used as a model for red cell biology, biochemistry and erythropoiesis.

2.1.1 Cell culture

Cell line	Tissue of origin	Media	Culture
K562	CML – blast crisis	RPMI+5% FBS	Suspension
GM12878	Human B cell	RPMI+15% FBS	Suspension
293T	Human kidney	DMEM+10% FBS	Adherent

Table 2.1 Cell lines. List of cell lines used and details of culture condition. The suspension cell lines were cultured in RPMI 1640 medium (Sigma-Aldrich), supplemented with varying concentrations of FBS and penicillin 500 IU/ml - streptomycin 500µg/ml

Adherent cell lines were cultured in 75cm² culture flasks in Dulbecco's Modified Eagle Medium (DMEM) (Sigma-Aldrich). Cultures are maintained at approximately 4-8x10⁵ cell/ml to ensure optimal growth and were fed every three days. They were supplemented with 10% Fetal Bovine Serum (FBS) (Sigma-Aldrich), 2mM L-glutamine and penicillin 500 IU/ml - streptomycin 500µg/ml (Stem Cell Technologies). They were incubated at 37°C and 5% CO₂. Once culture achieved approximately 70% confluency, the culture cells were detached using 1x trypsin-EDTA solution (Sigma-Aldrich), split and re-plated into fresh culture flasks.

The exclusion method was used to determine viable cell density. Cells were stained with trypan blue (Sigma-Aldrich) and live cells were counted on a Neubauer haemocytometer. The average number of live cells counted in two sets of 16 corner squares was multiplied by 10,000. They were then multiplied by 2 to correct for the 1:2 dilution from the addition of dye.

All cell lines were tested for mycoplasma contamination on a monthly basis.

2.1.2 Large-scale cell culture

Quantitative MS analysis of global histone PTMs will be used to identify histone modifications. A collaborator at Imperial College has developed a targeted CHIP-MS workflow that provides holistic label-free quantification of more than 30 histone PTMs. The unprecedented sensitivity of this technology with CAPTURE to characterise the locus-specific changes, in an unbiased manner, in histone PTMs enriched from the *G6PD*, *TPI* and non-targeting control (*GAL4*) promoters for quantitative MS analysis of all histone H3 and H4 PTMs.

This protocol generally require starting material in the order of 10^7 cells (due to the dynamic range of histone PTM abundance) (202). However, to characterise the locus-specific changes in promoter regions of *G6PD*, 1-2Kb in size (**Figure 2.5**), an estimated 10^9 cells would be required (203, 204). Therefore, large-scale cell culture techniques were adopted. CAPTURE cell lines were cultured in 1L, non-baffled, Erlenmeyer flasks. These cells were cultured in RPMI 1640 medium, supplemented with 5% FBS and penicillin 500 IU/ml - streptomycin 500µg/ml. They were subjected to continuous shaking at 100 rpm in the New Brunswick™ S41i incubator shaker (Eppendorf®) maintained at 5% CO₂ and 37°C. Cultures were maintained at approximately $5-10 \times 10^5$ cell/ml to ensure optimal growth, and were fed every three days, with inclusion of the required antibiotics for selection.

2.1.3 Histone deacetylase inhibitor induction

Four drugs were used in this study, two different HDAC inhibitors and two additional, modified, forms of suberoylanilide hydroxamic acid (SAHA). These were prepared in accordance with the manufacturer's advice.

Sodium butyrate was purchased from Sigma-Aldrich as a crystalline solid. A stock solution was made by dissolving the powder in PBS to a concentration of 600mM; this was stored at -20°C for a maximum of one week.

SAHA was sourced from Sigma-Aldrich as a powder that was reconstituted in DMSO at a concentration of 15mg/ml. The stock solution was prepared by diluting in PBS to a concentration of 600μM and was stored at -20°C. Immediately prior to use, it was further diluted to 100μM in PBS.

Biotinylated SAHA was designed and produced by Sai Life Sciences, on request of the Karadimitris group. The modified SAHA was reconstituted and diluted following the same protocol as parental SAHA.

SAHA BPyne, a clickable form of SAHA (Cambridge Bioscience) arrived as a solution in methanol, at a concentration of 1μg/2μl. Immediately prior to use it was diluted to a concentration of 100μM in PBS.

In all cases any working solutions were brought to room temperature prior to use, and discarded if not used within 4 hours.

2.1.4 HDAC inhibitor induction assays

Optimal conditions for selective G6PD induction by HDAC inhibitors were established by induction assays.

Makarona *et al.* recognised that HDAC inhibitors – such as sodium butyrate and SAHA – selectively induce *G6PD* transcription with resultant increase in protein expression, and restoration of G6PD enzymatic activity to near normal levels in non-haematopoietic cells and haematopoietic cell lines (74). To replicate this induction assay, dose response and time course experiments were performed to identify optimal conditions for 2-fold increase in transcription compared to a reference gene, β -actin.

Cells were cultured in 75cm² flasks, plated into a 6 well TC-treated plate (Corning®), at concentration of approximately 1x10⁶ cells/ ml, and made up to 2ml with fresh media. The plated cells were then incubated for varying durations with fixed HDAC inhibitor concentrations alongside equal concentrations of vehicle – time course. The cells were then harvested, RNA extracted, and cDNA synthesised as described in section 2.4.1.1. Finally, RT-qPCR was performed, and data analysed (section 2.4.1.3). A similar protocol was followed for dose response experiments. Here, the incubation time was fixed and drug concentration altered.

2.2 DEVELOPMENT OF CAPTURE CELL LINES

The establishment of cell lines expressing dCas9 (deactivated Cas9) and sequence-specific sgRNA (single guide RNA) enabled sequence specific pulldown of targeted genomic loci in an

unbiased manner. Initially, cell constructs were designed as described by Fujita, *et al*: a two-plasmid system (202). This consists of a lentivirus plasmid of approximately 15Kb, pLVhUbC-dCas9T2aGFP (**Figure 2.1**), obtained from Addgene. This expresses 3xFLAG tagged nuclease deficient Cas9, with a nuclear localization signal (NLS) sequence, and GFP selectable marker. The second plasmid, pU6-sgX-NT1 (**Figure 2.2**), also sourced from Addgene, is a sgRNA expression vector with a mCherry cassette (205). Four guide RNA inserts were designed, **Figure 2.5**; two targeting different regions of the *G6PD* gene. A third, *TPI* (Triose phosphate isomerase) directed guide RNA. TPI is a GPPP enzyme that is not induced by HDAC inhibitors (74). A final guide RNA, *Gal4*, is a fungal transcription factor, so has a non-cognate target in the human genome (205). The targeting guide RNAs were each cloned into the pU6-sgX-NT1 plasmid, creating four distinct constructs.

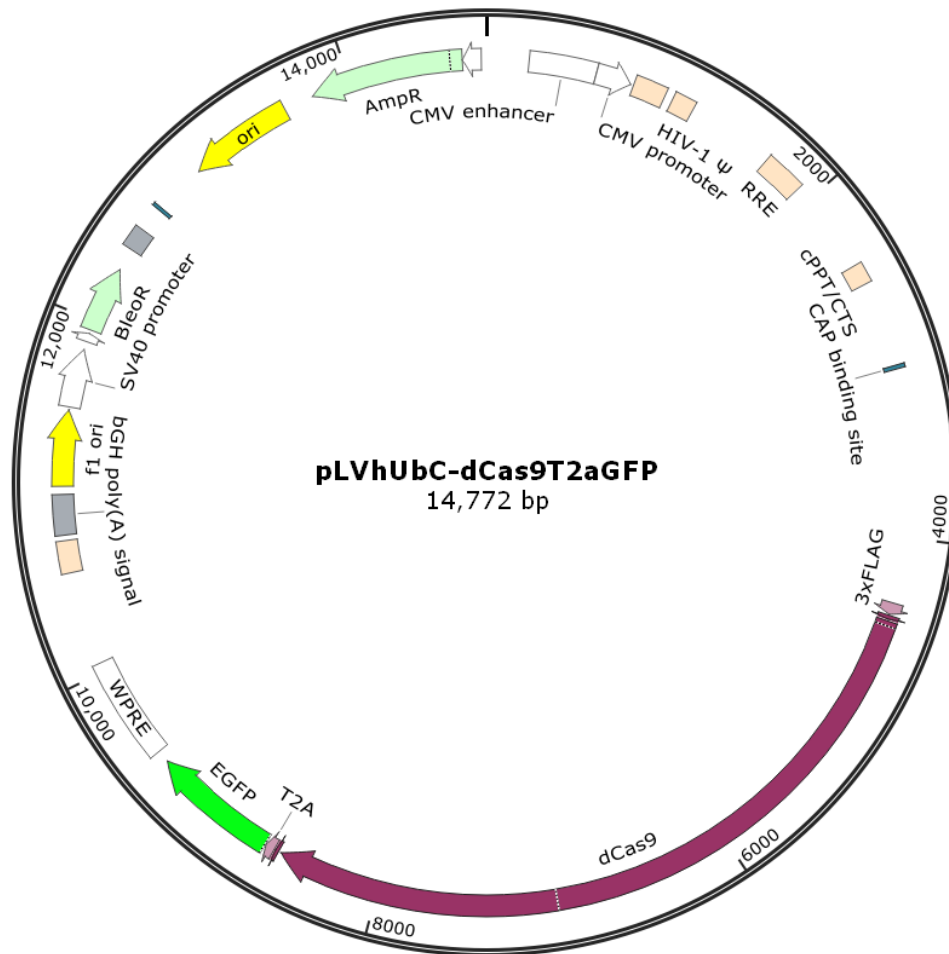


Figure 2.1 map of pLVhUbC-dCas9T2aGFP lentiviral vector that co-expresses deactivated Cas9 and enhanced GFP (eGFP). Key features include dCas9 (purple) - catalytically 'dead' mutant of the Cas9 endonuclease from the *Streptococcus pyogenes* Type II CRISPR/Cas system. EGFP (green) is the fluorescent reporter of expression. 3xFLAG tag - three tandem FLAG® epitope tags. 5' LTR - truncated 5' long terminal repeat flanks the transgene sequence that facilitates the integration of transfer plasmid sequences into host genome. The transgene sequence also contains WPRE, RRE and cPPT/CTS which promote transgene expression. Woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) increases transgene expression by facilitating mRNA transcript maturation. Central polypurine tract/central termination sequence (cPPT/CTS) improves vector integration and transduction efficiency. Light green – antibiotic, bleomycin (BleoR) and ampicillin (AmpR) resistance genes. Each replicon for antibiotic resistance gene has an origin of replication (*ori*) which enables plasmid self-replication. These genes have an associated Poly(A) signal which is important for the nuclear export, translation and stability of mRNA; and a CAP binding site which activates transcription in presence of cAMP.

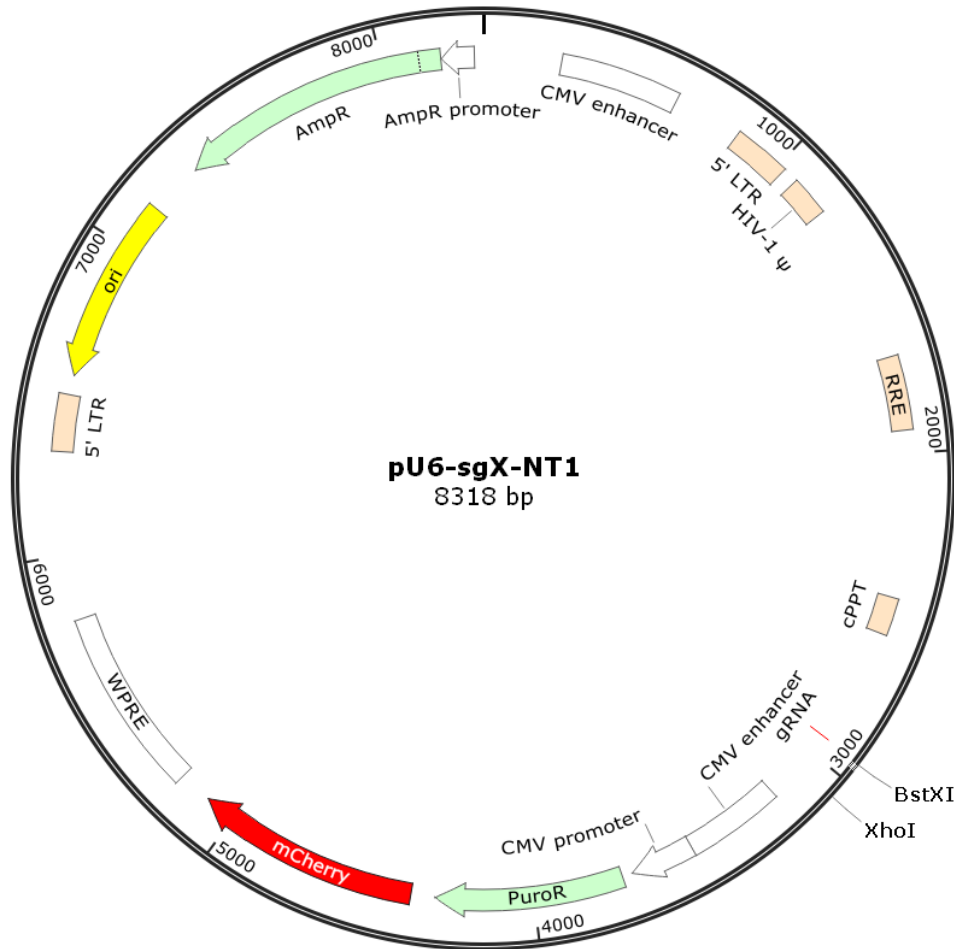
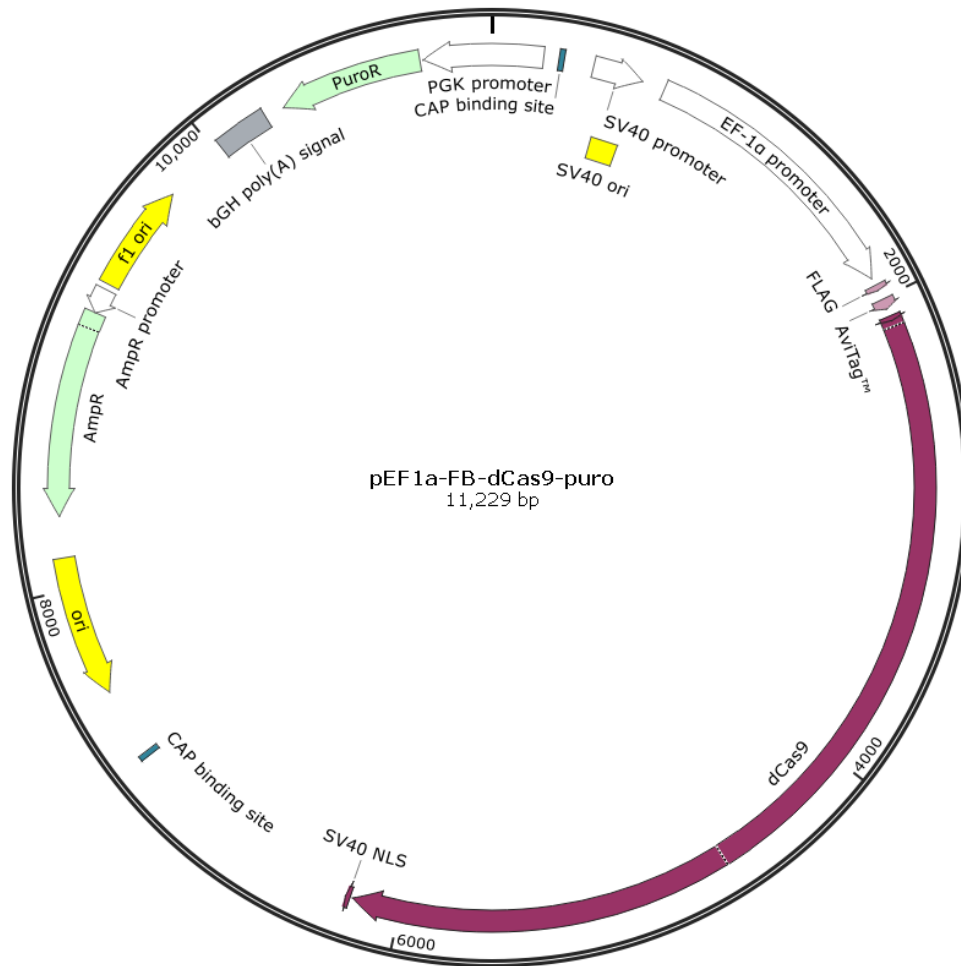


Figure 2.2 pU6-sgx-NT1 - expresses a sgRNA under U6a promoter with mCherry and puromycin selectable markers. Key features include insertion site for guide RNA (gRNA) flanked by XhoI and BstXI restriction sites. MCherry (red) is the fluorescent reporter of expression. 5' LTR - truncated 5' long terminal repeat flanks the transgene sequence that facilitates the integration of transfer plasmid sequences into host genome. The transgene sequence also contains WPRE, RRE and cPPT that promote transgene expression. Woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) increases transgene expression by facilitating mRNA transcript maturation. Central polypurine tract (cPPT) improves vector integration and transduction efficiency. Light green – antibiotic, puromycin (PuroR) and ampicillin (AmpR) resistance genes. Each replicon for antibiotic resistance gene has an origin of replication (ori) which enables plasmid self-replication. These genes have an associated Poly(A) signal which is important for the nuclear export, translation and stability of mRNA; and a CAP binding site which activates transcription in presence of cAMP.

The size of the pLVhUbC-dCas9T2aGFP vector was prohibitive, limiting transduction to target cells. To overcome this problem, it was replaced by two smaller vectors to create a three-plasmid system, as described by Liu, *et al* (195). This novel system employed *in vivo* biotinylation of dCas9-pEF1a-FB-dCas9-puro (**Figure 2.3A**) by a biotin ligase - pEF1a-BirA-V5-neo (**Figure 2.3B**), and sequence specific gRNAs (**Figure 2.2**).

A



B

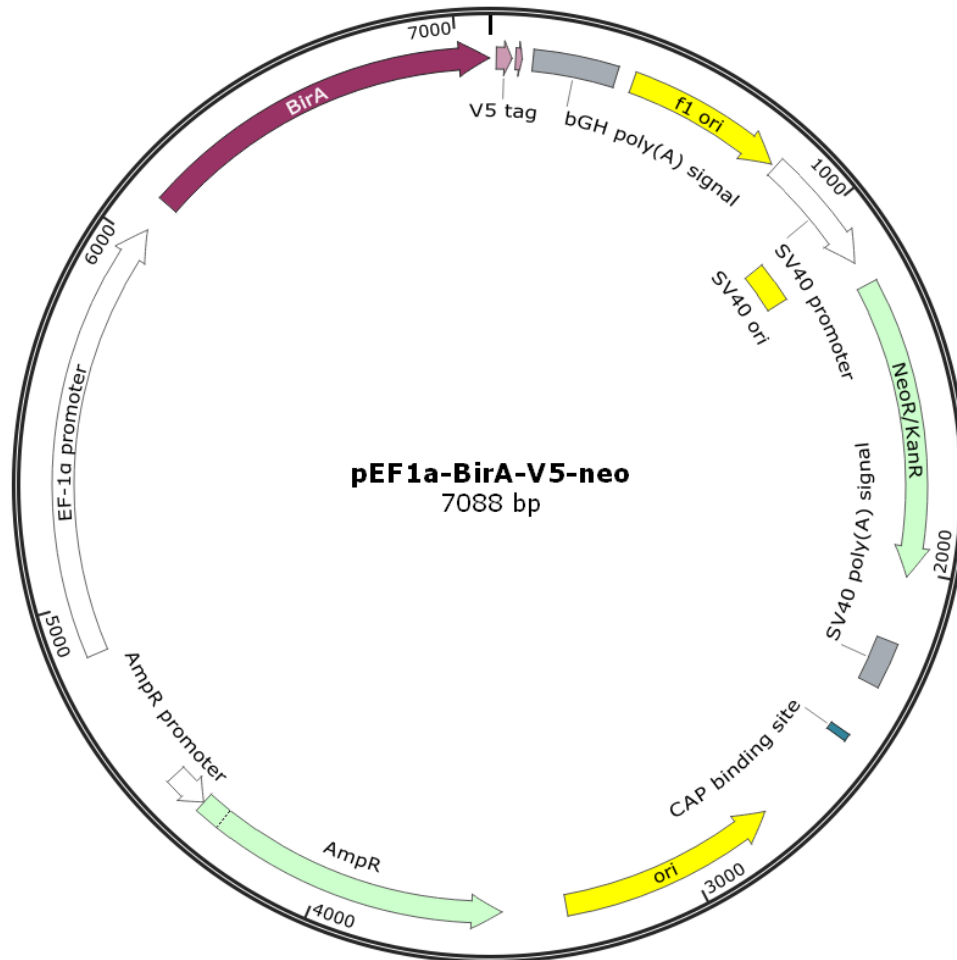


Figure 2.3 CAPTURE vectors.

(A) pEF1a-FB-dCas9-puro contains N-terminal Flag and biotin-acceptor-site (FB)-tagged dCas9 protein driven by EF1a promoter, selectable marker – puromycin resistance. Key features include dCas9 (purple) - catalytically dead mutant of the Cas9 endonuclease from the *Streptococcus pyogenes* Type II CRISPR/Cas system. This is flanked by FLAG tag (epitope tag), and AviTag™, a peptide tag that allows for enzymatic biotinylation, at one end. The other end is flanked by SV40 NLS, nuclear localization signal. The ampicillin (AmpR) and puromycin (PuroR) resistance genes (light green) are accompanied by ori, unique promoters, polyadenylation signals and CAP binding sites.

(B) pEF1a-BirA-V5-neo expresses C-terminal V5-tagged BirA biotin ligase driven by EF1a promoter with neomycin resistance cassette, selectable marker – G418 resistance. Key features include BirA (purple) - enzymatic biotinylation with *E. coli* biotin ligase which is highly specific in covalently attaching biotin to the 15 amino acid AviTag peptide. This is associated with a V5 tag (light purple). The ampicillin (AmpR) and G418, a neomycin analogue, (NeoR/KanR) resistance genes (light green) are accompanied by ori, unique promoters, polyadenylation signals and CAP binding sites.

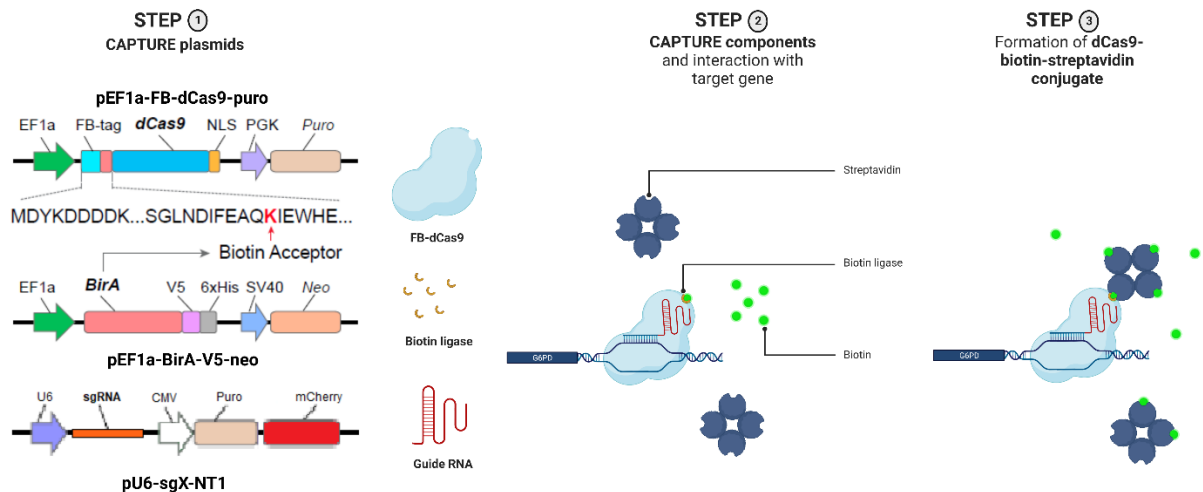


Figure 2.4 Schematic of dCas9-mediated CAPTURE system

The CAPTURE protocol describes a modification of enChIP optimising the purification of locus specific chromatin. The system consists of a fusion protein – deactivated Cas9, FLAG tag, biotin acceptor and nuclear localization signal - encoded by pEF1a-FB-dCas9-puro plasmid. pEF1a-BirA-V5-neo expresses biotin ligase (BirA) and neomycin resistance selection marker. pU6-sgX-NT1 plasmid, encodes for a locus specific guide RNA that direct the dCas9 to the gene of interest. These constructs interact to form a dCas9-biotin-chromatin complex which is purified following streptavidin pull down. Molecules (DNA, proteins, and others) associated with the target genomic region were isolated and characterised.

Adapted from *In Situ* Capture of Chromatin Interactions by Biotinylated dCas9, Liu X; *et al.* Cell. 2017.

Image created in BioRender by author.

2.2.1 Insert design

Work by Davie, *et al* identified butyrate response elements (185). This was elegantly illustrated and further defined using promoter deletion analysis, in the Karadimitris lab (74). For this reason, a 200bp sequence upstream of the HDAC inhibitor responsive region was selected as early work indicated that guides proximate to this region result in transcriptional interference by dCas9 and a significant reduction in G6PD induction (**Figure 2.5**).

An online filtering tool, <http://crispr.mit.edu/guides/>, was used to generate potential sgRNAs for the 200bp region described above. This tool predicted genomic off-target

binding allowing selection of guides highly specific for the region of interest (206, 207). Each sgRNA was then integrated into the generic forward primer (**Table 2.2**); this table also details the primers purchased from Sigma-Aldrich Company Ltd.

Generic insert sequence

GTATCCCTTGGAGAACCACCTTGTTGGNNNNNNNNNNNNNNNNNNNNGTTTAAGAGCTATGCTGG
AAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT
GCTTTTTTTCTCGAGTACTAGGATCCATTAGG

G6PD sgRNA 22 insert sequence

GTATCCCTTGGAGAACCACCTTGTTGGGGAAGGGCGACCGCGAAACTGTTTAAGAGCTATGCTGGA
AACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTG
CTTTTTTTCTCGAGTACTAGGATCCATTAGG

G6PD sgRNA 23 insert sequence

GTATCCCTTGGAGAACCACCTTGTTGGGAGCGGGCGGACTGTGAACGCGTTTAAGAGCTATGCTGGA
AACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTG
CTTTTTTTCTCGAGTACTAGGATCCATTAGG

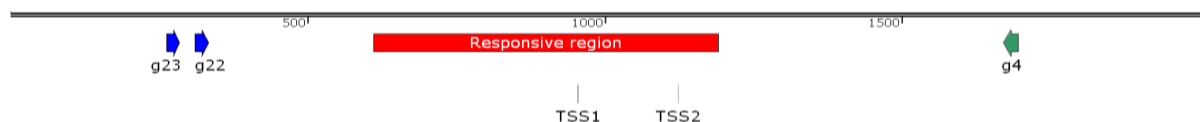


Figure 2.5 Genetic map of *G6PD* promoter

G6PD regulatory region (red) is the site that HDAC inhibitor binds to alter gene expression. The *G6PD* gene has two transcription start sites, TSS1 and TSS2. In blue are sites of the gRNAs downstream of the HDAC responsive region and therefore do not interfere with HDAC inhibitor induction of *G6PD* expression. Upstream of responsive region is g4 (green) that does interfere with upregulation of *G6PD* expression.

TPI sgRNA insert sequence

GTATCCCTTGGAGAACCACCTTGTTGGTTATATAGAGCGCCGCGAAGGTTTAAGAGCTATGCTGGAA
ACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
TTTTTCTCGAGTACTAGGATCCATTAGG

Gal4 sgRNA insert sequence

GTATCCCTTGGAGAACCACCTTGTTGGGAACGACTAGTTAGGCGTGTAAGTTTAAGAGCTATGCTGGA
AACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTG
CTTTTTTCTCGAGTACTAGGATCCATTAGG

Name	sgRNA	Primer
pU6 sgRNA X forward	-	GTATCCCTTGGAGAACCACCTTGTTG GNNNNNNNNNNNNNNNNNNNNNNNGTT TAAGAGCTATGCTGGAAAC
pU6 sgG6PD 22 forward	GGAAGGGCGACCGCGAAACT GGG	GTATCCCTTGGAGAACCACCTTGTTG GGGAAGGGCGACCGCGAAACTGTTT AAGAGCTATGCTGGAAAC
pU6 sgG6PD 23 forward	GAGCGGCGGACTGTGAACGC TGG	GTATCCCTTGGAGAACCACCTTGTTG GGAGCGGCGGACTGTGAACGCGTTT AAGAGCTATGCTGGAAAC
pU6 sgTPI forward	TTATATAGAGCGCCGCGAAG TGG	GTATCCCTTGGAGAACCACCTTGTTG GTTATATAGAGCGCCGCGAAGGTTT AAGAGCTATGCTGGAAAC
pU6 sgGAL 4 forward	GAACGACTAGTTAGGCGTGTA	GTATCCCTTGGAGAACCACCTTGTTG GGAACGACTAGTTAGGCGTGTAAGTT TAAGAGCTATGCTGGAAAC
pU6 sgRNA universal reverse primer	-	TAATGGATCCTAGTACTCGAGAAAA AAAGCACCGACTCGGTGCCACTTTTT CAAGTTGATAACGGACTAGCCTTATT TAACTTGCTATGCTGTTTCCAGCAT AGCTCTT

Table 2.2 sgRNA specific pU6-sgX-NT1 primers for inserts. PAM sequence in green, integrated sgRNA sequence in red.

2.2.2 Cloning of pU6-sgX-NT1 vector

The four sgRNAs were each cloned into the lentiviral U6-driven expression vector by amplifying the insertions using a common reverse primer and unique forward primers as shown in **Figure 2.6**. This was achieved by linearising pU6-sgX-NT1 vector with BstXI and XhoI restriction enzymes (Thermo Scientific) to produce unique restriction sites complementary to the insert. The linearised vector was then isolated from intact plasmid and purified by agarose gel electrophoresis followed by extraction using the Monarch® DNA Gel Extraction Kit (New England BioLabs). The insert was amplified by PCR with high-fidelity DNA polymerase and the primers detailed in **Table 2.2**, then digested with the two restriction enzymes used above, BstXI and XhoI. Following purification the linearised vector and fragmented insert were quantified using Nanodrop™ Lite spectrophotometer (Thermo Scientific), then ligated using the T4 DNA ligase (Thermo Scientific), in accordance with the manufacturer's protocol (incubation at 22°C for 10 minutes).

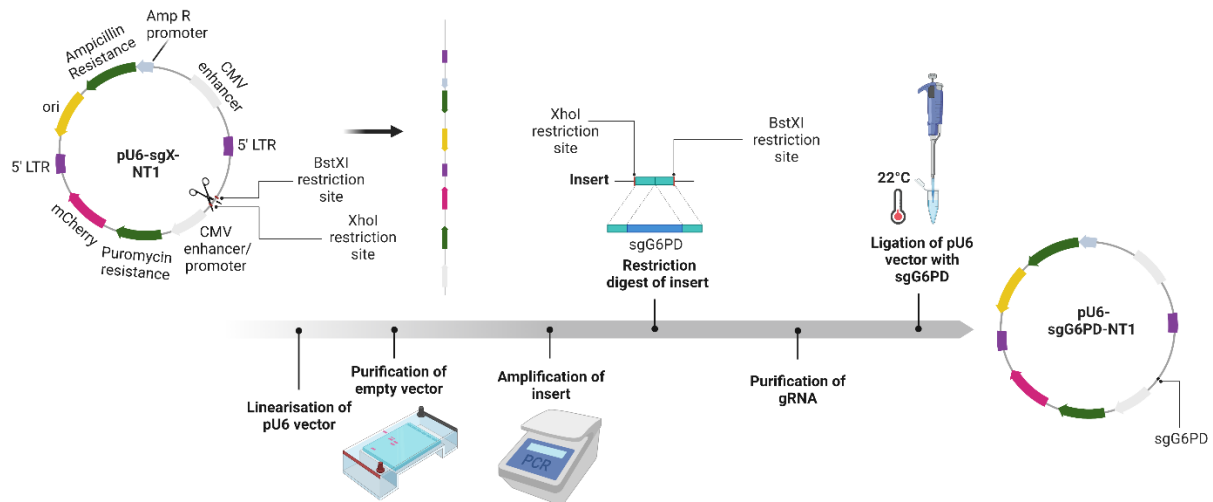


Figure 2.6 Schematic of pU6-sgG6PD-NT1 vector cloning strategy

pU6-sgX-NT1 and specific guide RNA insert were double digested with BstXI and XhoI restriction enzymes, creating complementary cohesive ends. Following purification of the gel extracted linearised plasmid and amplification of the digest, the fragments were ligated, to create a construct containing the required gRNA for transduction to the CAPTURE cell line.

Image created in BioRender by author

2.2.3 Transformation of competent cells

Following validation of successful ligation of gRNA to plasmid backbone by Sanger sequencing, the plasmid was amplified using DH5α competent *E. coli* cells (New England BioLabs). Immediately, after thawing on ice, 10ng of ligation reaction was introduced to the bacterial culture and carefully mixed. The mixture was incubated on ice, then heat shocked for 30 seconds at 42°C. After a brief period of cooling, the transformed cells were incubated for 1 hour at 37°C with constant shaking. The culture was spread on a Luria-Bertani (LB) agar plate containing ampicillin (100ng/ml) overnight. The following day, the plates were reviewed, and 8-12 colonies identified.

2.2.4 Amplification of pU6 vector

Each colony was isolated, picked and incubated overnight at 37°C in LB broth with ampicillin (100ng/ml). The cells were harvested and plasmid DNA was extracted with GeneJET Plasmid Miniprep Kit (Thermo Scientific), as per manufacturer's protocol. The resulting DNA samples were screened by diagnostic digest with BstXI and KpnI (Thermo Scientific) for positive clones which were confirmed by Sanger sequencing (**Appendix D Figure S 1-2**).

Following confirmation of transformation of competent cells with cloned pU6 plasmid the positive colonies were cultured overnight to an optical density of 2-2.5. The cells were then harvested, and endotoxin-free plasmid DNA was extracted with the GeneJET Plasmid Maxiprep Kit (Thermo Scientific), as per manufacturer's protocol. The plasmid DNA was quantified by Nanodrop™ Lite spectrophotometer.

2.2.5 Electroporation of plasmids for CAPTURE cell lines

The Neon® Transfection System (Invitrogen) was used to simultaneously transfect native K562 cells with the two pEF1a plasmids. The protocol is outlined as follows: high quality plasmid DNA, 1–5µg/µl, was prepared in elution buffer. 1×10^7 cells were then harvested, washed with PBS, resuspended in resuspension buffer R and the transfection system was setup, as per manufacturer's instructions. The cells were then electroporated using the parameters detailed in **table 2.3**, below.

Pulse voltage (V)	Pulse width (ms)	Pulse number	Cell density (cells/ml)	Transfection efficiency	Viability	Tip type
1,450	10	3	10 million	90%	90%	10µl

Table 2.3 Electroporation parameters used to transduce cell lines with pEF1a-FB-dCas9-puro and pEF1a-BirA-V5-neo. Optimised for K562 cell line

Following transfection, the cells were plated in a 10cm culture dish, incubated in RPMI media which was supplemented with 5% FBS and 1% pen-strep, without selection antibiotics, at 37°C in a humidified CO₂ incubator for 24 hours, after which the media was replaced. A further 24 hours later, the media was changed and selection antibiotics, puromycin and G-418 at final concentrations of 3µg/ml and 600µg/ml respectively, were added. Media with antibiotics were changed every three days.

2.2.6 Lentiviral production and concentration

Here, the pU6-sgX-NT1 plasmid was integrated into lentiviral pseudoparticles using HEK 293T cells, by transfection (**Figure 2.7**), by the calcium phosphate method (208).

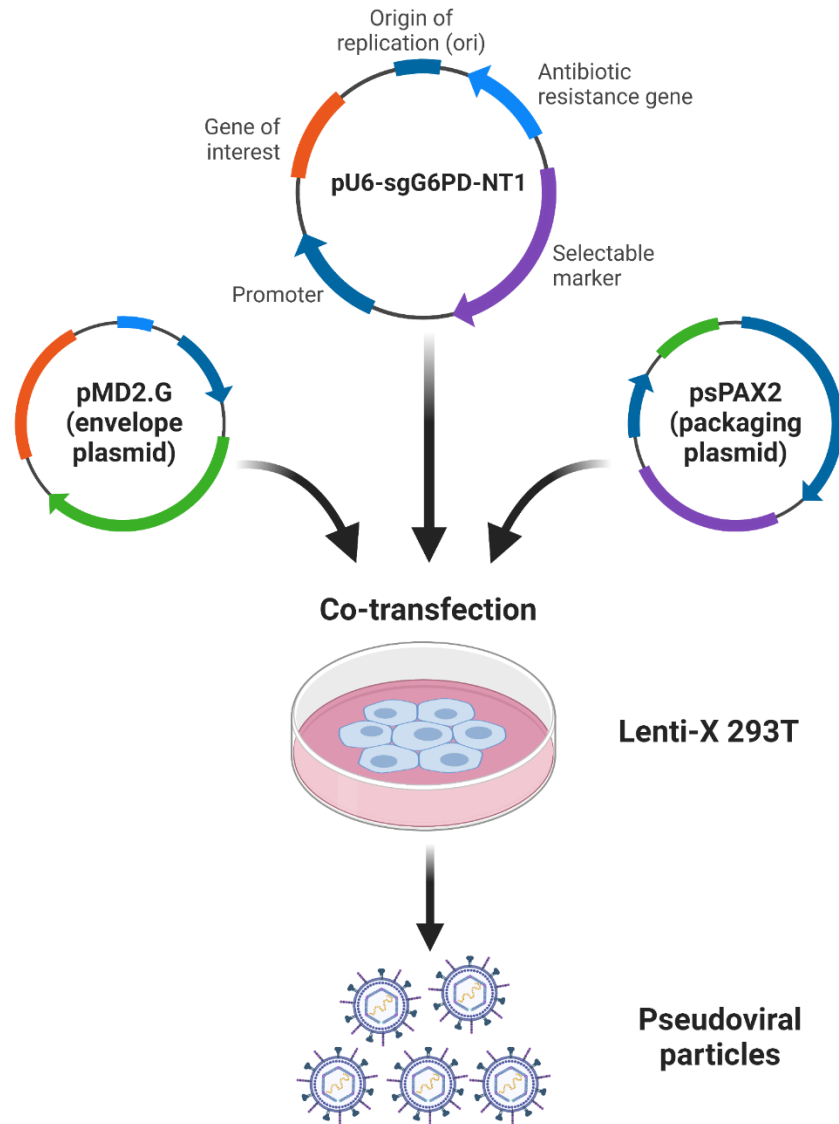


Figure 2.7 Schematic illustrating transfection of Lenti-X 293T cells with pU6-sgG6PD-NT1
 2×10^5 293T cells were harvested, transfected with the target plasmid (pU6-sgG6PD-NT1) and two helper plasmids, then incubated with calcium phosphate cocktail. Following confirmation of virus production by fluoroscopy. The virus was then harvested, pooled, snap frozen and stored for future use.
 Image created in BioRender by author

Early passage (<P12) Lenti-X 293T were seeded in 75cm² culture flasks to approximately 70% confluency then plated 3×10^6 cells into a 10cm culture dish; five dishes were required for a single round of virus production, and incubated for 24-30 hours prior to transfection. Next, culture medium was replaced, and plasmid cocktail prepared, detailed in **Table 2.4**.

Component	Quantity
<i>Transfer plasmid</i>	10µg
pMD2.G	3µg
psPAX2	6µg
Low TE buffer	up to 450µl
CaCl ₂ (2.5M)	50µl
2xHBS	500µl
Total volume	1ml

Table 2.4 Calcium phosphate methods for virus production

14-16 hours later, the Calcium-Pi-precipitated plasmid DNA was replaced by fresh media. At 48- and 72-hours post transfection, the viral supernatant was collected, centrifuged and filtered with a 0.45µm PVDF membrane filter to remove cell debris. The supernatant was centrifuged at 23,000rpm for 90 minutes, using the Thermo Sorvall Ultracentrifuge, to pellet the viral particles which were then resuspended in FBS free media. The concentrated virus was agitated at 180rpm, pooled, aliquoted, snap frozen and stored at -80° C.

2.2.7 Cellular transduction

2x10⁵ K562 expressing cells pEF1a-FB-dCas-puro and pEF1a-BirA-V5-neo plasmids were seeded in a 24-well plate. Concentrated lentivirus was removed from a -80°C freezer and allowed to thaw at room temperature. 100µl of freshly thawed virus was diluted with warmed RPMI media and were added to each well. Polybrene (Sigma-Aldrich) was then added to the target cells to a final concentration of 8µg/ml. Spinoculation was performed by centrifuging plate at 2,000g for 2 hours at 32°C and the plate was placed in the incubator overnight. At 16-20 hours post spinoculation the media was replaced and returned to the

incubator. At approximately 72 hours after transduction, the transduction rate was checked by fluorescent microscopy, with a suitable filter, to confirm sufficient transduction was achieved. The BD FACSAria™ Fusion was used to sort, using the mCherry selectable marker to identify the transduced cells with pU6 plasmid (Figure 2.2) and cultured in RPMI 1640 supplemented with 20% FBS and 1% Penicillin-Streptomycin.

2.3 PROTEIN EXPRESSION ASSAYS

2.3.1 Western blot

2.3.1.1 Sample preparation

2×10^6 cells were harvested, washed with cold PBS, and lysed using RIPA buffer (Sigma-Aldrich) and phenylmethylsulfonyl fluoride (PMSF) (Sigma-Aldrich) in a 1:100 ratio.

Following shaking, the lysate was incubated at -80°C , then protein quantified using Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). Equal quantities of protein were aliquoted from each sample and combined with appropriate volumes of 3x SDS loading buffer then samples were denatured by boiling at 100°C for 5-10 minutes.

2.3.1.2 Running gel and transfer of protein to membrane

Following high-speed centrifugation to remove debris, samples were loaded onto a pre-cast 4-12% Tris-Glycine gel (Novex™ WedgeWell™) and ran at 100V, alongside PageRuler™ Plus Prestained Protein (Thermo Scientific) and Blue Prestained Protein Standard (New England Biolabs) ladders to determine molecular weights. Electric current was applied at appropriate settings, increasing after the dye front had reached the stacking gel, first demarcation, until the dye front reached the bottom of the gel. Once the gel had run, the protein was transferred to a nitrocellulose membrane, in a semi-dry approach, using the Trans-Blot® Turbo™ Transfer System (Bio-Rad), as per manufacturer's instructions.

2.3.1.3 Antibody staining

The membrane was rinsed with 0.1% Tween20 in PBS (PBST) then blocked overnight in blocking buffer (5% skim milk (Sigma-Aldrich) in PBST) at 4°C. The membrane was rinsed again three times with PBST and incubated overnight at 4°C in primary antibody diluted in blocking buffer (**Table 2.5**). The membrane was once again rinsed three times, as above, then incubated for 1 hour at room temperature in the secondary antibody diluted in blocking buffer (**Table 2.5**). After rinsing the membrane for a fourth, and final time in PBST, the membrane was incubated for 10 minutes with Enhanced Chemiluminescence detection reagent (Thermo Fisher Scientific) and visualised using the Bio-Rad Blot Imager.

PRIMARY ANTIBODY	SPECIES	DILUTION	SUPPLIER
α -tubulin (62204)	Mouse monoclonal	1:5000	Thermo
Cas9 (C15310258)	Rabbit polyclonal	1:1000	Diagenode
Cas9 (7AG-3A3)	Mouse monoclonal	1:500	CST
GAPDH (SC-25778)	Rabbit polyclonal	1:3000	Santa Cruz Biotechnology
FLAG (F1804)	Mouse monoclonal	1:1000	Sigma
G6PD (ab993)	Rabbit polyclonal	1:10,000	Santa Cruz Biotechnology
Streptavidin- HRP (405210)	-	1:3000	BioLegend
Anti V5-HRP (46-0708)	Mouse monoclonal	1:5000	Invitrogen

SECONDARY ANTIBODY	SPECIES	DILUTION	SUPPLIER
α -rabbit HRP	Goat polyclonal	1:3000	Dako, Ltd
Anti-CRISPR/Cas9-FITC clone (SAB420735)	Mouse monoclonal		Sigma
α -goat HRP	Rabbit polyclonal	1:3000	Dako, Ltd

Table 2.5 Antibody dilutions for western blotting

2.3.2 Co-immunoprecipitation

Co-immunoprecipitation (Co-IP) is an extension of immunoprecipitation (IP) that is based on the potential of IP reactions to capture and purify the primary target (the antigen) as well as other macromolecules that are bound to the target by native interactions in the sample solution. This process is based on the principle that if a known protein is a member of a large protein complex, then the whole protein complex will be pulled down from the solution using its specific antibody and can be used to identify other members of this complex using a secondary antibody in a technique, such as western blot. Here Co-IP was used to demonstrate the *in vivo* biotinylation of dCas9.

2.3.2.1 Sample preparation

1×10^7 cells were harvested and washed three times with pre-chilled PBS. They were then resuspended in a 1:100 mixture of PMSF in RIPA lysis buffer (Sigma-Aldrich™), agitated briefly at 4°C, followed by incubation at -80°C. Cell debris was pelleted by high-speed centrifugation and the lysate collected.

2.3.2.2 Immunoprecipitation

The immune complex was formed overnight by gently rotating the lysate at 4°C with polyclonal Cas9 antibody (Diagenode). Equal volumes of washed Dynabeads-protein A and G beads (Thermo Fisher Scientific) were added to the lysate/antibody mix, then incubated overnight at 4°C with gentle rotation. The beads were washed three times with RIPA lysis buffer and 1:100 SIGMAFAST protease inhibitor cocktail (Sigma-Aldrich™). The beads were then resuspended in 3x SDS buffer, boiled and the supernatant was collected. Finally, Western Blot probing for biotin was performed, as described above.

2.3.3 Immunofluorescence

Immunofluorescent labelling was employed to visualise subcellular distribution of biotinylated dCas9 within the CAPTURE cell lines.

1×10^6 K562 cells were harvested by centrifugation at 800g. The cells were fixed with 3.7% methanol free formaldehyde, then permeabilized with a mixture of 0.05% Triton X-100 and 1% w/v milk in PBS (209). Prior to immunostaining, the cells were blocked with 2% BSA, then incubated overnight, in the dark, with primary antibody diluted to desired concentration

with 0.1% BSA (**Table 2.6**). Following washes with PBST, the cells were further incubated in secondary antibody diluted to the desired concentration in 0.1% BSA (**Table 2.6**), once again ensuring samples are protected from light. After washes with PBST, 1 drop of ProLong® Silver Antifade Reagent with DAPI (Life Technologies Ltd) was added to the cell pellet. The cells were then mounted on a clean microscope slide and covered with a 22x22mm coverslip, cured and sealed. The cells were imaged using deconvolution microscope (Zeiss Axiovert 100) and analysed using ImageJ software.

ANTIBODY	SPECIES	DILUTION	SUPPLIER
STREPTAVIDIN-CY3	-	1:500	BioLegend
CAS9 (C15310258)	Rabbit polyclonal	1:200	Diagenode
ANTI-RABBIT/FITC	-	1:250	BioLegend

Table 2.6 Antibody dilutions for Immunofluorescence

2.3.4 Antibiotic resistance assay

The first critical step for stable cell line generation is determining the optimal antibiotic concentration for selecting stable cell colonies; the optimal concentration is cell type dependent. A kill curve is a dose-response experiment where the cells are subjected to increasing amounts of antibiotic to determine the minimum antibiotic concentration needed to kill all the cells that do not contain the resistance gene (210). pEF1a-FB-dCas-puro vector co-expresses a catalytically inactive Cas9 and puromycin N-acetyltransferase that confers puromycin resistance. While pEF1a-BirA-V5-neo vector expresses both biotin ligase and aminoglycoside phosphotransferase that confers neomycin, kanamycin and G-418 (Geneticin®) resistance.

2.5×10^5 of native K562 cells were plated in a 24-well tissue culture plate, with appropriate media. Increasing concentrations of G-418 antibiotic (Scientific Lab Supplies Ltd.) were added, as illustrated in **Figure 2.8**. Every 2-3 days the media containing selection antibiotic was replaced. Cell viability was assessed by flow cytometry until the lowest concentration of antibiotic that was lethal to >99% of cells was identified.

This process was repeated for puromycin using the concentrations below.

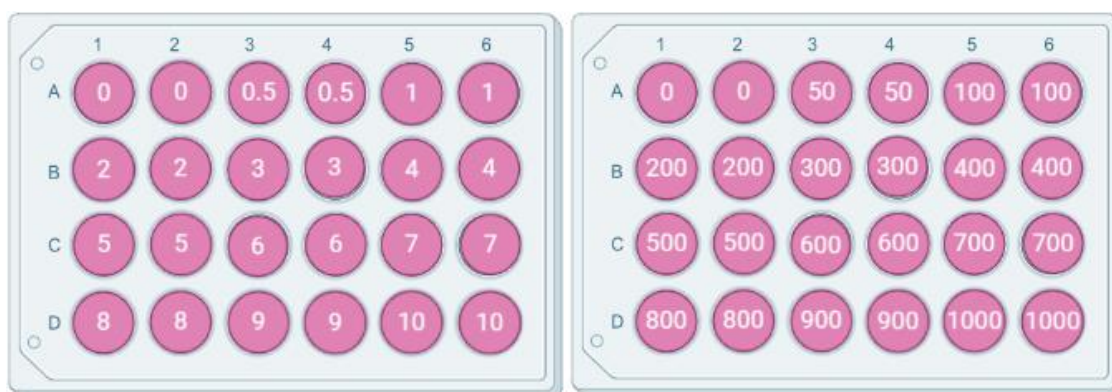


Figure 2.8 24-well plate illustrating antibiotic dilutions for resistance assay, G-418 - (left) and puromycin (right).

To ensure expression of the molecular machinery required for CAPTURE. The CAPTURE cell lines were transiently transduced with pEF1a-BirA-V5-neo – expressing biotin ligase, and pEF1a-FB-dCas-puro – expressing dCas9 and biotin acceptor, along with the stably transduced pU6-sgX-NT1 plasmid. The pEF plasmids are maintained by antibiotic selection pressures. Plate on left shows serial dilutions to determine optimal concentration of G-418 (μg/ml) to ensure pEF1a-BirA-V5-neo expression. And on the right, dilutions to determine optimal concentration of puromycin (μg/ml) to maintain pEF1a-FB-dCas-puro expression.

Image created in BioRender by author.

2.4 GENE EXPRESSION ASSAYS

2.4.1 RT- qPCR assays

2.4.1.1 RNA purification and cDNA synthesis

Harvested cells were washed with cold PBS, then extraction was performed using ReliaPrep™ RNA Purification kit (Promega Ltd.), following manufacturer's protocol. Upon completion, the RNA was quantified, spectrophotometrically, using Nanodrop™ 2000 (Thermofisher Scientific). Immediately following RNA purification, reverse transcription of

1µg of RNA from each sample was carried out using GoScript™ Reverse Transcription kit with random primers (Promega), in accordance with manufacturer's guidance, to generate cDNA for RT-qPCR assays. This is the first step of a two-step RT-PCR process.

2.4.1.2 TaqMan® probe testing

At least two TaqMan® probes (Thermo Fisher Scientific) were tested for each target gene. These were assessed by creating a standard curve to determine the efficiency and linearity. The most efficient probe for each amplicon was selected (**Table 2.7**).

Target gene	Function	Selected TaqMan® probe
G6PD	Inducible GPPP enzyme	Hs00166169_m1
B-actin	Reference gene	Hs01060665_g1
TPI	Non-inducible GPPP enzyme	Hs03806547_s1
ARAP 1	Negative control	Hs00362929_m1
SERPINI 1	Positive control	Hs01115397_m1

Table 2.7 List of TaqMan probes

TaqMan probes used for ChIP-qPCR, time and dose response experiment. Each probe, selected from ≥2 alternatives, following testing

2.4.1.3 RT-qPCR assay and data analysis

The second step of the two step PCR process involves gene expression quantification was performed for each probe in triplicate. cDNA templates generated above were diluted with double distilled water, mixed with TaqMan Fast Universal PCR Master Mix (Applied Biosystems) and one of the above probes. PCR reaction was performed, in 96 well plates with each reaction condition tested in triplicate, using the StepOnePlus™ Real-Time PCR System. The following program settings were used:

Holding Stage – initial denaturation (disrupt RNA/cDNA strands and primer dimer formation)

1. 50°C for 2 minutes
2. 95°C for 10 minutes

Cycling Stage (45 cycles)

1. 95°C for 15 seconds – denaturation (disrupt cDNA/primer formation at beginning of each cycle)
2. 60°C for 1 minute – anneal (promote primer/cDNA binding)

Transcript levels for each gene were compared to reference gene, B-actin, using the delta-delta threshold cycle ($\Delta\Delta CT$) method (211).

2.5 CHROMATIN IMMUNOPRECIPITATION ASSAYS

To address the two main objectives of this project, four different chromatin immunoprecipitation (ChIP) protocols were developed based on two techniques, CAPTURE (195) and Click-ChIP (198). Each ChIP technique was refined to ensure optimal pull down of both DNA and protein. ChIP DNA sequencing (ChIP-seq) was used for DNA pull down and ChIP mass spectrometry for protein pull down.

2.5.1 Click ChIP-seq

2.5.1.1 HDAC inhibitor induction and chromatin cross linking

4×10^7 native K562 cells were simultaneously treated with either clickable SAHA, SAHA or vehicle to achieve a concentration of 10 μ M and incubated at 37°C for 4 hours. These cells

were then harvested by centrifugation at 300g. Aliquots were collected and prepared for RNA sequencing (section 2.7.1) and RT-qPCR to confirm HDAC inhibitor induction (section 2.4.1.3). The remaining culture was cross-linked with 1% methanol free formaldehyde. Following thrice washing with cold PBS and PMSF, the cross-linked cells were snap frozen and stored at -80°C.

2.5.1.2 Chromatin isolation and fragmentation

After thawing on ice, the fixed cells were lysed using 1ml of cell lysis buffer (10mM Tris-HCl, 1mM EDTA, 0.1% sodium deoxycholate, 0.1% SDS, 1% Triton X-100 (pH 8.0)) and rotated for 15 minutes at 4°C. The lysate was centrifuged at 2,300g for 5 minutes at 4°C to isolate the nuclei. The nuclei were resuspended in 1.7ml of nuclear lysis buffer (0.5% SDS; 10mM EDTA; 50mM Tris-HCl, pH 8.0) and subjected to sonication for 22 cycles, 30 seconds ON /30 seconds OFF, using Bioruptor® Pico (Diagenode) to achieve chromatin fragments of approximately 500 bp, confirmed by agarose gel electrophoresis. The chromatin fragments were centrifuged at 13,000 rpm for 10 minutes. Samples for input and sonication control were collected at this point.

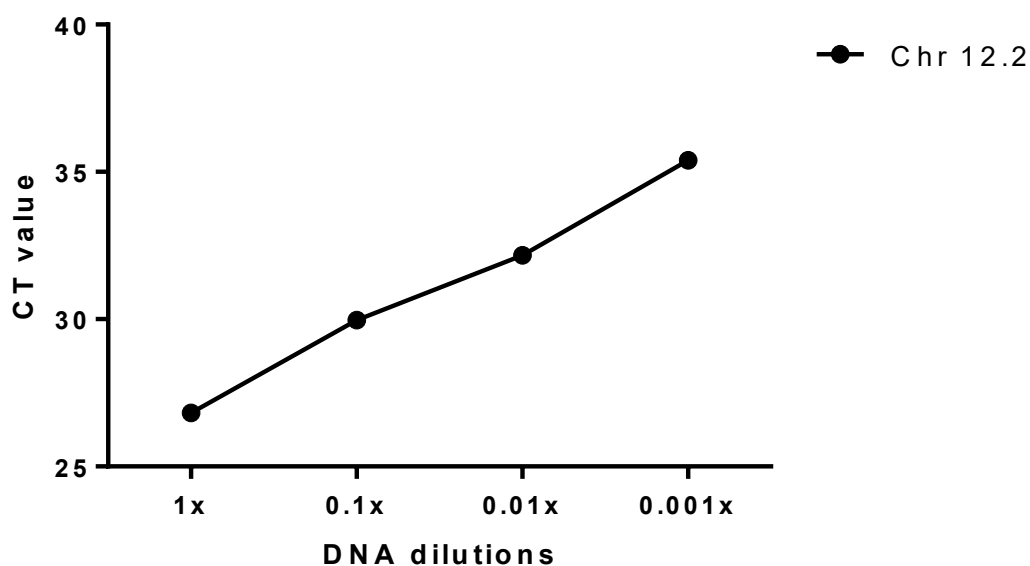
2.5.1.3 Click reaction and chemical precipitation

The sheared chromatin was added to freshly made azide-biotin mix (50µM biotin-azide, 5mM sodium L-ascorbate, 4mM copper (II) sulphate) and rotated for 1 hour at room temperature to enable the click reaction. Washed magnetic Streptavidin Dynabeads® (Thermo-Fisher) were added to chromatin/clicked SAHA mixture and rotated at 4°C overnight, to allow the biotin-streptavidin coupling to occur (**Figure 1.9C**). The beads then

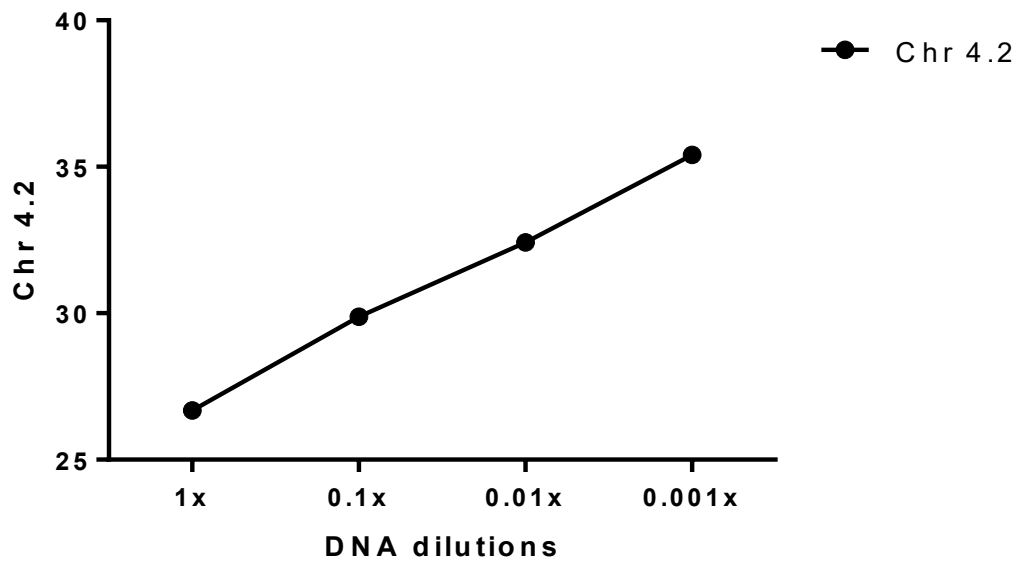
underwent sequential washes, each performed in duplicate. The first with 2% SDS, after rotating for 5 minutes at 4°C. The sample was placed on a magnet stand and supernatant was removed. This was repeated with RIPA buffer with 500mM sodium chloride (10mM Tris-HCl, 1mM EDTA, 0.1% sodium deoxycholate, 0.1% SDS, 1% Triton X-100, 0.5M sodium chloride, pH 8.0), lithium chloride buffer (250mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1mM EDTA, 10mM Tris-HCl, pH 8.0) and, finally, with TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0). Between these final two washes the microcentrifugation tubes were replaced to ensure that chromatin that had adhered to the tube was not eluted. The precipitated DNA was eluted using 150µl of freshly prepared elution buffer (0.1M Sodium bicarbonate and 1% SDS). Samples were shaken overnight at 65°C, then placed on a magnet stand and eluate collected in fresh microcentrifugation tubes. Cross-linking was reversed by adding sodium chloride to eluate, to achieve a final concentration of 200mM. Samples were incubated at 65°C overnight, 1µl of 10 mg/ml RNase A was added and samples were then incubated at 45°C for 45 minutes. Finally, the samples were deproteinised using 2µl of 0.5M EDTA (pH 8.0), 4µl of 1M Tris (pH 6.8), and 1µl of 20 mg/ml proteinase K (Thermo-Fisher Scientific) and incubated at 45°C for 90 minutes. DNA was purified with Agencourt AMPure XP Beads (Beckman Coulter) in accordance with manufacturer's instructions. Aliquots were taken for DNA quantification by Qubit High Sensitivity DNA kit, sonication control by agarose gel electrophoresis, RT-qPCR to confirm pull down, as described in section 2.5.1.5, and the remaining sample was prepared for DNA sequencing.

2.5.1.4 Primer design and testing

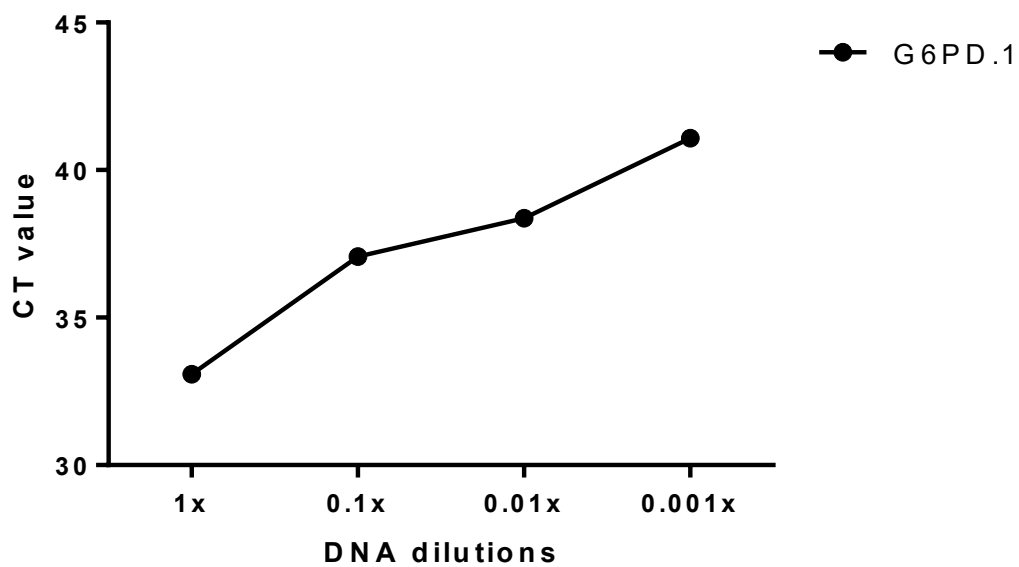
Primer sets were designed using Primer3web (<http://primer3.ut.ee/>). At least four pairs were produced for each amplicon. The main criteria were that each were between 18-23mer in length with melting temperatures (T^m) between 57-63°C and GC% of 30-70%. These primers were assessed for specificity, by reviewing melting curve, efficiency and linearity, with standard curve, and limits of detection and quantification. Specifically, linearity was tested by performing serial dilutions of random DNA. The resulting threshold cycle (CT) values, after performing qPCR, were plotted and linear regression (R^2) determined. R^2 of greater than 0.9 and less than 1.0 indicated linear amplification by the primer pair (**Figure 2.9 (A-E)**).



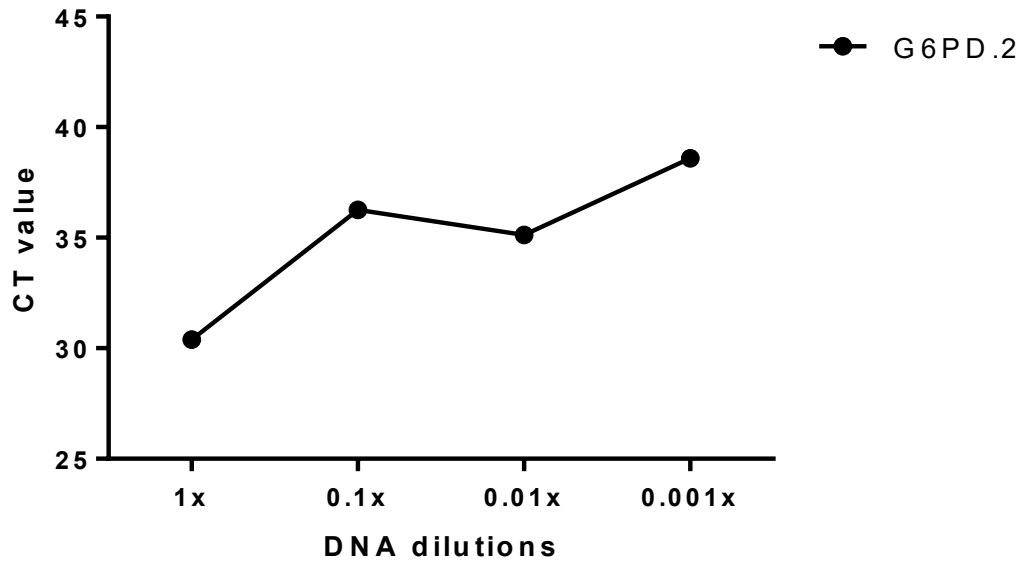
(A) RT qPCR to assess linearity of primer - chromosome 12.2 (negative control).
 $R^2 = 0.9949$



(B) RT qPCR to assess linearity of primer - chromosome 4.2 (negative control).
 $R^2 = 0.9982$

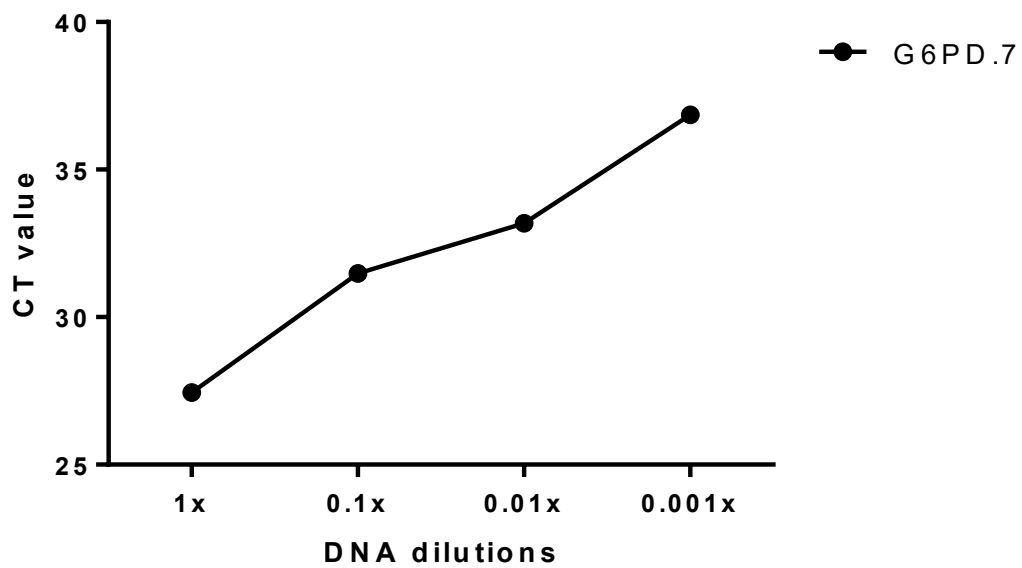


(C) RT qPCR to assess linearity of primer – *G6PD* 1.
 $R^2 = 0.9629$



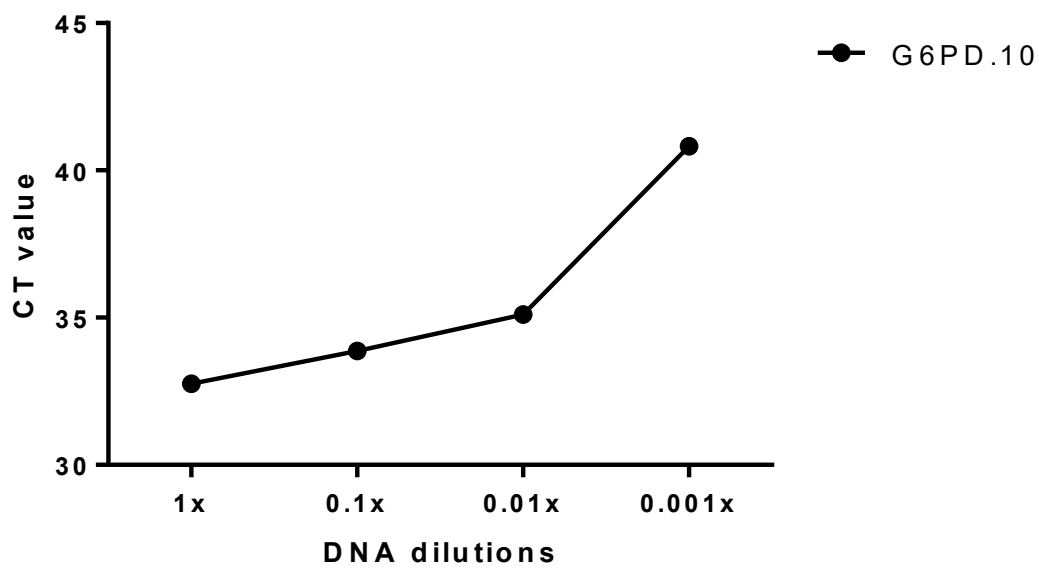
(D) RT qPCR to assess linearity of primer – *G6PD.2*.

$R^2 = 0.7717$



(D) RT qPCR to assess linearity of primer – *G6PD.7*.

$R^2 = 0.9792$



(E) RT qPCR to assess linearity of primer – *G6PD.10*.

$R^2 = 0.8379$

Figure 2.9 (A-E) Primer optimisation testing using serial dilutions of random DNA. n=3.

The primers that best fulfilled these criteria was selected for each amplicon (**Table 2.7**)

(212).

Target gene	Forward primer	Reverse primer	Annealing Temperature
G6PD.1	GGGAGCGGCGGACTGTGAA C	GGGGGCGGGGCTTGTGTTTT	60°C
G6PD.5	TGTTGTGCTTGAGAACCGAG C	TTGCCAAGCTGGGTGACCC	60°C
G6PD.7	CAGCCGTTCAACAAGGAGTGA	CTGGTGGAAGAAAGGCTCGT	60°C
TPI	GCCGGAGCTCACAGGTCT	GCACTGTTCCGACGTTCCC	60°C
Chromosome 4.2	CACATCGAGCGACTGACCTT	GCCTGGAGAGACAAAGCCAA	60°C

Table 2.7 List of ChIP probes

2.5.1.5 ChIP qPCR

Following DNA clean up, using Agencourt AMPure XP beads, and quantification, quantitative PCR was performed with Select Master Mix (Applied Biosystems). The reaction was run using the StepOnePlus™ Real-Time PCR System with the primers, designed and tested in

Table 2.7.

The following protocol was used:

Holding Stage – initial denaturation (dissociate the DNA strands)

1. 50°C for 2 minutes
2. 95°C for 2 minutes

Cycling Stage (45 cycles) - denaturation and annealing

1. 95°C for 3 seconds
2. 60°C for 30 seconds

Melt Curve Stage – records the change in fluorescence as the two DNA with incorporated dye molecules dissociates.

1. 95°C for 15 seconds
2. 60°C for 1 minute
3. 95°C for 15 seconds

DNA enrichment was calculated using the $\Delta\Delta\text{CT}$ method, comparing to input DNA.

2.5.1.6 Preparation of samples for RNA seq

Following quantification of RNA by Nanodrop™ Lite spectrophotometer and assessment of the RNA integrity using Agilent RNA 6000 Pico Kit and 2100 Bioanalyzer, targeting RNA integrity number (RIN) >7. Samples were prepared for RNA sequencing using the NEBNext® rRNA depletion kit (New England BioLabs), following manufacturer's instructions. After library preparation, the samples were sequenced in the Imperial BRC Genomics Facility on an Illumina HiSeq 4000 platform to obtain 75bp paired-end reads.

2.5.1.7 Preparation of samples for ChIP-seq

The enriched ChIP DNA samples were indexed with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following the manufacturer's protocol. After library preparation, the samples were purified using Agencourt AMPure XP Beads. Library DNA was quantified with the Qubit High Sensitivity DNA kit and library quality and size were assessed using the Bioanalyzer High Sensitivity DNA kit. Libraries were then diluted to a minimum of 2nM and

were sequenced in the Imperial BRC Genomics Facility on the Illumina HiSeq 2500 platform to obtain 50bp single-end reads.

2.5.2 Click ChIP-MS

The Click ChIP-MS protocol was adapted for CAPTURE ChIP-proteomics.

2.5.2.1 HDAC inhibitor induction and chromatin cross linking

The protocol detailed in section 2.5.1.1 was used; this involved treating native K562 cells with either clickable SAHA, SAHA or methanol and incubating for four hours. Aliquots were taken to confirm HDAC inhibitor induction. Unlike the Click ChIP-seq samples, these samples were cross linked with 2% methanol free formaldehyde.

2.5.2.2 Chromatin isolation and fragmentation

After thawing on ice, the fixed cells were lysed using 10ml of cell lysis buffer (25mM Tris-HCl, 85mM potassium chloride, 0.1% Triton X-100, pH 7.4) and freshly added dithiothreitol 1mM (Sigma-Aldrich) and 1:200 protease inhibitor cocktail. Samples were rotated for 15 minutes at 4°C. The lysate was centrifuged at 2,300g for 5 minutes at 4°C to isolate the nuclei. The nuclei were resuspended in 5ml of nuclear lysis buffer (50 mM Tris-HCl, 10 mM EDTA, 4% SDS, pH 7.4), with freshly added dithiothreitol 1mM and 1:200 protease inhibitor cocktail. The samples were incubated at room temperature for 10 minutes, then 15ml of 8M urea lysis buffer was added to nuclei suspension and centrifuged at 16,100g for 25 minutes. This wash was repeated for a total of three nuclear washes. Pelleted chromatin was then washed twice with 5ml cell lysis buffer. Prior to sonication the nuclei were suspended in IP

binding buffer (20mM Tris-HCl, 1mM EDTA, 0.1% NP-40, 10% glycerol, pH 7.5). Samples were subjected to sonication for 22 cycles, 30 seconds ON /30 seconds OFF, using Bioruptor® Pico (Diagenode) to achieve chromatin fragments of approximately 500bp, confirmed by agarose gel electrophoresis. The chromatin fragments were centrifuged at 13,000 rpm for 10 minutes. Samples for input and sonication control were collected at this point. Sodium chloride was added to the sheared chromatin to a final concentration of 150mM.

2.5.2.3 Click reaction and chemical precipitation

The sheared chromatin was added to freshly made azide-biotin mix (50µM biotin-azide, 5mM sodium L-ascorbate, 4mM copper (II) sulphate) and rotated for 1 hour at room temperature to facilitate the click reaction. Washed streptavidin agarose resin (Thermo-Fisher) was added to chromatin-azide-biotin mix and rotated overnight at 4°C. The resin was collected by centrifugation and washed five times in detergent free IP binding buffer (20mM Tris-HCl, 1mM EDTA, 150mM NaCl, 10% glycerol, pH 7.5) with 1:200 of protease inhibitor.

2.5.2.4 In-Solution Digestion and Peptide Isolation

The conjugated beads were resuspended in 0.5M Tris buffer. Samples were reduced with Tris (2-carboxyethyl) phosphine (TCEP), to a final concentration of 20mM, to break disulphide bonds in proteins and peptides. Following this, iodoacetamide was added to samples and incubated in the dark at room temperature. This covalently binds the thiol group of cysteine to prevent reformation of these disulphide bridges. After 30 minutes, the reaction was quenched with dithiothreitol. The beads suspension was then digested with

trypsin overnight at room temperature. The digested peptides were purified using cellulose acetate filter spin cup (Thermo-Fisher Scientific), then boiled with 3M sodium chloride to reverse cross-linking. The peptides were dried and reconstituted in 70% acetonitrile (Fisher Scientific Ltd) and desalted using Peptide Desalting Spin Columns (Pierce™). The samples were then subjected to Liquid Chromatography coupled Mass Spectrometry.

2.5.3 CAPTURE-seq

2.5.3.1 HDAC inhibitor induction and chromatin cross-linking

4×10^7 of each of the four CAPTURE cell lines were simultaneously treated with either SAHA or vehicle to achieve a concentration of $10 \mu\text{M}$ and incubated at 37°C for 4 hours. These cells were then harvested by centrifugation at 300g. Aliquots were taken and prepared for RNA sequencing (section 2.7.1) and RT-qPCR to confirm HDAC inhibitor induction (section 2.4.1.3). The remaining culture was cross-linked with 1% methanol free formaldehyde. Following thrice washing with cold PBS and PMSF, Sigma-Aldrich. The cross-linked cells were snap frozen and stored at -80°C .

2.5.3.2 Chromatin isolation and fragmentation

After thawing on ice, the fixed cells were lysed using 1ml of cell lysis buffer (10mM Tris-HCl, 1mM EDTA, 0.1% sodium deoxycholate, 0.1% SDS, 1% Triton X-100 (pH 8.0)) and rotated for 15 minutes at 4°C . The lysate was centrifuged at 2,300g for 5 minutes at 4°C to isolate the nuclei. The nuclei were resuspended in 1.7ml of nuclear lysis buffer (0.5% SDS; 10mM EDTA; 50mM Tris-HCl (pH 8.0)). Samples were subjected to sonication for 10 cycles, 5 seconds

ON/90 seconds OFF, using Bioruptor® Pico (Diagenode) to achieve chromatin fragments of approximately 2Kb, confirmed by agarose gel electrophoresis. The chromatin fragments were centrifuged at 13,000 rpm for 10 minutes. Samples for input and sonication control were collected at this point.

2.5.3.3 Immunoprecipitation

Sodium chloride was added to the sheared chromatin to achieve a final concentration of 300mM. Washed magnetic Streptavidin Dynabeads® were added to fragmented chromatin and rotated at 4°C overnight to allow the dCas9-biotin-streptavidin coupling to occur. The following day, the beads underwent sequential washes, each performed in duplicate. The first with 2% SDS, after rotating for 5 minutes at 4°C. The samples were placed on a magnet stand and supernatant removed. This was repeated with RIPA buffer with 500mM sodium chloride (10mM Tris-HCl, 1mM EDTA, 0.1% sodium deoxycholate, 0.1% SDS, 1% Triton X-100, 0.5M sodium chloride, pH 8.0), lithium chloride buffer (250mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1mM EDTA, 10mM Tris-HCl, pH 8.0) and, finally, with TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0). Between these final two washes the microcentrifugation tubes were replaced to ensure that chromatin that had adhered to the tube was not eluted. The immunoprecipitated DNA was eluted using 150µl of freshly prepared elution buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl, pH 8.0). Samples were shaken overnight at 65°C, then placed on a magnet stand and eluate collected in fresh microcentrifugation tubes. Reverse cross-linking was performed by adding sodium chloride to eluate, to achieve a final concentration of 200mM. Samples were incubated at 65°C overnight, 1µl of 10 mg/ml RNase A was added and samples were then incubated at 37°C for

45 minutes. Finally, deproteinization of samples was achieved using 2 µl of 0.5M EDTA (pH 8.0), 4 µl of 1M Tris (pH 6.8), and 1µl of 20 mg/ml proteinase K, incubated at 45°C for 90 minutes. DNA was purified with Agencourt AMPure XP Beads in accordance with manufacturer's instructions. Aliquots were taken for DNA quantification by Qubit, sonication control by agarose gel electrophoresis, RT-qPCR to confirm pull down, as described in section 2.5.1.5, and the remaining sample was prepared for DNA sequencing.

2.5.3.4 Primer design and testing

Detailed in section 2.6.1.4.

2.5.3.5 ChIP qPCR

Detailed in section 2.6.1.5.

2.5.4 CAPTURE-MS

The CAPTURE-seq protocol was adapted for ChIP-proteomics.

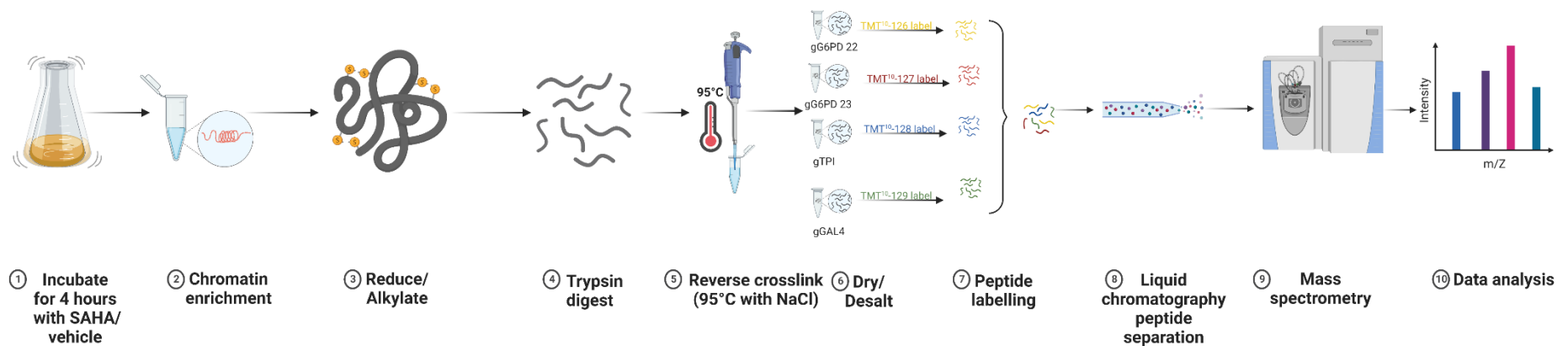


Figure 2.10 CAPTURE-MS workflow

Following incubation of modified K562 cells with either SAHA or vehicle. The chromatin was crosslinked with formaldehyde and the treated cells were subjected to a series of washes to isolate chromatin. After sonication, the chromatin fragments were selectively precipitated on streptavidin beads. Samples were then alkylated with iodoacetamide after cystine reduction to prevents the formation of disulphide bonds. The peptides were recovered by digestion with trypsin followed by reversal of crosslinking. The samples were then labelled with a TMT mass tag labelling kit to enable multiplex relative quantitation liquid chromatography in tandem with mass spectrometry.

Image created in BioRender by author.

2.5.4.1 HDAC inhibitor induction and chromatin cross linking

The protocol outlined in section 2.5.3.1 was followed incubating four CAPTURE cell lines with SAHA or vehicle for 4 hours. Aliquots were taken to confirm HDAC inhibitor induction. Unlike the CAPTURE-seq samples, these samples were cross-linked with 2% methanol free formaldehyde.

2.5.4.2 Chromatin isolation and fragmentation

After thawing on ice, the fixed cells were lysed using 10ml of cell lysis buffer (25mM Tris-HCl, 85mM potassium chloride, 0.1% Triton X-100 (pH 7.4)) and freshly added 1mM dithiothreitol and 1:200 protease inhibitor cocktail. Samples were rotated for 15 minutes at 4°C. The lysate was centrifuged at 2,300g for 5 minutes at 4°C to isolate the nuclei. The nuclei were resuspended in 5ml of nuclear lysis buffer (50 mM Tris-HCl, 10 mM EDTA, 4% SDS (pH 7.4)), with freshly added dithiothreitol 1mM and 1:200 protease inhibitor cocktail. The samples were incubated at room temperature for 10 minutes. 15ml of 8M urea lysis buffer was added to nuclei suspension, then centrifuged at 16,100g for 25 minutes. This wash was repeated for a total of three nuclear washes. Pelleted chromatin was then washed twice with 5ml cell lysis buffer. Prior to sonication the nuclei were suspended in IP binding buffer (20mM Tris-HCl, 1mM EDTA, 0.1% NP-40, 10% glycerol, pH 7.5). Samples were subjected to sonication for 10 cycles, 5 seconds ON/90 seconds OFF, using Bioruptor® Pico to achieve chromatin fragments of approximately 2Kb, confirmed by agarose gel electrophoresis. The chromatin fragments were centrifuged at 13,000 rpm for 10 minutes.

Samples for input and sonication control were collected at this point. Sodium chloride was added to the sheared chromatin to a final concentration of 150mM.

2.5.4.3 Immunoprecipitation

Washed Streptavidin agarose resin was added to the fragmented chromatin and rotated at 4°C overnight, to allow the dCas9-biotin-streptavidin coupling to occur. The resin was collected by centrifugation and washed five times in detergent free IP binding buffer (20mM Tris-HCl, 1mM EDTA, 150mM NaCl, 10% glycerol (pH 7.5)) with 1:200 of protease inhibitor.

2.5.4.4 In-Solution Digestion and Peptide Isolation

The conjugated resin was resuspended in 0.5M Tris buffer. Samples were reduced with TCEP, to a final concentration of 20mM. Following this, iodoacetamide was added to samples and incubated in the dark at room temperature. After 30 minutes, the reaction was quenched with dithiothreitol. The beads suspension was then digested with trypsin overnight at room temperature. The digested peptides were purified using cellulose acetate filter spin cup; they were then boiled with 3M sodium chloride to reverse crosslinking. The peptides were dried and reconstituted in 70% acetonitrile and desalted using Peptide Desalting Spin Columns. Then labelled with multiplexed isobaric tags using the TMT 10plex Isobaric Mass Tag Labelling Kit (Thermo Scientific) prior to analysis by Liquid Chromatography coupled Mass Spectrometry.

2.6 STATISTICAL ANALYSIS

Statistical analyses were performed using GraphPad Prism v9.3.1. Following QQ plots to ensure normality, two-way ANOVA were performed, with a p value significance threshold of <0.05. All graphs illustrate mean and standard deviation (SD), unless otherwise specified.

2.7 BIOINFORMATIC ANALYSIS

Bioinformatic analysis of RNA-seq was performed by Dr Nikolaos Trasanidis, Imperial College London. Computational workflows used are outlined below.

2.7.1 RNA-seq

After size selection, 75bp paired-end reads per RNA sequencing experiment were demultiplexed and quality checked by the Imperial BRC Genomics Facility. The reads were aligned against the GRCh38 human reference genome using “STAR 2.5.3a” (213). Raw counts were obtained for mapped individual genes with the “Rsubread” Bioconductor package (214). Identification of differentially expressed genes between different biological conditions was determined using the “DESeq2” Bioconductor package (215). Biological interpretation of differentially expressed genes using functional enrichment and pathway analysis was performed using two online tools, Reactome (<https://reactome.org/>) and EnrichR (<https://maayanlab.cloud/Enrichr/>) (216) (**Appendix C Figure S1**).

2.7.2 Proteomic analysis

Bioinformatic analysis of proteomic datasets (**Appendix C Figure S2**) was performed by Dr Harry Whitwell, Imperial College London, except for general proteomics performed on

CAPTURE samples. This was processed by Proteomics Facility at the Centre of Excellence for Mass Spectrometry, King's College London.

Liquid chromatograph-mass spectrometry (LC-MS) analysis was performed using bottom-up proteomic approach, in which proteins are digested to peptides, coupled with data-dependent acquisition mass spectrometry. The data was initially searched and processed through commercial software to identify preliminary targets and perform quantitation. However, histone post-translational modification analysis required manual validation and annotation using in-house scripts to interrogate the annotated mass spectrometry data and unidentified peaks. Retention time and masses (m/z) of target peptides were used to quantify modified peptides using PeakPanther – R package for performing accurate peak integration of LC-MS data (217).

For histone PTMs (to compare the quantitation of modified peptides in progenesis versus PeakPanther) a histone dilution series was analysed by LC-MS and processed using the developed script and Progenesis. Peptides were scored by their confidence (identification) and linearity (quantitative suitability) to derive a list of target histone modifications.

Contaminants were identified using CRAPome online database (<https://reprint-apms.org/?q=reprint-home>) and excluded (218). Biological interpretation of differentially expressed peptides using functional enrichment and pathway analysis was performed using two online tools, Reactome (<https://reactome.org/>) and PantherDB (pantherdb.org).

2.8 OTHER SOFTWARE

Flow cytometry analysis was performed using FlowJo X.07 software. Cloning strategies were planned using SnapGene 3.2.1 software.

2.8.1 Online software

BioRender (<https://app.biorender.com/>) was used to generate illustrations.

STRING networks were created at (<https://string-db.org/>) and modified using the complimentary program Cytoscape 3.7.2 (219).

Chapter Three

Systems Design and Development

3 SYSTEMS DESIGN AND DEVELOPMENT

3.1 INTRODUCTION

A broad and unbiased approach was needed to decipher the transcriptional and epigenetic mechanism by which *G6PD* activity is selectively induced by HDAC inhibitors. I developed two distinct yet complementary chemo-purification techniques to isolate the chromatin of interest.

This chapter describes the development of the two techniques used to answer the central questions of this project. The evolution from Chem ChIP to Click ChIP was mainly due to difficulty identifying optimal conditions for *G6PD* induction by the tagged SAHA. There was also the progression from enChIP to CAPTURE for locus specific pull down due to difficulty transducing cells with required plasmids. Attempts to reduce transduction payload of original 2-plasmid enChIP system were unsuccessful. So, I progressed to a 3-plasmid CAPTURE protocol. The CAPTURE protocol has several other advantages over its predecessor. However, I was unable to create a stable CAPTURE cell line. So, the cell lines created were transiently transduced with the required plasmids and therefore required ongoing antibiotic exposure to ensure ongoing expression. Finally, as well as evolving the ChIP techniques I also had to change the cellular model from GM12878 to K562 because of difficulty achieving optimal selective induction of the GM12878 cell line.

As the data obtained from these two distinct modalities will be used to answer the same central question there was a clear imperative to ensure uniformity of results for validation.

3.2 CELLULAR MODEL

The selection of an appropriate cellular model was essential to the success of the project.

There were numerous factors that needed be considered when selecting a cell line. In particular, the cell line should be:

1. Of human origin, ideally with a normal karyotype.
2. Similar phenotype to the erythroid lineage that is being investigated.
3. Amenable to manipulation, i.e. receptive to transduction to facilitate introduction of the cellular machinery required for dCas9 pull down.
4. The model would also need to be robust enough to tolerate the toxic effects of SAHA.
5. It would need to be scalable.
6. Finally, the cell line should be a tier 1 ENCODE cell line to allow easy access to high-quality epigenetic data.

With respect to the last point, it was crucial that the selected cell line was one for which high quality epigenetic data was readily available. The ENCODE project is an ongoing international collaboration launched by the National Human Genome Research Institute following the completion of the Human Genome Project (which aimed to identify all functional elements in the genome). This consortium of research groups developed a free-to-access database that includes epigenetic information (220). To aid comparison and integration of these data generated from a variety of technologies and platforms, the

consortium assigned cell lines commonly used, easily cultured, and offered the greatest diversity, tier 1 status. In this project, tier 1 cell lines are used in all experiments, where possible, and therefore a diversity of datasets were available. The two haematopoietic cell lines are GM12878 and K562 (the former was selected for this project as it has a normal karyotype). GM12878 is a B cell lymphoblastoid cell line derived from the blood of a female donor of European Caucasian ancestry.

Previous work in the Karadimitris lab established a stable 293T enChIP cell line by cloning pLVhUbC-dCas9T2aGFP which expresses 3xFLAG nuclease deficient Cas9 with GFP fluorochrome and pU6-sg4-NT1 which expresses an mCherry cassette that encodes a G6PD-specific gRNA, g4. The expression of these constructs was confirmed by immunoblotting and flow cytometry, respectively. Preliminary optimisation experiments to establish ideal conditions for HDAC inhibitor induction, sonication and chromatin pull down used this clone.

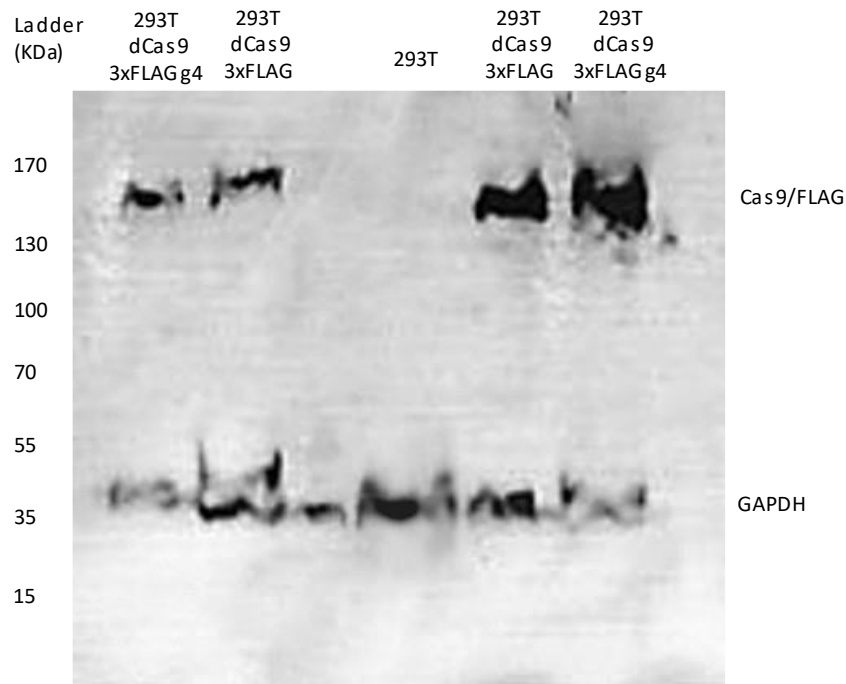


Figure 3.1 Western blot showing expression profile of 293T, 293T dCas9 and 293T dCas9 g4

This blot confirms that 293T dCas9 and 293T dCas9 g4 cell lines express Cas9 (left) and FLAG proteins (right). Immunoblot performed on a 4-12%Tris-Glycine pre-cast gel with PageRuler Prestained Protein and PageRuler Plus Prestained Protein Ladders.

This 293T cell line was robust, easy to transduce and tolerant of a wide range of SAHA doses. However, this clone had limitations, as a tier 3 ENCODE cell line there are limited datasets available. The other important limitation was the impact of dCas9 binding on *G6PD* expression, see section 2.2.1.

Therefore, when considering factors critical to the establishment of an enChIP clone, GM12878 was my first choice as it is the only hematopoietic tier 1 ENCODE cell line with a normal karyotype. I also concluded that gRNAs would need to be redesigned to ensure non-interference with *G6PD* transcription.

3.2.1 Establishment of clones

3.2.1.1 Modification of enChIP vector

Initial attempts to establish a new enChIP cell line with GM12878 proved unsuccessful.

Following the redesign of the *G6PD* gRNA and integration into the pU6-sgX-NT1 vector, my focus turned to transduction of both this plasmid and the pLVhUbC-dCas9T2aGFP into the GM12878 cell line.

Viral transduction of the pLVhUbC-dCas9T2aGFP vector yield was <0.1%. Unsurprisingly, attempts to expand such a small culture were unsuccessful.

To address this problem, I first attempted to reduce the transduction burden on the target cell by limiting insert redundancy, particularly of the pLVhUbC-dCas9T2aGFP construct (the larger of the two plasmids). Lentiviral vectors have a large packaging capacity of approximately 8Kb (221). However, as the Cas9 gene alone is more than 4Kb in size, with the vector almost 9Kb, exceeding the lentiviral packaging capacity made transduction challenging. The plasmid was redesigned using both restriction ligation and Gibson Assembly to reduce the plasmid size to approximately 12Kb and the transduction payload to <8Kb (**Appendix B Figure S1**). Despite these modifications transduction efficiency was persistently low. The transgene sequence is flanked by long terminal repeat (LTR) sequences, and it is these LTRs that facilitate integration of the transfer plasmid sequences into the host genome.

3.2.1.2 Gene transfer system

Next, to address the transduction problem I interrogated the gene transfer system. Initially, a 3rd generation system was employed. This requires a transfer vector and therefore does not require a HIV Tat protein. This decreases the probability of creating replication-competent lentivirus in the target cells, so is safer than predecessors. It also separates the packaging system into two plasmids: one encoding Rev and one encoding Gag and Pol, increasing the number of recombination events necessary to create a replication-competent virus. This complexity, which necessitates a three-plasmid helper system, is more cumbersome and leads to a lower viral titre due to the higher number of plasmids required for transduction. Conversion to a 2nd generation system would theoretically increase virion number and potentially transduction efficiency. Unfortunately, this change to the packaging system resulted in only a marginal increase in transduction efficiency, not enough to impact clonal expansion.

3.2.1.3 Transduction enhancement

Despite the measures outlined in section 3.2.1.2, I was unable to develop enChIP GM12878 cell lines. It is well documented that transduction of lymphocytes requires stimulation to exit the G₀ phase for transduction competence by VSVG- pseudotyped lentiviruses (222). With this information, I turned my attention to augmentation of the transduction process.

RetroNectin® is a well-established transduction enhancer (223). This recombinant fibronectin fragment binds virions to sequences in the heparin-binding domain and target cell integrins maintaining them in close proximity. This concept of co-localisation to promote

transduction was further explored with spinoculation. Spinoculation uses a centrifugal force to achieve the same goal of bringing target cell and viral particles into close vicinity to promote transduction. Once again, neither of these measures had a significant impact on transduction efficiency. As RetroNectin® could not be used in conjunction with any other transduction enhancing measures, I opted to abandon RetroNectin® but did integrate spinoculation to the protocol.

Next, I explored the option of proprietary enhancers, such as ViralPlus®, reported to suppress the cellular anti-viral state by mimicking the activity of viral virulence gene products, and thus significantly increasing lentivirus transducing efficiency by upto 10-fold (compared with traditional enhancers such as polybrene). Despite trialling a range of concentrations, 1:100 to 1:50, with or without polybrene, this also had no significant impact on transduction.

Finally, given that stimulating B cells out of quiescence would theoretically improve transduction efficiency. I speculated that the introduction of interleukin-6 to induce cells to enter G₁ phase of the cell cycle could be beneficial. Unfortunately, neither alone nor in combination with other measures, this improved viral transduction of the pLVhUbC-dCas9T2aGFP vector into GM12878 cells.

I now had only one choice, to select an alternative cell line which is more amenable to transduction.

3.3 OPTIMISATION OF TARGET GENE ENRICHMENT

3.3.1 Chromatin pull down

Fujita *et al.* described a technique which allowed examination of a specific chromatin region in an agnostic manner. EnChIP enables the purification of a specific region of chromatin using a locus specific guide RNA. This protocol utilises a gRNA that anneal the genomic region of interest directing the deactivated Cas9 which tethers the target gene and the associated transcriptional machinery. The chromatin of interest can be extracted using antibody directed to either 3xFLAG tag or Cas9. The isolated chromatin from the specific genomic region of interest retains molecular components and interactions for downstream biochemical analysis (202). The quality of downstream analysis is dependent on antibody-antigen interaction with high sensitivity and specificity to limit non-specific binding and that the affinity of interaction is sufficient to enable stringent washes and minimal losses.

There are three antibody-antigen options in this protocol:

1. Antibody directed to 3x FLAG tag
2. Monoclonal anti-Cas9 antibody
3. Polyclonal anti-Cas9 antibody

To evaluate the most effective antibody-antigen interaction for this purpose pull down experiments were performed to compare the three approaches. 293T dCas9 g4 cells were used in ChIP experiments employing the enChIP protocol. These optimisation experiments were performed in duplicate, and the results are shown in **Figure 3.2**.

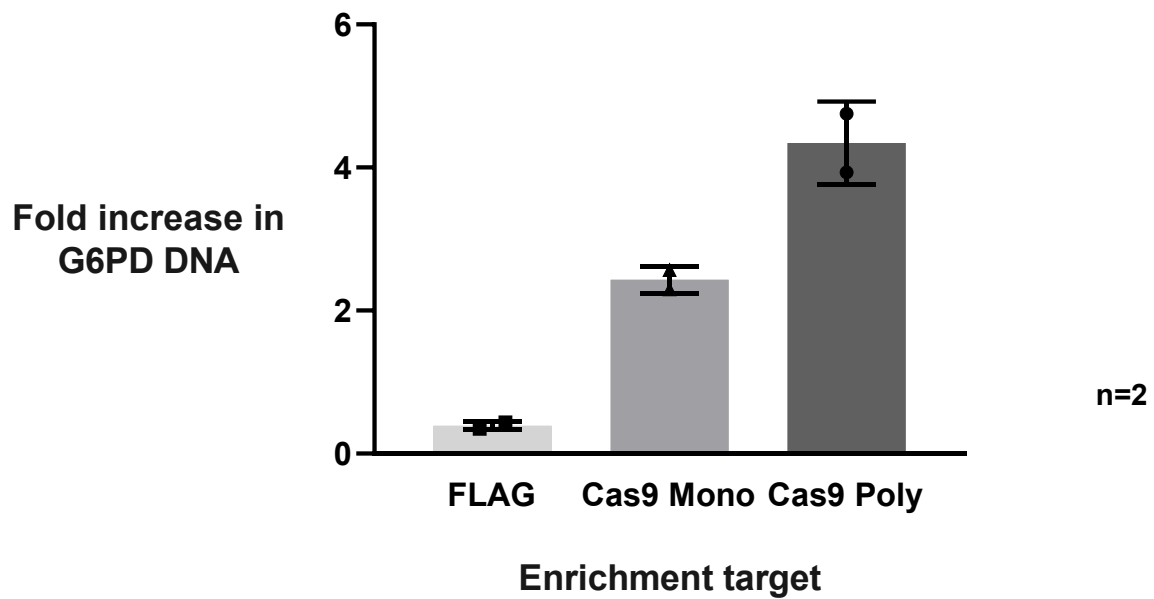


Figure 3.2 Fold change in *G6PD* DNA using different ChIP pull down approaches.

Chart represents fold enrichment of *G6PD* DNA, normalised to control, using either FLAG or Cas9 antibodies. The polyclonal Cas9 antibody was 4 times more efficient than FLAG mediated pull down, and 1.5 times more effective than the monoclonal Cas9 antibody. Each experiment was performed in duplicate and filled shape represent individual values. Error bars depict mean and standard deviation.

These data showed a greater avidity of anti-Cas9 antibody over anti-FLAG antibody with >4 times greater enrichment. Furthermore, this investigation demonstrated the polyclonal anti-Cas9 antibody showed almost 2-fold greater pull down efficacy than the monoclonal variant.

3.3.2 EnChIP to CAPTURE

The transition from enChIP to CAPTURE protocol, for reasons described in section 3.2.1, was to facilitate successful plasmid transduction. This modification utilises an alternative pull down technique, rather than the classical antibody-antigen interaction. The CAPTURE protocol employs *in vivo* biotinylation of dCas9 which enables biotin-streptavidin pull down.

The biotin-streptavidin interaction offers several advantages in isolating chromatin-associated complexes over conventional antibody-based affinity purification.

First, the biotin-streptavidin interaction is extremely avid, $K_d = 10^{-14}$ mol/L (224), this equates to an affinity 1,000-fold greater than antibody-mediated interactions. Next, this approach allows for more stringent washes and, therefore, efficient and stable capture of protein-DNA complexes. This reduces co-purified protein contamination and better interpretation of results due to fewer off-target effects. Finally, as few naturally biotinylated proteins are located within the nucleus, streptavidin affinity purification is unlikely to result in significant unwanted binding, as is more commonly associated with antibody-based approaches.

3.3.2.1 CRISPR affinity purification in situ of regulatory elements (CAPTURE)

Liu *et al.* described a modification of the enChIP protocol (195). From a transduction perspective, this protocol replaces the 2-plasmid system, **Figure 1.10**, with a 3-plasmid system, **Figure 2.4**. Although the number of plasmids is increased, the plasmid sizes are significantly reduced, and the combined sequence payload remains within the lentiviral packaging capacity. The two CAPTURE plasmids, pEF1a-BirA-neo and pEF1a-FB-puro, lack viral long terminal repeats which ensure integration of cDNA of gene of interest to the host genome, and so can be only transiently transduced. I attempted to modify both vectors, removing redundancy while adding LTRs and substituting antibiotic resistance selectable markers with fluorochromes (**Figures 3.3 and 3.7**).

3.3.3 Modification of pEF plasmids

There were limitations to the pEF1a-FB-dCas9-puro and pEF1a-BirA-V5-neo plasmids:

1. Inability to utilise FACS sorting as no fluorescent selectable markers
2. Viral transduction was not possible as they lack LTRs
3. It was necessary to maintain selection with two antibiotics.

I attempted the redesign of the two pEF plasmids by introducing a fluorescent tag and LTRs flanking the proviral DNA, with appropriate promoters.

Cloning of pEF1a-BirA-V5-BFP-virus v7 plasmid began with simulation of the cloning process using SnapGene® program (**Figure 3.4**), illustrated below.

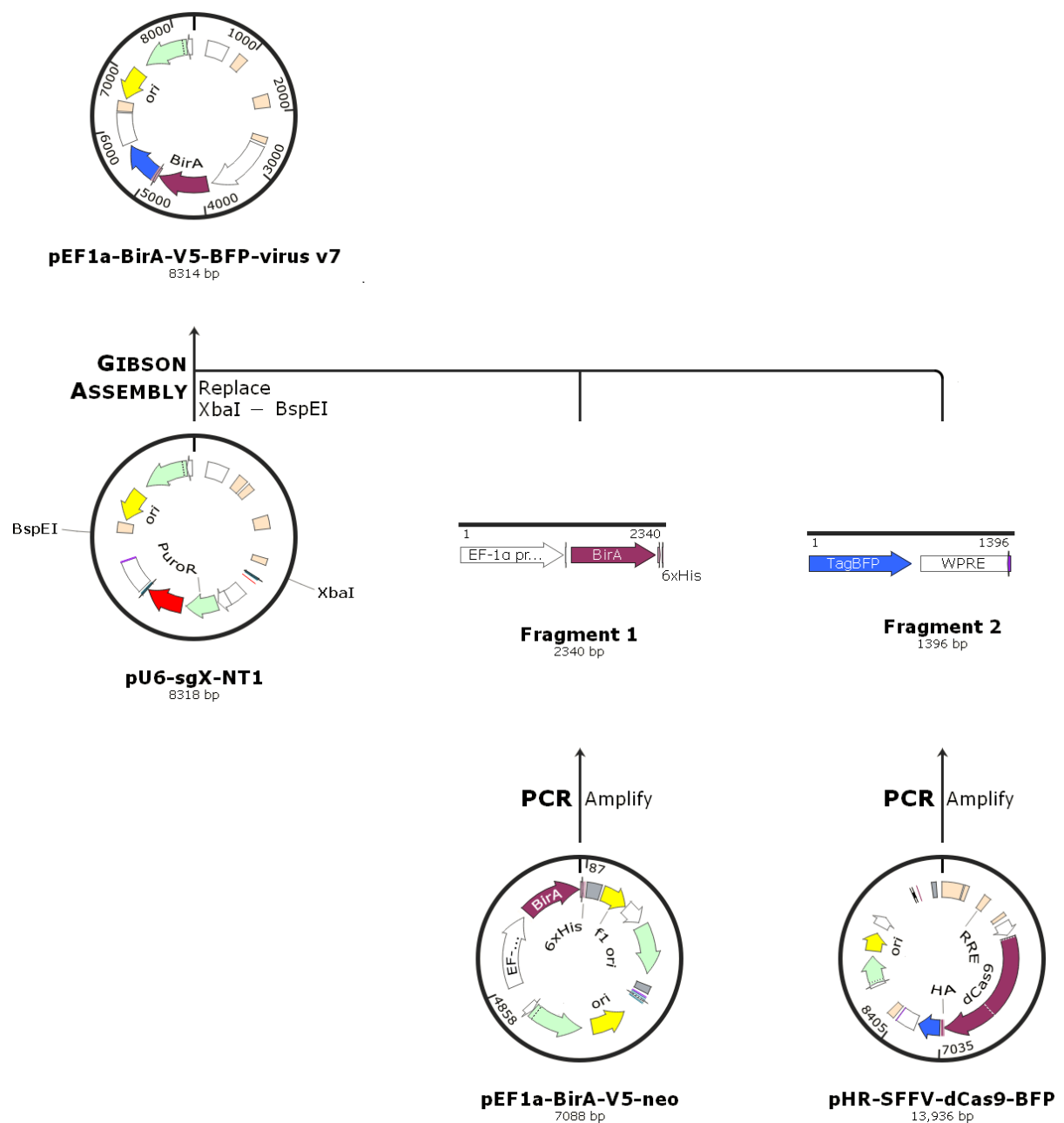


Figure 3.3 pEF1a-BirA-V5-neo → pEF1a-BirA-V5-BFP-virus v7 for viral transduction

The modifications to pEF1a-BirA-V5-neo enabled viral transduction and allow selection by flow cytometry. Fragments containing biotin ligase (BirA) and its promoter, and another with BFP fluorescent tag/WPRE sequence were amplified using two plasmid templates by PCR. These fragments were integrated to pU6-sgX-NT1 construct following linearised BspEI and XbaI using Gibson Assembly. The resultant construct contained flanking LFTs to allow integration to host genome following viral transduction. The plasmid also expresses BFP fluorescent tag to enable selection by flow cytometry.

pU6-sgX-NT1 plasmid was digested with XbaI and BspEI (Thermo Fisher Scientific) (**Figure 2.2**). The 4632bp fragment was then isolated by agarose gel electrophoresis and extracted

using Monarch[®] DNA Gel Extraction Kit. Two fragments: one encoding for EF-1a promoter, BirA ligase and polyhistidine tag, and the other Blue Fluorescent Protein (BFP) marker and Woodchuck Hepatitis Virus Posttranscriptional Response Element (WPRE) were amplified by PCR using PCRBIO HiFi Polymerase (PCR Biosystems Ltd.); templates and primers detailed in **table 3.1**. The fragments were quantified by Nanodrop™ Lite spectrophotometer and fragments were annealed using the Gibson Assembly process (NEB). Following incubation for 15 minutes at 50°C with the Gibson Assembly Master Mix, as per the manufacturer's instructions, they were transformed into NEB DH5α competent cells. Successful cloning was confirmed by diagnostic digest and agarose gel electrophoresis, then Sanger sequencing.

Name	Size (bp)	Function	Template	Primers
Vector		Backbone	pU6-sgX-NT1	
Fragment 1	2340	BirA ligase, Histidine tag and EF-1a promoter	pEF1a-BirA	Forward: tagtaccgggcccgttggtccggtg cccg
				Reverse: tcgctggaatggtgatggtgatgatga ccg
Fragment 2	1396	Blue fluorescent protein	pHR-SFFV-dCas9-BFP	Forward: atcaccattccagcgagctgattaagg agaaca
				Reverse: agagagaccagtagctcgaggtcg acggtatcgatgc

Table 3.1 Components of pEF1a-BirA-V5-BFP and associated primers

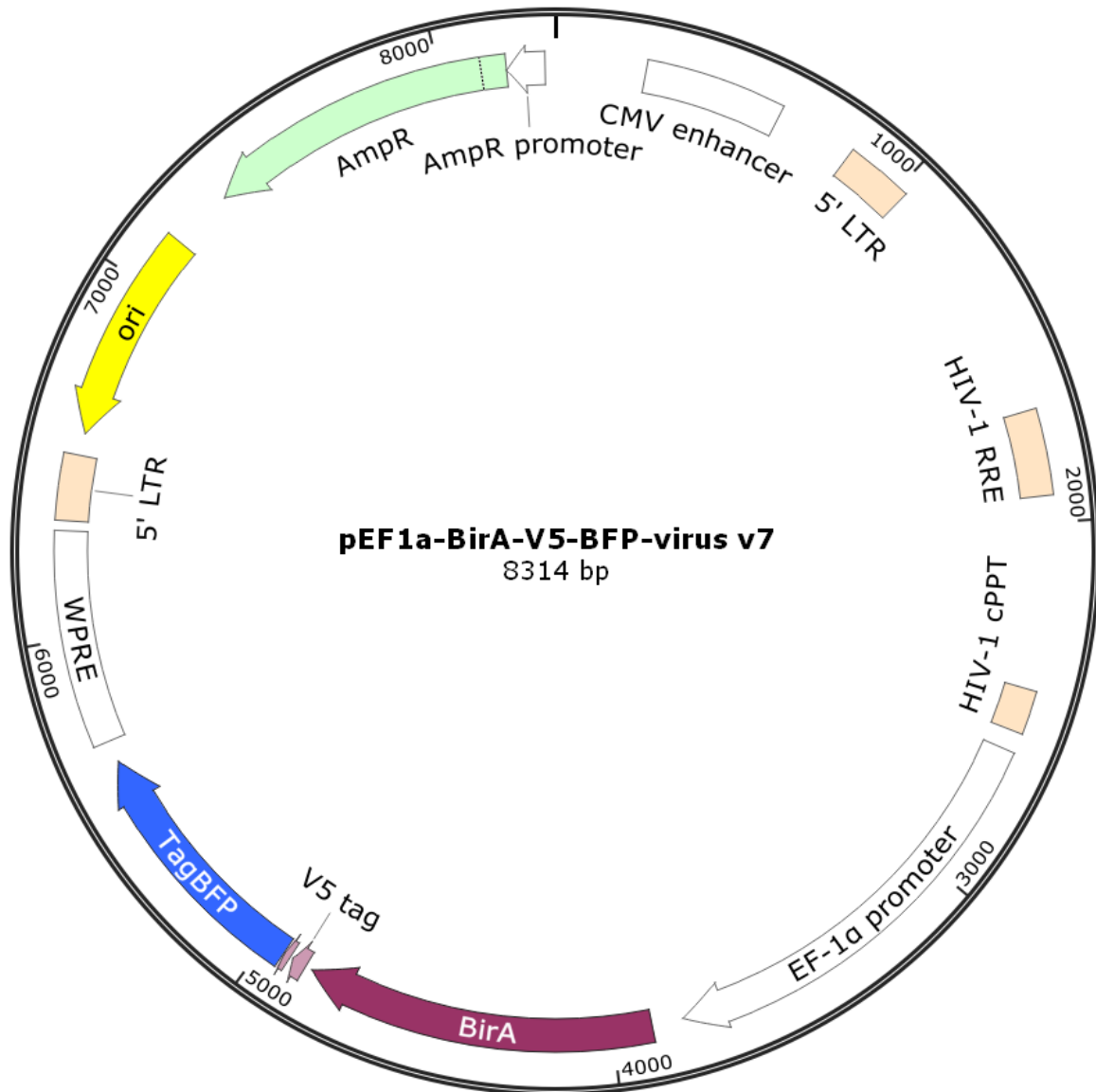


Figure 3.4 Modification of pEF1a-BirA-V5-neo plasmid to pEF1a-BirA-V5-BFP-virus v7

BirA ligase expression cassette driven by the EF-1a promoter, chimeric LTRs with a CMV promoter and a gene for the Blue Fluorescent Protein (blue). 5' LTRs flanks the transgene sequence to facilitates the integration of transfer plasmid sequences into host genome. The transgene sequence contains WPRE, V5 and histidine tags, BFP and biotin ligase. Each replicon also contains antibiotic resistance gene (AmpR), Central polypurine tract (cPPT) improves vector integration and transduction efficiency and an origin of replication (*ori*) which enables plasmid self-replication.

The construction of pEF1a-FB-dCas9-GFP-virus v4 was significantly more complicated than pEF1a-BirA-V5-BFP-virus v7, requiring a two-step approach. The first step was to integrate

the fluorescent tag and the second to add the LTRs and additional enhancers, **Figures 2.10** and **2.11**.

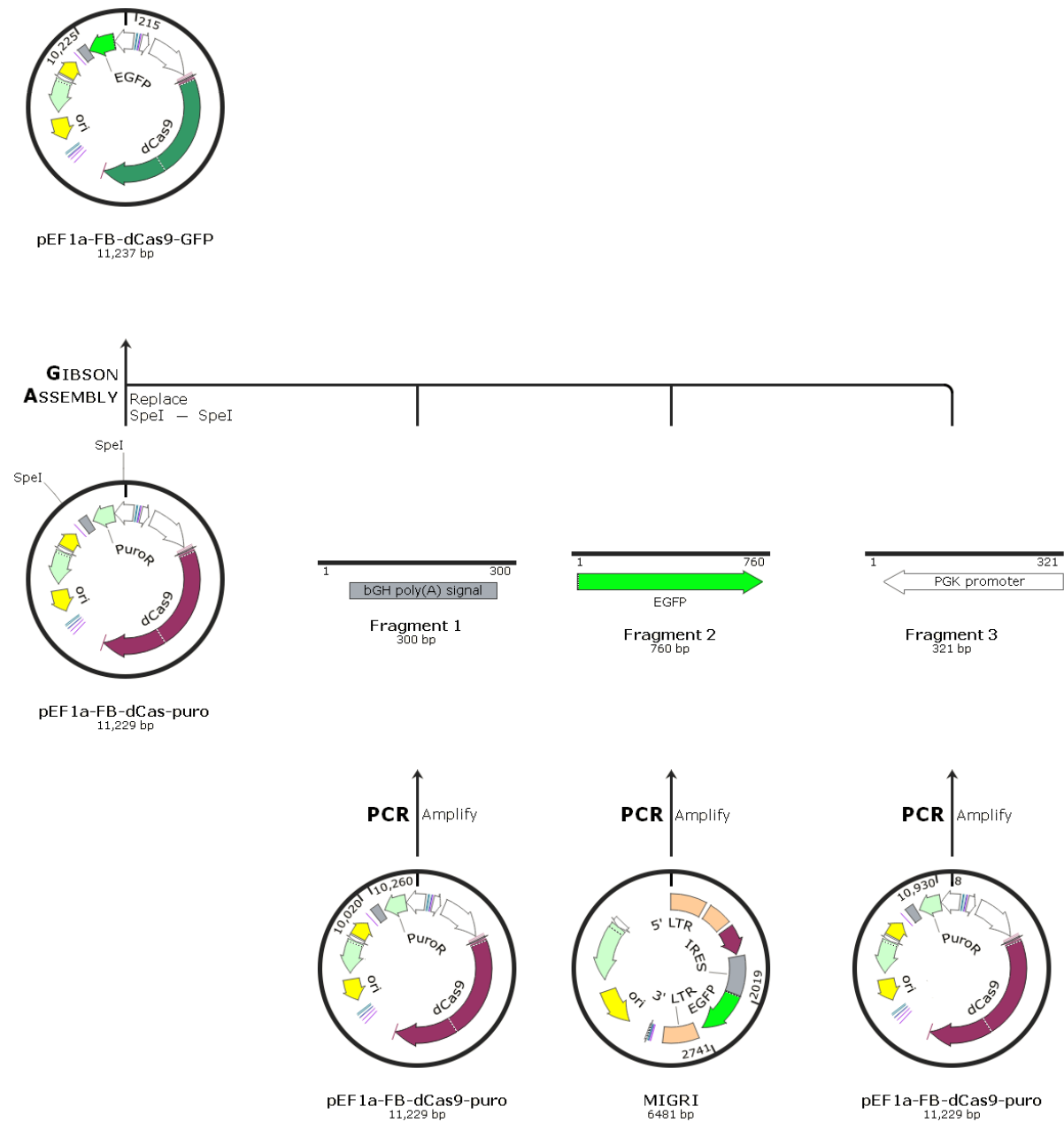


Figure 3.5 Step 1 of pEF1a-FB-dCas9-puro → pEF1a-FB-dCas9-GFP-virus v4 for viral transduction

Illustrating the initial steps in the construction of pEF1a-FB-dCas9-GFP-virus v4. This process using PCR to amplify three fragments:

1. polyadenylation signal – ensures nuclear export, translation and stability of mRNA.
2. EGFP – green florescent protein that enables selection by flow cytometry.
3. PGK promoter – GFP promoter.

Then using Gibson Assembly these fragments were integrated to the pEF1a-FB-dCas9-puro plasmid linearised using SpeI restriction plasmid to form the pEF1a-FB-dCas9-GFP, intermediate construct. This intermediate construct brings GFP and promoter in close proximity, so that it can be used as a template in the formation of pEF1a-FB-dCas9-GFP-virus v4

Step 1: pEF1a-FB-dCas9-puro plasmid was double digested by SpeI (**Figure 3.5**) which has two restriction sites. The larger (10,020 bp) fragment was then isolated by agarose gel electrophoresis and extracted using Monarch[®] DNA Gel Extraction Kit. Next, fragments encoding for Enhanced Green Fluorescent Protein (eGFP) and the second for a phosphoglycerate kinase (PGK) promoter were amplified by PCR using PCRBIO HiFi Polymerase, templates and primers detailed in **Table 3.2**. The fragments were quantified by Nanodrop[™] Lite spectrophotometer and annealed using the Gibson assembly process. Following incubation for 15 minutes at 50°C with the Gibson Assembly Master Mix (NEB), as per the manufacturer's instructions, they were transformed in NEB DH5α competent cells. Successful cloning was confirmed by diagnostic digest and agarose gel electrophoresis, then Sanger sequencing.

Name	Size (bp)	Function	Template	Primers
Vector		Backbone	pEF1a-FB-dCas9-puro	
Fragment 1	760	Enhanced GFP	MIGRI	Forward: ggcctttcgaccagcttaccatggcgagca aggg
				Reverse: ctagaaggcacagtgcaggcttactgtac agctcgtccatgccg
Fragment 2	321	PGK promoter	pEF1a-FB-dCas9-puro	Forward: tcctcgcccttgctcgccatggtaagctggt cgaaaggccc
				Reverse: gcgtcgtgcaggacgtgacaaatggaagt agcacgtctcatctcgtgcagat

Table 3.2 Primers and vectors used in first step of pEF1a-FB-dCas9-puro → pEF1a-FB-dCas9-GFP-virus v4

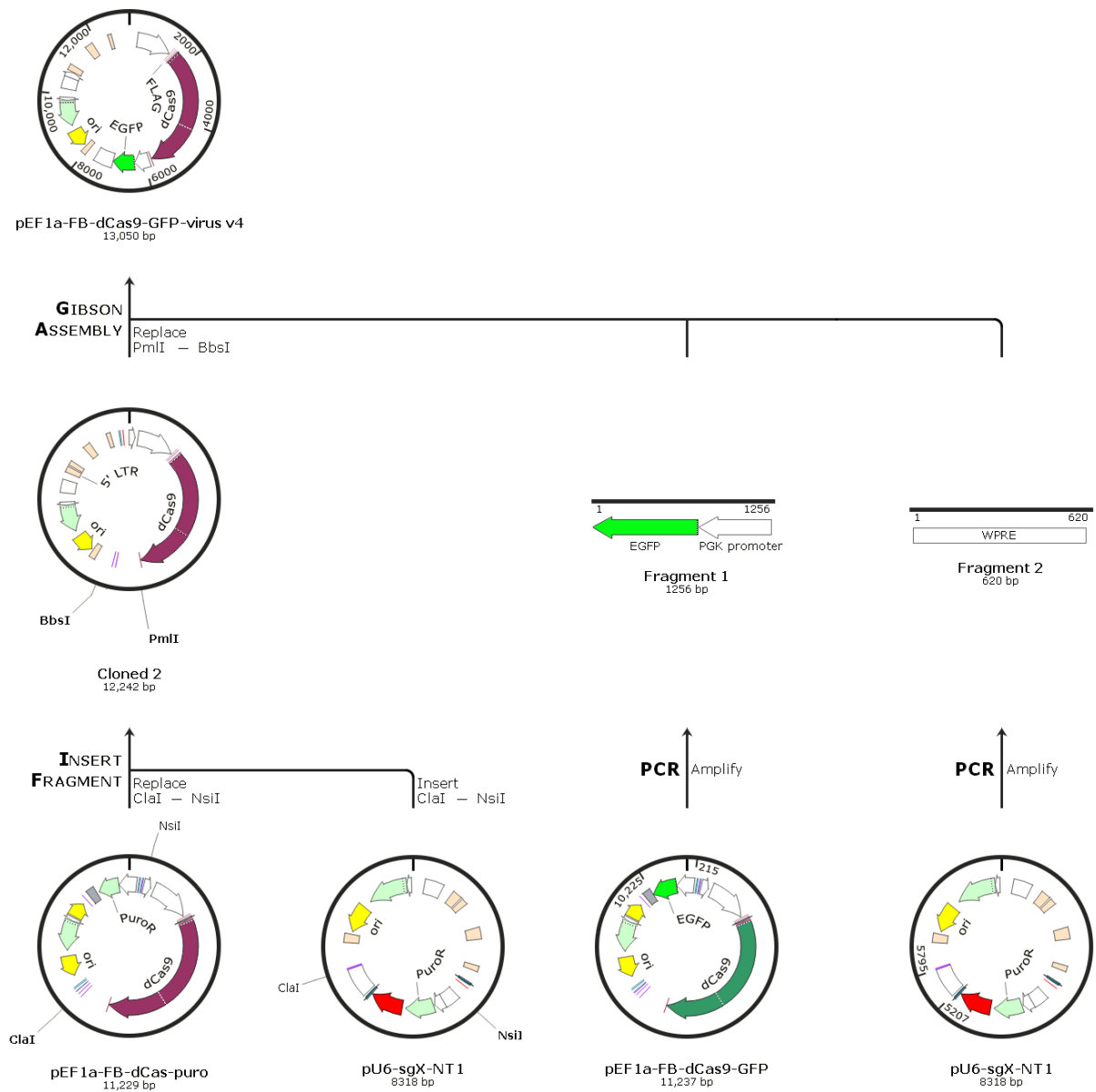


Figure 3.6 Step 2 of pEF1a-FB-dCas9-puro → pEF1a-FB-dCas9-GFP-virus v4 for viral transduction

The intermediate construct, pEF1a-FB-dCas9-GFP, was used as a template for modification of the original pEF1a-FB-dCas9-puro plasmid. pEF1a-FB-dCas9-puro and pU6-sgX-NT1 plasmids were restriction digested and ligated creating 'cloned 2'. EGFP and promoter, and WHP Posttranscriptional Response Element were amplified by PCR and integrated to the linearised cloned 2 plasmid by PmlI and BbsI restriction enzymes using Gibson Assembly. The modifications of pEF1a-FB-dCas9-puro to pEF1a-FB-dCas9-GFP-virus v4 ensures additional expression of flanking LTRs to enable viral transduction and GFP to allow selection by flow cytometry.

Step 2 began with double digests of pEF1a-FB-dCas9-puro and pU6-sgX-NT1 plasmids by ClaI and NsiI (**Figure 3.6**) restriction enzymes. The 6608 bp fragment pEF1a plasmid and 5634 bp fragment from pU6 plasmid were then isolated by agarose gel electrophoresis and extracted using Monarch® DNA Gel Extraction Kit. The fragments were ligated using the T4 DNA ligase, following the manufacturer's protocol, incubation was at 22°C for 10 minutes. The ligation reaction products were transformed in NEB DH5α competent cells. Successful cloning was confirmed by diagnostic digest and agarose gel electrophoresis, then Sanger sequencing (**Appendix D figures 1-2**). Next, the new clone was digested with BbsI and PmlI restriction enzymes. The 11,238 bp fragment was then isolated by agarose gel electrophoresis and extracted using Monarch® DNA Gel Extraction Kit. Two fragments encoding for Enhanced Green Fluorescent Protein (EGFP) and its promoter and WHP Posttranscriptional Response Element were amplified by PCR using PCRBIO HiFi Polymerase, templates and primers detailed in **Table 3.3**. The fragments and plasmid backbone were quantified by Nanodrop™ Lite spectrophotometer and then annealed using the Gibson Assembly process. Following incubation for 15 minutes at 50°C with the Gibson Assembly Master Mix, as per the manufacturer's instructions, they were transformed into NEB DH5α competent cells. Successful cloning was confirmed by diagnostic digest and agarose gel electrophoresis, then Sanger sequencing.

Name	Size (bp)	Function	Template	Primers
Fragment 1	1256	EGFP and PGK promoter	pEF1a-FB-dCas9-GFP	Forward: agaggttgattttacttgtagctcggtccatgcc
				Reverse: gagtatcaggtctccacgggtaggggagcgcg
Fragment 2	620	WRPE	pU6-sgX-NT1	Forward: tgtacaagtaaaatcaacctctggattaca aaatttgtaaaga
				Reverse: gtagatccacagatcaaggagcggggagcgcg

Table 3.3 Primers and vectors used in second step of pEF1a-FB-dCas9-puro → pEF1a-FB-dCas9-GFP-virus v4

These fragments provide PCR templates for amplification. The resultant products were combined with the intermediate construct 'cloned 2' by Gibson Assembly.

Fragment 1 includes the sequence for GFP, fluorescent selectable marker, and associated promoter.

Fragment 2 contains the sequence for the Woodchuck Hepatitis Virus Post-transcriptional Response Element.

It increases mRNA stability and protein yield

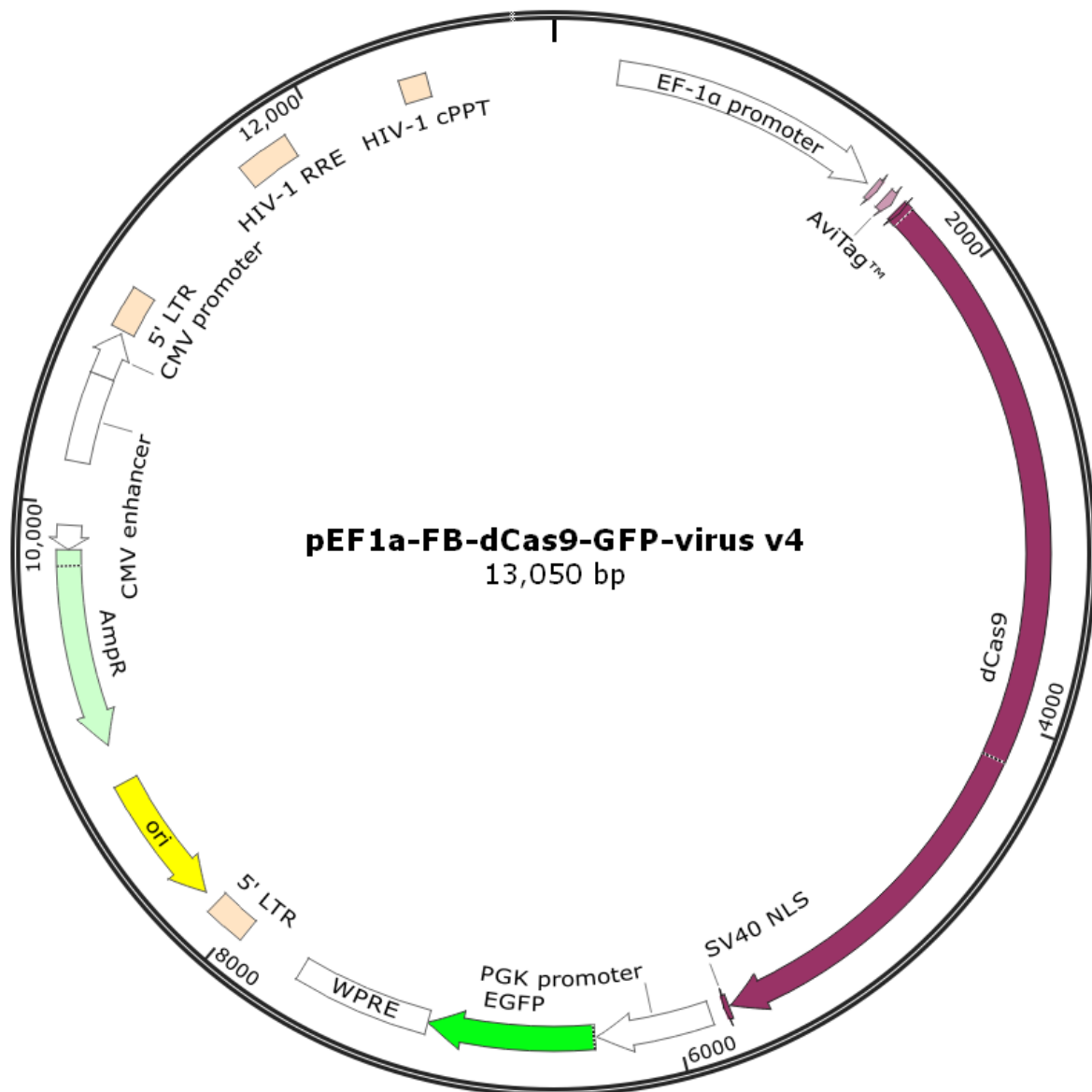


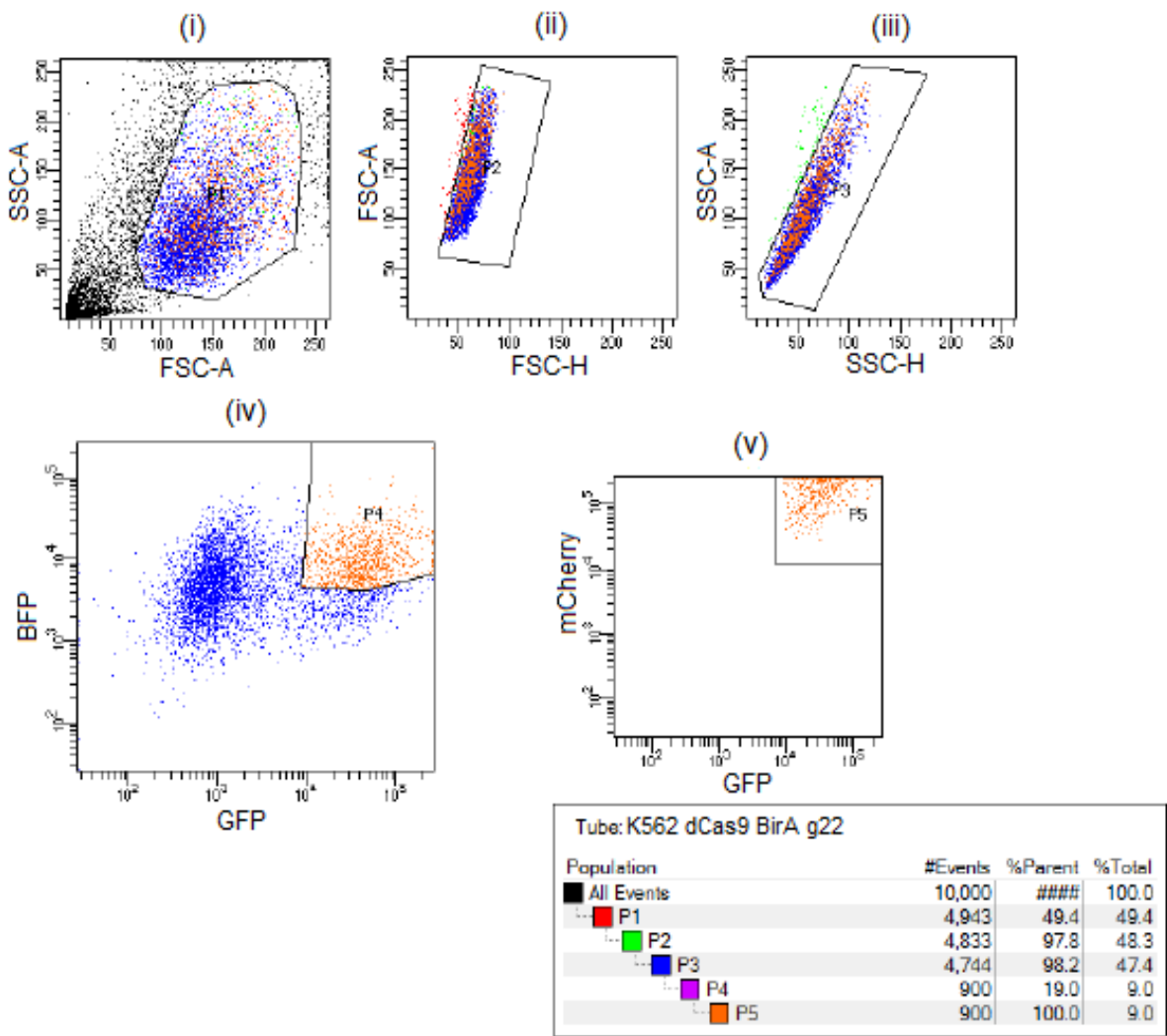
Figure 3.7 Modified pEF1a-FB-dCas9-GFP-virus v4 expressing EGFP selectable marker and flanking LTRs for viral transduction

Chimeric long terminal repeats flank the transgene sequence that facilitates the integration of transfer plasmid sequences into host genome. The transgene sequence includes a BirA ligase expression cassette driven by the EF-1a promoter, CMV promoter and a gene for the Green Fluorescent Protein under control of PGK promoter. The transgene sequence also contains WPRE, RRE, AviTag and cPPT promote transgene expression. Woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) increases transgene expression by facilitating mRNA transcript maturation. AviTag facilitates biotinylation of deactivated Cas9 by biotin ligase. Central polypurine tract (cPPT) improves vector integration and transduction efficiency. Other sequence include ampicillin (AmpR) resistance genes. Each replicon for antibiotic resistance gene has an origin of replication (*ori*) which enables plasmid self-replication

This was an ambitious endeavour but would simplify the scaling up of CAPTURE cell lines.

The G418 resistance gene of pEF1a-BirA-neo plasmid was replaced with BFP and LTRs added, while the pEF1a-FB-puro plasmid was modified to replace puromycin resistance with GFP fluorochrome and incorporate LTRs. Diagnostic digests and selection by flow cytometry indicated successful modification, **Figure 3.3**. Unfortunately, western blot analysis confirmed that neither pEF vector expressed the protein of interest.

A



B

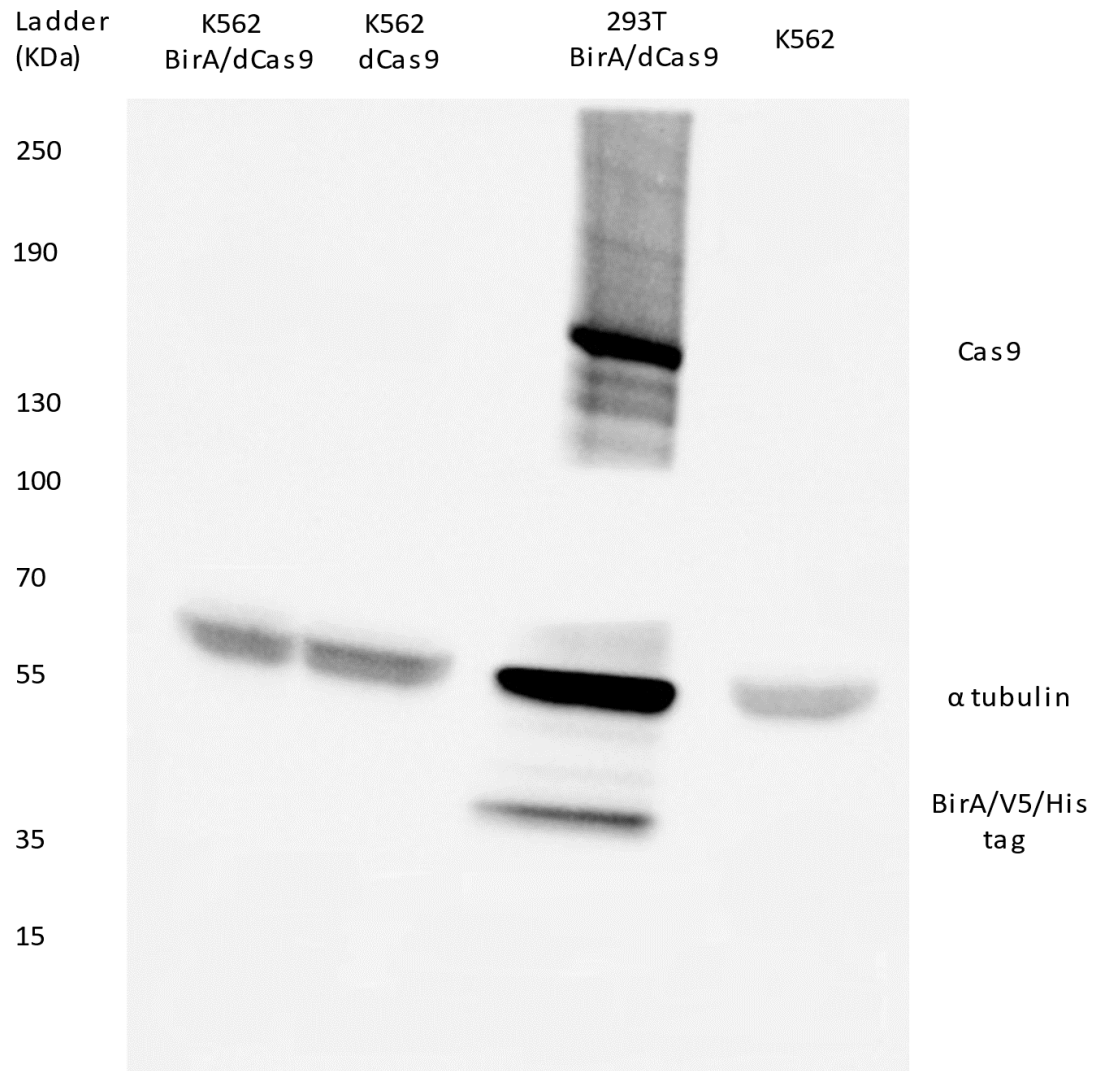


Figure 3.8 Expression profile of modified CAPTURE plasmids by flow-cytometry and western blot

(A) Gating strategy for flow-cytometry-based sorting of K562 cells transduced with three CAPTURE plasmids (P1-5). Gates depict events in a hierarchical order, (P1) FSC-A/SSC-A (cell size and granularity) - excludes debris, (P2) FSC-H/FSC-A (single cells) - doublet discrimination, (P3) SSC-H/SSC-A (2nd doublet discrimination), (P4) GFP/BFP (selection for pEF1a-FB-dCas9-GFP-virus/pEF1a-BirA-V5-BFP-virus positive cells), (P5) GFP/mCherry (selection for pEF1a-FB-dCas9-GFP-virus/pU6-sgX-NT1 expressing cells). Bottom right, population hierarchy.

(B) Western Blot assessing expression of pEF1a-FB-dCas9-GFP-virus and pEF1a-BirA-V5-BFP-virus constructs. Negative control – native K562 cells. Positive control – 293T cells transiently transfected by lipofection with pEF1a-FB-dCas9-GFP-virus and pEF1a-BirA-V5-BFP-virus plasmids. K562 cells transduced with modified pEF1a plasmids did not express Cas9 nor BirA protein

I postulated that the transgene sequence exceeded the lentiviral capacity and incomplete sequences were integrated to the host genome.

3.3.4 Establishment of CAPTURE cell lines

Following the difficulty modifying the two pEF plasmids for viral transduction and fluorescent selection. I opted to return to pEF1a-FB-dCas-puro and pEF1a-BirA-V5-neo, the original pEF plasmids. Here, I would transiently transduce the above plasmids by electroporation and determine optimal concentrations of the selection antibiotics to ensure continuous expression of the genes required for CAPTURE.

Section 2.2.5 outlines how continuous expression of these transiently transfected plasmids was achieved. This required optimisation of the transfection conditions using Neon™ Transfection system and the antibiotic concentrations required to maintain selection pressure to ensure pEF plasmid expression. This is outlined in **table 3.4**.

To establish optimal transduction of the two pEF plasmids using the neon transfection system in K562 cells required a series of optimisation experiments. The manufacture's recommendations, **table 2.3**, were generic to suspension cells and gave broad ranges on parameters such as cell density, plasmid and buffer concentrations and Neon™ Transfection system programme. This required a series of optimisation experiments to identify the best

machine settings and reagent concentrations for K562 cell transfection. This process increased transfection efficiency by approximately 20%.

Pulse voltage (V)	Pulse width (ms)	Pulse number	Cell density (cells/ml)	Transfection efficiency	Viability	Tip type
1000 - 1450	10-50	1-4	$1-1.51 \times 10^7$	83-90%	88-90%	10 μ l

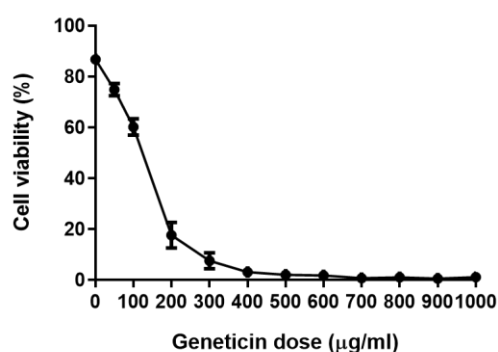
Format	DNA (μ g)	Pulse number	Cell density (cells/ml)	Transfection efficiency	Viability	Tip type
10 cm	5-30	1-4	$2-5 \times 10^6$	83-90%	88-90%	10 μ l

Table 3.4 (A) Generic electroporation parameters for transduction of suspension cell lines. (B) Outline of generic cell culture and transduction conditions.

Both electroporation parameters and culture condition were optimised for K562 cell line, **table 2.3**.

After the transfection of pEF1a-FB-dCas-puro and pEF1a-BirA-V5-neo plasmids. The pU6-sgX-NT1 vector, after successfully cloning the appropriate gRNA, was virally transduced to the modified K562 cells. Thereafter, the clones were maintained by selection with G418 and puromycin at optimal concentrations, **figure 3.9**, to ensure expression of dCas9 and BirA proteins.

A



B

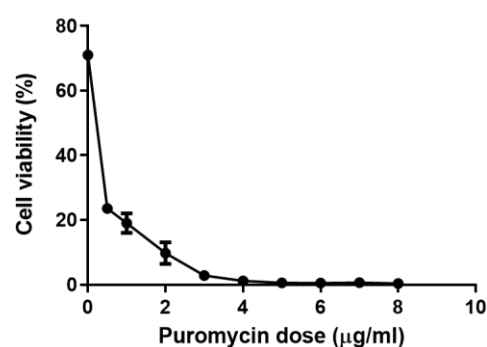


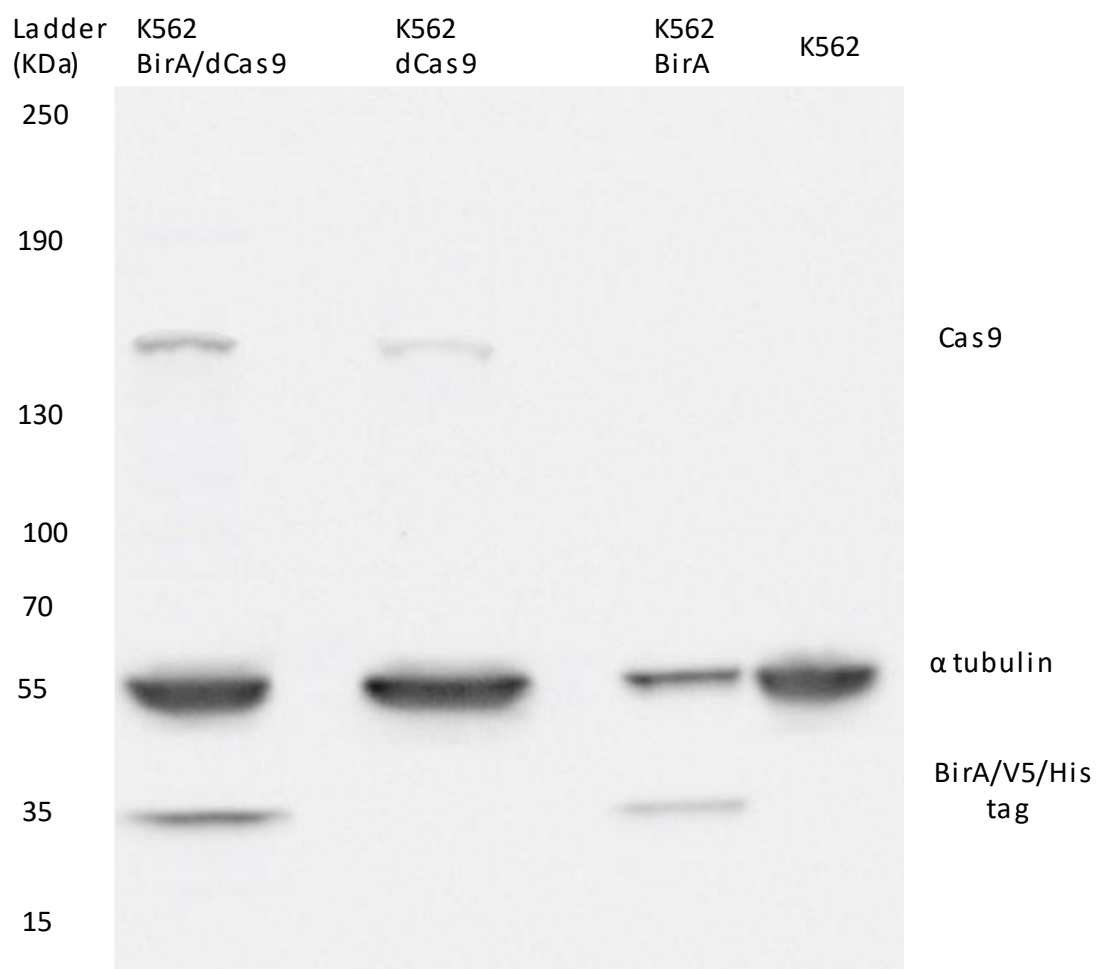
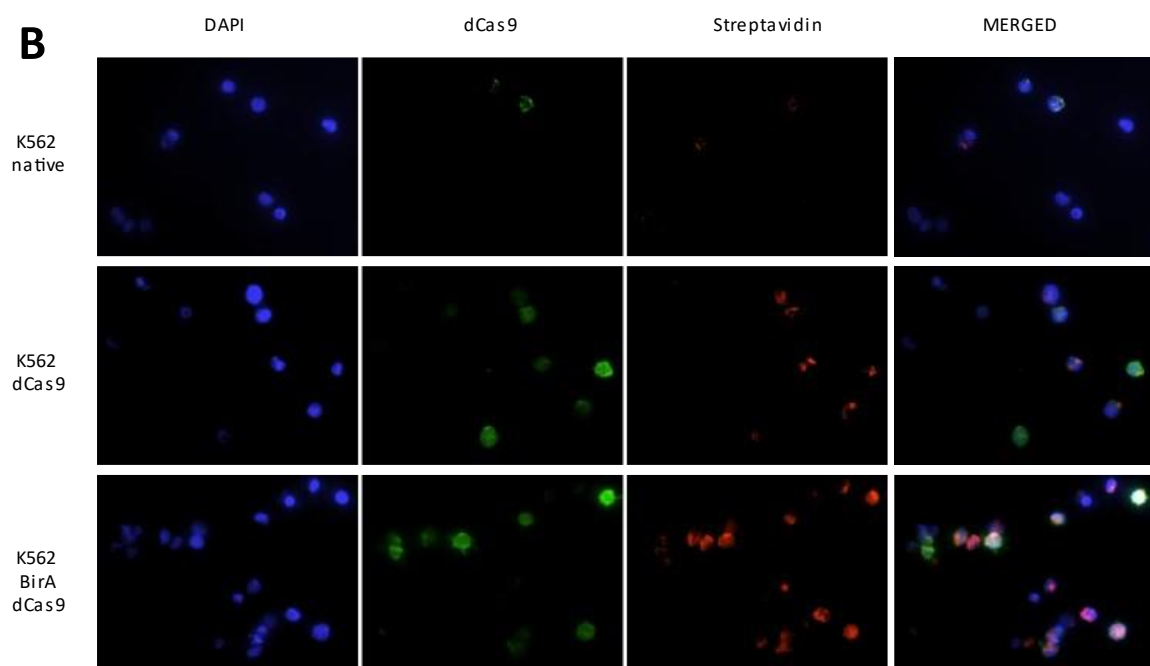
Figure 3.9 Geneticin (A) and puromycin (B) toxicity curve for K562 cell line.

2.5x10⁵ of native K562 cells were suspended in media treated with G-418 at varying concentrations, as illustrated in figure 2.8. Each concentration was assessed in duplicate and incubated for 10 days. Assessment of viability was assessed every two days by flow cytometry.

Mean concentrations are represented, with mean error bars. n=2.

Optimal concentration was the minimum concentration at which all cells are killed by selection antibiotic.

By following a methodical approach, I was able to increase transduction efficiency 20-fold, generating cell lines capable of expressing the necessary components to facilitate locus-specific pull down of the genetic targets.

A**B**

Deconvolution microscope (x40 magnification)

Figure 3.10 *In vitro* confirmation of expression profile for CAPTURE K562 cell lines

(A) Western blot assessing expression profiles of modified K562 cells transfected with pEF1a-FB-dCas9-puro and pEF1a-BirA-V5-neo plasmids to express dCas9 and biotin ligase, BirA (V5 affinity tag), respectively. Internal positive control - α tubulin and negative control cell line – native K562 cells were used. There was no positive control cell line used in this experiment.

(B) Deconvolution microscopy showing intracellular immunofluorescence of native K562 cells (top panels), pEF1a-FB-dCas9-puro transfected K562 cells, expressing ‘deactivated’ Cas9 (middle), and K562 cells transfected with pEF1a-FB-dCas9-puro and pEF1a-BirA-V5-neo, co-expressing dCas9 and biotin ligase.

Left to right: DAPI stain (blue) identifies fixed/live cells. Cas9-FITC (green). Streptavidin-Cy3 (red). Panels on right, MERGED, are composite images illustrating the fluorescence pattern of Cas9-FITC and Streptavidin-Cy3 which confirms the co-expression of biotinylated Cas9 in K562 cells transfected with both pEF plasmids. This pattern is not seen in native nor dCas9, only expressing cell lines.

3.4 dCAS9 INTERFERENCE

3.4.1 Guide RNAs

The gRNAs were another key component of the pulldown machinery required to ensure locus specific purification of the regions of interest: *G6PD*, *TPI* and *GAL4*. The selection and cloning of these guides required specific considerations. The first and most important consideration was the design, particularly of the *G6PD* guides. **Figure 2.5** illustrates the position of the g4 downstream of the HDAC responsive region. Unpublished work by Makarona *et al* (replicated in **Figure 3.11**), illustrates the negative impact of the binding of dCas9 at this region on *G6PD* induction. This transcriptional interference results in a 1.5- to 2-fold reduction in expression in response to NaBu.

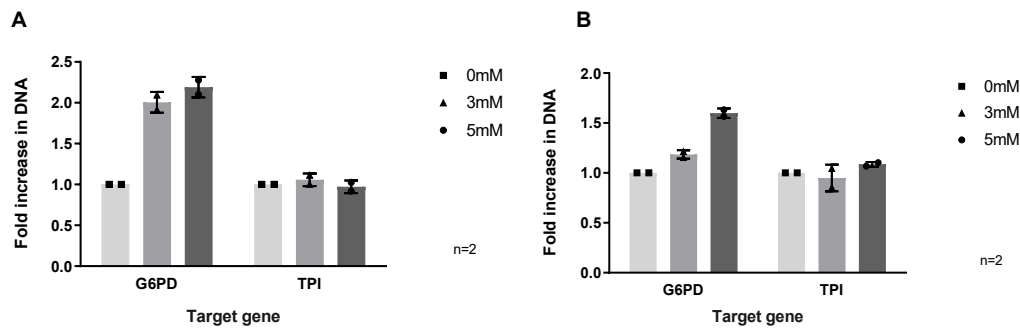


Figure 3.11 Comparative induction of inducible and non-inducible GPPP enzymes in 293T and 293T dCas9 g4 cells after treatment with sodium butyrate and incubation for 8 hours. (A) 293T and (B) 293T dCas9 g4.

Chromatin immunoprecipitation was performed using 2×10^7 native 293T/293T dCas9 g4 cells. The cells were incubated for 4 hours in 10 μ M SAHA/vehicle. After which the cells were chemically crosslinked with paraformaldehyde, followed by sonication to 2Kb fragments. Chromatin immunoprecipitation was performed using polyclonal anti-Cas9 antibody and *G6PD* DNA eluted. *G6PD* DNA was quantified by RT-PCR and normalised to IgG control.

The data shows the impact downstream gRNA has on *G6PD* transcription. Maximal selective induction was observed at 5mM sodium butyrate. This was reduced by 47% in the 293T dCas9 g4 cell line compared with the native 293T cells.

Each experiment was performed in duplicate and fold increase in *G6PD* expression was compared to control. Error bar show mean value with standard deviation, n=2.

I postulated that placing the gRNA downstream of the TSS and the responsive region interferes with *G6PD* transcription, limiting the increased gene expression expected following treatment with HDAC inhibitor.

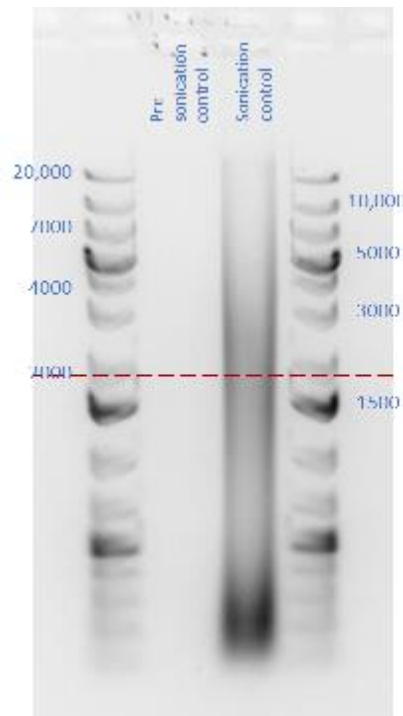
Therefore, subsequent gRNAs were designed to anneal to distal promoter areas upstream of the transcription start site and HDACi-responsive region to avoid transcriptional interference. *G6PD* gRNAs were designed using an online resource MIT CRISPR design tool (<http://crispr.mit.edu>) which predicted on-target activity and off-target effects. A proprietary algorithm generated a score that combined the probability of Cas9 off-target binding and on-target score, which represented the cleavage efficiency of Cas9 (225, 226). The higher the score, the greater the on-target activity and fewer off-target effects. Although the literature suggests that a score >60 offered sufficient specificity (227), I

selected three guides for cloning with scores >90. After successful cloning of two of the three *G6PD* gRNAs into the pU6-sgX-NT1 plasmid, I followed a similar process when cloning pU6-sgTPI-NT1. Here, only two suitable guides were identified, one was successfully cloned. One GAL4 guide was cloned into the pU6-sgX-NT1 plasmid.

3.4.2 Chromatin fragmentation

As outlined in section 3.4.1, the large dCas9 protein gives rise to transcriptional interference when positioned downstream of TSS and responsive region. Therefore, the CAPTURE protocol was modified so that, unlike conventional ChIP protocols where fragments are usually 250-500bp in length, this protocol would use fragments of approximately 2Kb to accommodate the regulatory regions, proximal gene body and the upstream dCas9 protein (Appendix E Table S1) . **Figure 3.12** Illustrates that consistent fragmentation of chromatin was achieved using Covaris S220 Focused-ultrasonicator (protocol detailed in section 2.5.1.2).

A



B

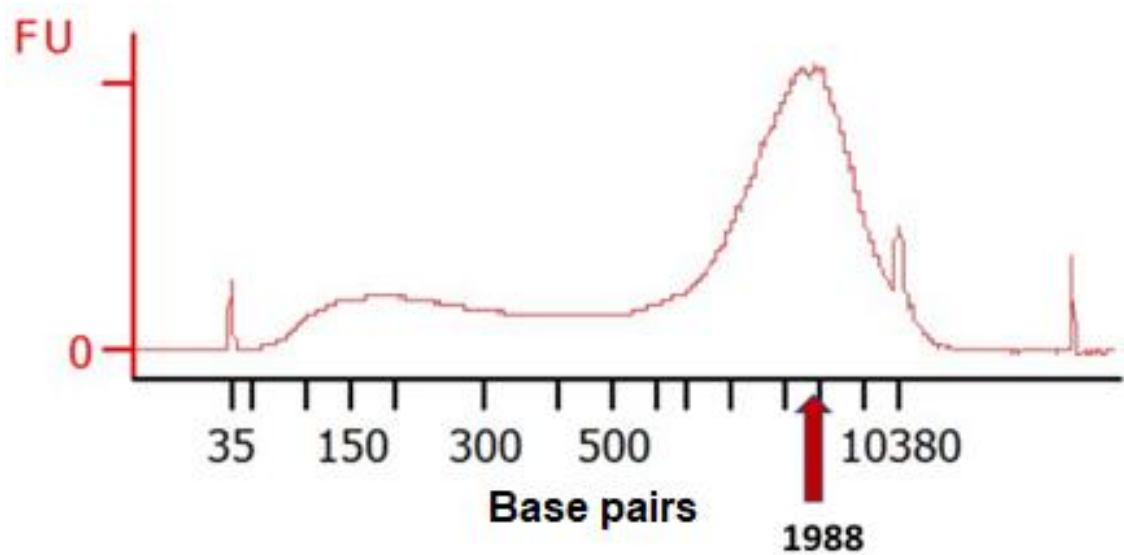


Figure 3.12 sonication control confirming optimisation of 2Kb fragments for CAPTURE protocol (10^7 K562 cells). (A) Gel electrophoresis (B) Bioanalyzer trace showing fragment size of approximately 2Kb.

Height Threshold (FU): is a measure of whether a peak is maintained. It represents the minimal peak height.

3.5 OPTIMAL CONDITIONS FOR SELECTIVE INDUCTION OF *G6PD* EXPRESSION BY SAHA

The next major challenge was to identify the optimal conditions at which *G6PD* activity is selectively induced by SAHA and its tagged derivative. A bespoke biotinylated SAHA was synthesised by Sai Life Sciences Ltd. This Bio SAHA was used to monitor the functional state of HDAC complexes to decipher the mechanism by which its parent drug selectively induces *G6PD* expression. This would require consideration of both ChIP protocols so the datasets could be compared.

To ensure comparison of CAPTURE and Chem ChIP data, the conditions at which a tagged SAHA phenocopies SAHA in *G6PD* induction were explored. Makarona *et al*'s pioneering work in this lab demonstrated that following treatment with HDAC inhibitor there was a three-fold increase in *G6PD* activity from baseline (74). These data were reproduced using SAHA (**Figure 3.9**) but not with the biotinylated derivative. I noted that the modified SAHA was cytotoxic at either high dose or following prolonged exposure. And with lower concentrations or shorter incubation times, insufficient induction was achieved. **Figure 3.13** shows dose response and time course experiments using biotinylated SAHA. Despite attempts at numerous incubation times and drug doses, it became clear that I would be unable to mimic the selective induction of *G6PD* activity using Bio SAHA. The conditions at which ≥ 2 *G6PD* activity was achieved also resulted in non-specific induction of other GPPP enzymes, **Figure 3.13**. Prolonged drug exposure had cytotoxic effects and ultimately resulted in reduced cell viability. The GM12878 cell line was especially susceptible to these effects. Evidence of non-specific GPPP enzyme induction, likely due to cellular stress, was also noted. The activation of these pathways in response to cellular stress is well

documented (228). This revelation coincided with the discovery detailed in section 3.2.1, relating to transduction difficulties with GM12878 cell line.

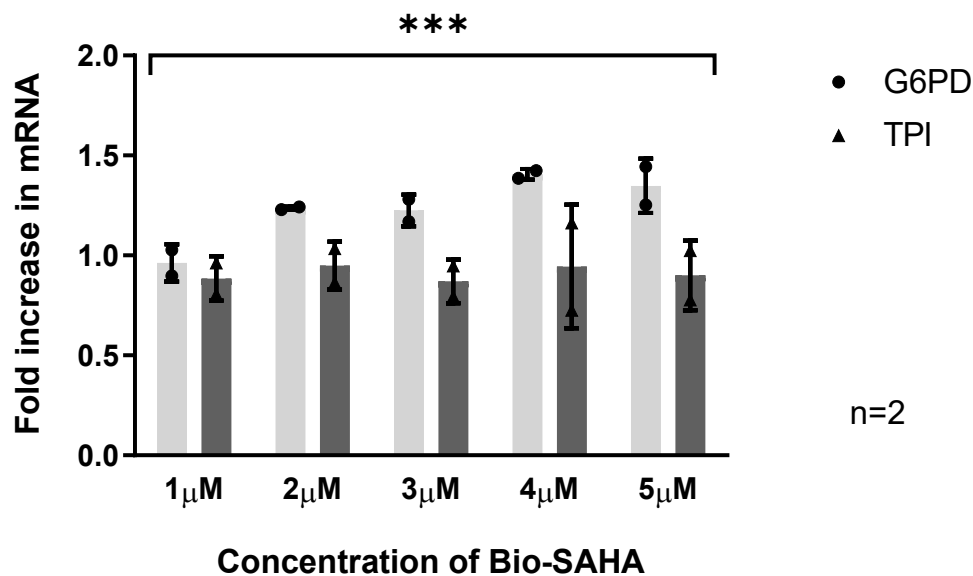


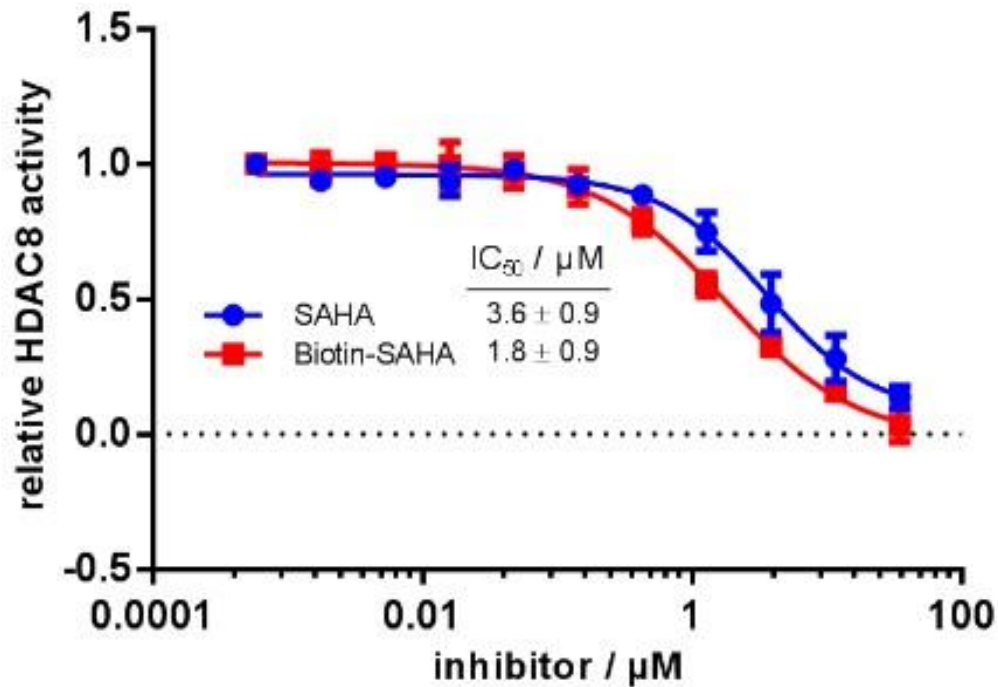
Figure 3.13 Biotinylated SAHA dose response experiment at 8 hours, normalised to control

Dose response experiments were performed to determine the optimal conditions at which Bio-SAHA selectively induces *G6PD* expression. 1×10^6 GM12878 cells were incubated with vehicle or varying concentration (1-5 μM) of Bio-SAHA for 8 hours. After which RNA was extracted followed by cDNA synthesis. RT-PCR quantified *G6PD* and *TPI* expression at each concentration. The results were normalised to vehicle control and performed in duplicate (n=2).

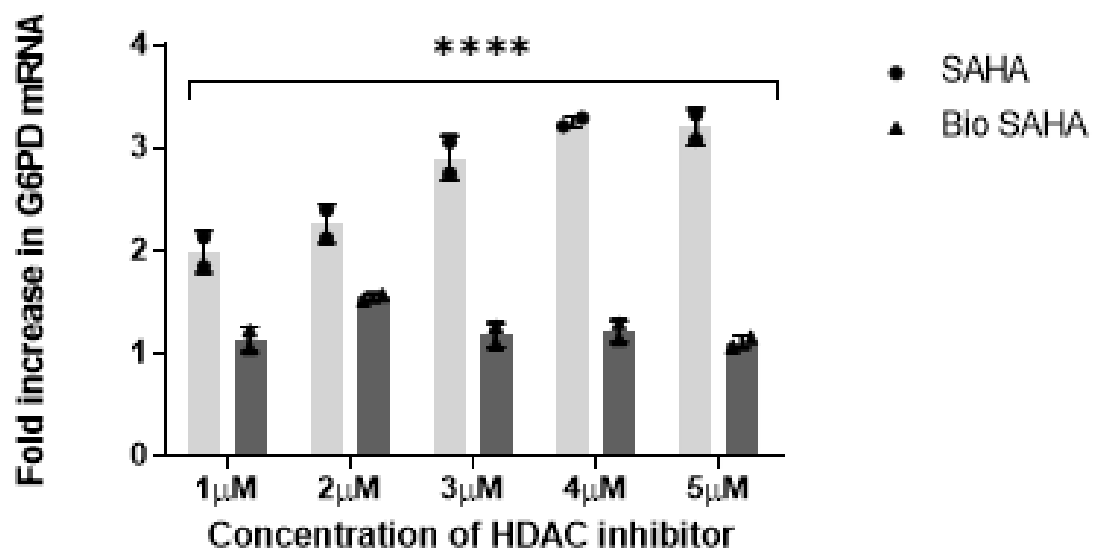
P value (2 way ANOVA): 0.0004. There is statistical difference expression of *G6PD* and *TPI* in Bio SAHA treated groups. However, *G6PD* induction in GM12878 cell line by biotinylated SAHA was consistently less than 1.5 times baseline. Therefore, is insufficient to proceed to downstream analysis. Note error bars represent mean with standard deviation

3.6 STERIC HINDRANCE DUE TO BIOTIN MOIETY

A



B



C

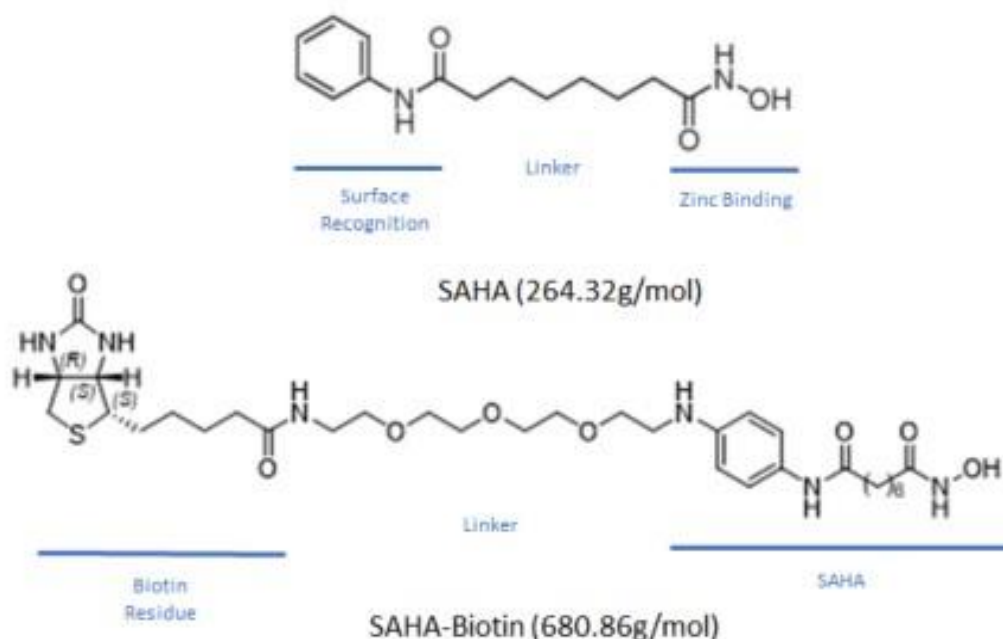


Figure 3.14 SAHA and biotinylated modifications

(A) *In vitro* activity of SAHA and biotinylated SAHA. IC₅₀ indicating biotinylated SAHA is two times more effective at inhibiting HDAC8 than the parental drug.

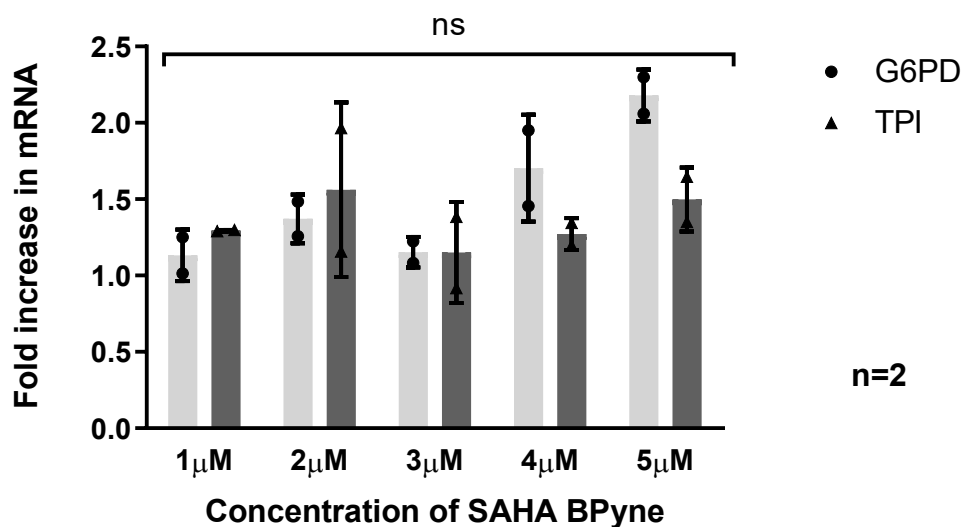
(B) Comparative dose response of SAHA and Bio-SAHA in GM12878 cells at 6 hours, normalised to DMSO control. *G6PD* induction is significantly greater than the modified drug, particularly at high concentrations where SAHA induces *G6PD* expression 3-fold greater than the Bio-SAHA. Also Bio-SAHA induction suboptimal at <2 times baseline.

Error bars indicate mean with standard deviation, n=2. 2 way ANOVA demonstrates statistical significance between *G6PD* induction in SAHA treated cells compared with those treated with Bio-SAHA (p value: <0.0001).

(C) Chemical structures and molecular weights of SAHA and Bio SAHA. The addition of the biotinylated tag increases the size of the drug almost three times and likely causes significant steric hindrance

3.6.1 Adoption of Clickable SAHA

A



B

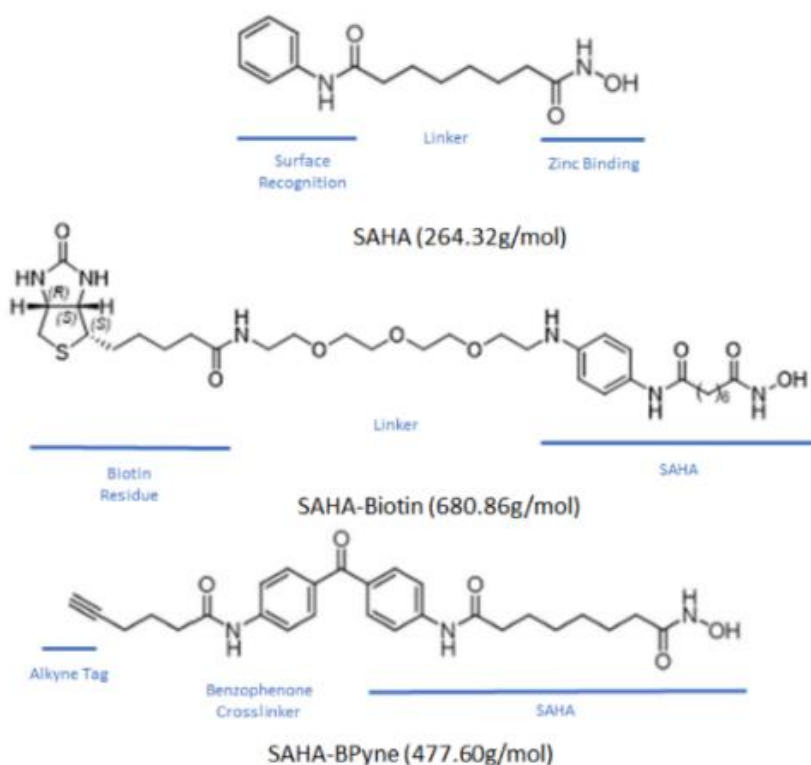


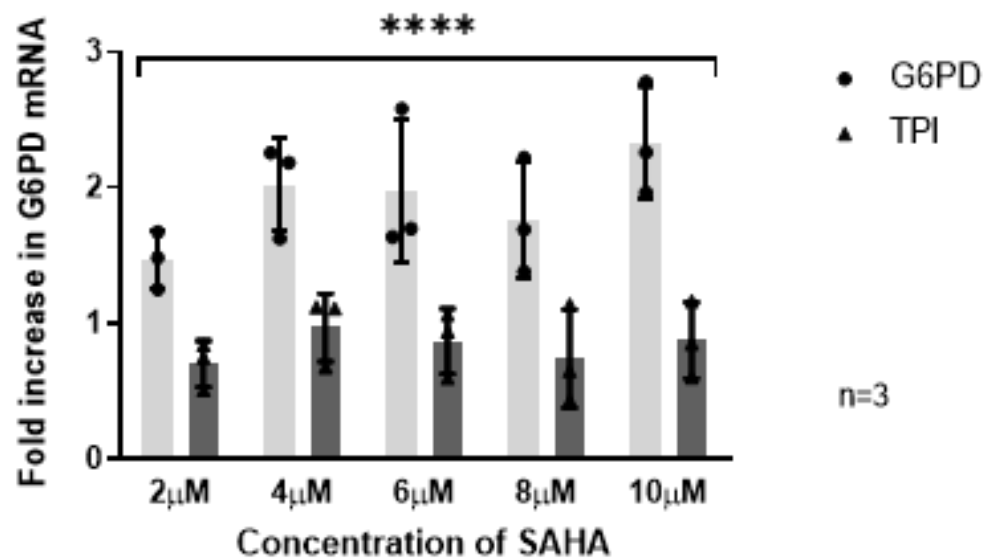
Figure 3.15 SAHA and tagged modifications

(A) Dose response showing the effect of SAHA BPyne on *G6PD* and *TPI* transcription in GM12878 cell line after 6 hours, normalised to vehicle control. At low SAHA BPyne concentration *G6PD* induction is sub-optimal at <1.5 times baseline. However, as drug concentration is increased, *G6PD* induction increases, but to achieve target 2x baseline induction, non-specific *TPI* induction was also observed. 2 way ANOVA demonstrates that

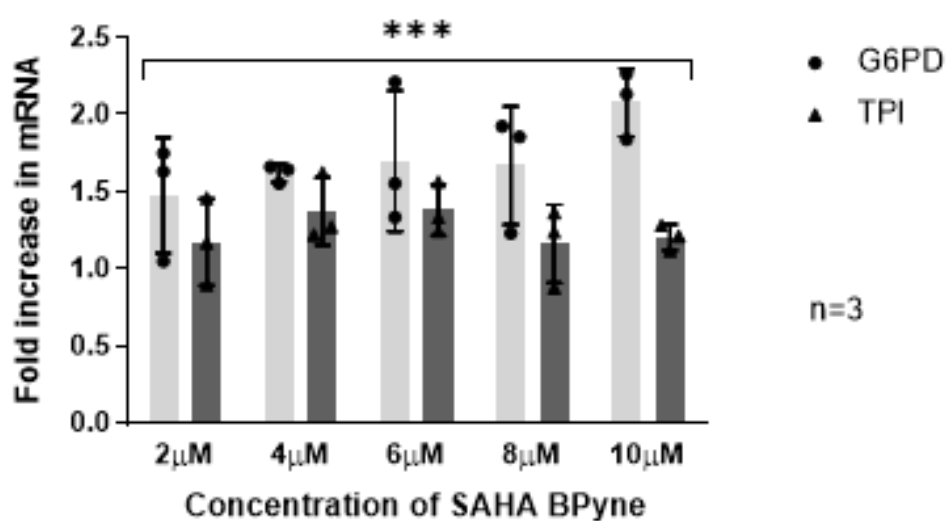
there is no statistically significant difference between *G6PD* and *TPI*, the non-inducible GPPP enzyme, expression after varying doses of SAHA BPyne following incubation for 6 hours, p value (0.2289). Error bars indicate mean with standard deviation, n=2.

(B) Chemical structures and molecular weights of SAHA, Bio SAHA and SAHA BPyne. Unlike the biotinylated form, the clickable moiety of SAHA BPyne increases the molecular weight by ~100g/mol and is therefore less likely to interfere with the drug's ability to interact with the tubular HDAC pocket

A



B



C

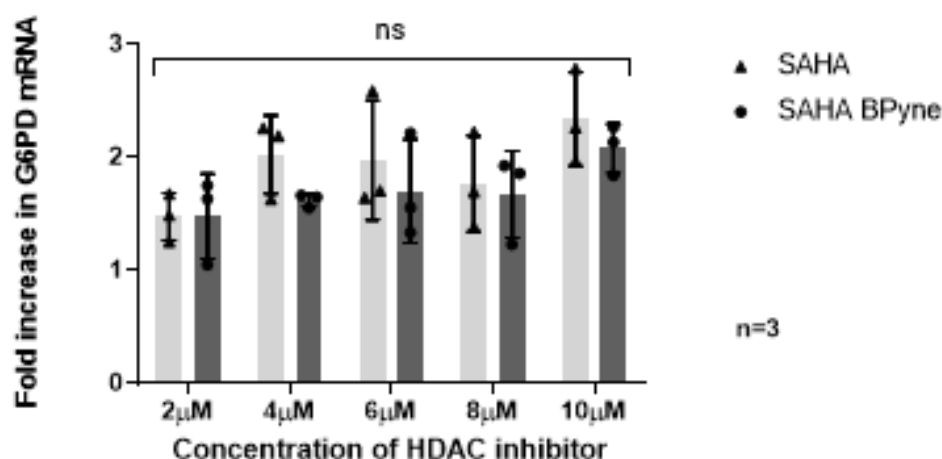


Figure 3.16 Optimisation of reaction conditions for selective HDAC inhibitor induced *G6PD* upregulation.

(A) illustrates *G6PD* and *TPI* expression in native K562 cells following incubation for 4 hours with parental SAHA. *G6PD* and *TPI* transcription was quantified by RT PCR and normalised to DMSO control. Dose response shows that at concentrations greater than 2μM SAHA induces *G6PD* expression to greater than 2x baseline. *TPI* induction remained at baseline. 2 way ANOVA demonstrates that there is statistically significant difference between *G6PD* and *TPI* expression at all concentrations of SAHA after incubation for 4 hours, p value (<0.0001). Error bars represent mean with standard deviation, n=3.

(B) shows *G6PD* and *TPI* expression in native K562 cells following incubation for 4 hours with SAHA BPyne. After which *G6PD* and *TPI* transcription was quantified by RT PCR and normalised to methanol control. Dose response shows that only at 10μM of SAHA BPyne induction of *G6PD* expression exceeded 2x baseline. Non-specific induction was observed to a greater extent than that seen with parental SAHA, evidence by *TPI* induction. Though marginal, consistently exceeded baseline. 2 way ANOVA demonstrates that there is statistically significant difference between *G6PD* and *TPI* expression at all concentrations of SAHA after incubation for 4 hours, p value (<0.005). Error bars represent mean with standard deviation, n=3.

(C) comparative dose response of *G6PD* expression following treatment of native K562 cells with HDAC inhibitor for 4 hours. This illustrates that at 10μM SAHA BPyne phenocopies SAHA, showing equivalence of *G6PD* expression, with minimal non-specific induction (A and B). P value (ANOVA): 0.1399, Error bars represent mean with standard deviation, n=3

3.7 SCALING UP

Quantitative MS analysis of proteomic changes due to HDAC inhibition generally requires a large amount of starting material as, in this case especially, the region of interest is located

at only two loci within the genome. To culture and process cells in numbers that would yield protein amounts sufficient for MS analysis, in four different CAPTURE cell lines, under two different conditions, in duplicate, was prohibitive and would have created multiple challenges.

3.7.1 Measures to limit cell numbers

Following discussion with Dr Whitwell, a collaborator with expertise in general and histone proteomics, it was estimated that the minimum histone protein required to ensure detection of post translational modifications using the mass spectrum technique employed was estimated at 1.5µg. To achieve this quantity of protein, I estimated this would require in the order of 10^{10} cells per condition. This would not have been achievable given the cellular system used. I explored isobaric labelling to multiplex samples to improve sensitivity and reduce cell numbers. The addition of these tags would enable pooling multiple biological samples (229). This reduces instrument time, reagents and M/S run to run reproducibility as fewer runs were required. But more importantly, the pooling of samples would potentially lead to an upto eight-fold increase of peptide signal, providing relative quantification. This reduces the starting sample requirement; this is particularly important for low abundance peptides.

Overall, in conjunction with TMT labelling and other general measures such as optimisation of cross-linking, sonication, pull down to reduce sample loss, I was able to reduce the starting cell number requirement by ten-fold.

3.7.2 Cell culture

Culturing 10^9 instead of 10^{10} cells for eight different reaction conditions was challenging.

Selection of the most appropriate culture system required consideration of multiple factors.

Conventional systems using “T” flasks in a static incubator are easy to use and do not require additional equipment. However, they are extremely labour intensive and require considerable incubator space. I estimated that I would have required forty-four T225 flasks to culture 10^9 cells, this was not a practical option. As an alternative I had the option to use Erlenmeyer flasks in a shaking incubator. This was an attractive option as mass cultures could be grown in a fraction of the space. And also suitable for suspension cells such as K562 – and high-cost equipment was available in the lab. However, this approach had the disadvantage of consuming large volumes of reagents and was still quite labour intensive due to regular media changes. I opted to proceed with this option.

The optimised shaking incubator conditions were 1L Erlenmeyer flasks at 100rpm with cultures of no more than 4×10^8 cells. Scaling up to 2L flasks resulted in significant cell death. I suspect the increased centrifugal force, even at reduced shaking speed, affected cell integrity. I also utilised baffled flasks to ensure optimal gas exchange without excessive agitation. Baffling increases agitation and therefore oxygenation, improving doubling time without causing foam formation. This modification reduces foaming which act as an oxygen barrier, offsetting the benefits of agitation.

The result of the optimisation steps was a three-fold increase in doubling time and a more compact culture.

3.7.3 Chromatin fragmentation

Previous optimised sonication protocol for CAPTURE-seq required further refinement for the CAPTURE-proteomics workflow. To enable sonication of chromatin from 10^9 K562 cells, I transitioned from Covaris S220 Focused-ultrasonicator (in which its 1ml sonication tubes have a maximum capacity of 1×10^7 cells) to Bioruptor® Pico sonication system with a capacity of 6×10^8 cells per run, making the sonication process feasible. The added benefit of this approach was that this sonication system used sonication beads to guarantee greater reproducibility and even distribution of the energy for efficient shearing throughout samples, **Figure 3.14**. The Bioruptor® Pico sonication system uses a gentle method of sonication to retain the integrity of chromatin, protein-DNA complexes and protein-protein interactions.

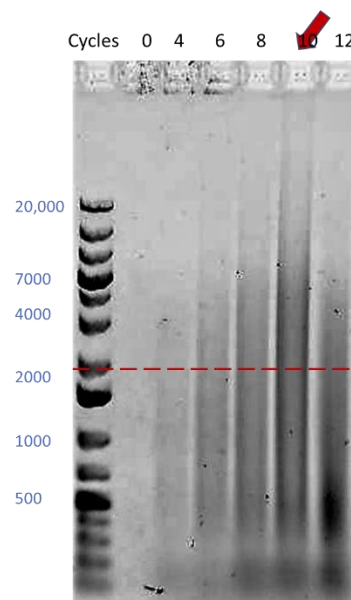


Figure 3.17 Sonication control confirming effective and uniform fragmentation of chromatin from 2.5×10^8 CAPTURE cell lines per run. This approach enabled 10 times more cells per run than previous approach

By significantly shortening the pulse time, extending the total time per cycle and limiting the number of cycles, I was able to maintain large fragments reliably and efficiently, with no unsonicated chromatin and minimal over-sonicated product.

3.8 SUMMARY OF PROCESS

The adaptation of Click ChIP and CAPTURE was particularly challenging and required a rigorous stepwise approach to develop working protocols. In both cases, these modalities replaced other less robust techniques.

The main challenges of this project were the establishment of CAPTURE cell lines and modification of plasmids; identification of optimal conditions for selective induction of *G6PD* expression with tagged SAHA; dCas9 interference with design of appropriate guide RNAs; and mass culture. Each of these issues was addressed and solutions developed.

The choice of the most appropriate biological system was particularly challenging. GM12878 was theoretically the most appropriate choice as a tier 1 ENCODE cell line of haematological lineage and normal karyotype. However, this B cell line was impractical as I was unable to successfully transduce the necessary constructs and identify the conditions at which SAHA and modified SAHA selectively upregulated *G6PD* transcription. This fundamental problem required significant evolution of the protocols which included substituting GM12878 for K562. Although K562 is a cell line with a malignant karyotype, it is of erythroid lineage and the cell line used in the seminal publication that describes the CAPTURE technique (195).

Next, the biotinylation of SAHA impacted its pharmacological efficacy *in vivo*. The change to SAHA BPyne enabled ≥ 2 selective induction of *G6PD* transcription, so the Click ChIP protocol was adopted.

By modifying the CAPTURE technique, I was able to establish stable cell lines that expressed the necessary molecular machinery to allow purification of *G6PD* chromatin for analysis. In addition, culturing these cell lines in a shaking incubator allowed mass production of these CAPTURE cell lines to ensure sufficient starting material for downstream proteomic analysis.

The annealing of *G6PD* gRNA downstream of the promoter region resulted in the dCas9 macromolecule inhibiting transcription. This issue was resolved by designing guide RNAs which bound upstream of the HDACi responsive region. These changes – in conjunction with a sonication protocol which generated 2Kb fragments – allowed dissection of all relevant epigenetic interactions.

In summary, I have adapted two novel ChIP protocols to explore the primary aim of my project. By replacing GM12878 with K562 cells, Chem ChIP with Click ChIP, and enChIP with CAPTURE, I was able to establish a cellular system that would allow me to address my primary biological questions from both a genome-wide and locus-specific perspective with two robust and functional modalities.

Chapter Four

Click ChIP

4 CLICK CHIP

4.1 INTRODUCTION

4.1.1 Background

The dynamic reversible acetylation of lysine residues on histone tails plays a critical role in transcriptional regulation (230). Histone deacetylases have been traditionally regarded as co-repressors that bind gene promoters, inhibiting transcription through histone deacetylation (231). However, a genome-wide study (77, 232) and work from the Karadimitris lab (74) have demonstrated that this characterisation is more nuanced, and that the role of histone acetylation and deacetylation of active, primed and inactive genes is more complex than previously described. These studies demonstrated that active genes, such as housekeeping genes, showed increased HAT and HDAC binding at the promoters and gene bodies, whilst primed genes showed increased binding at only the promoter regions. It has since been hypothesised that HAT/HDAC activity occurs in cycles requiring deacetylation at the end of each phase of transcription to prevent gene hyperacetylation (77, 232).

4.1.2 Current understanding of the epigenetic effects of SAHA

HDAC inhibitors disrupt the dynamic equilibrium described in section 4.1.1, giving rise to disproportionate HAT activity altering gene expression affecting cell proliferation, cell death pathways and other critical cellular functions (**Table 4.1**) (233).

It is also well established that H3 and H4 histone proteins are the primary histone deacetylase targets for HDAC inhibitors. Of these two histones lysine 9,14, 18 and 23 of H3 and 16 of H4 are targets for acetylation (81). Therefore, there is an assumption that these would be the main histone targets for SAHA. The interaction of these histone marks – and which of these are more important to the action of SAHA – is poorly described.

Understanding the changes in expression profiles will help unravel the impact that HDAC inhibitors have on transcription and the role they play in chromatin regulation via interaction with chromatin-modifying enzymes and transcription factors.

In this chapter, I explore the genome wide transcriptomic and proteomic changes that occur following treatment with SAHA to better understand the mechanism by which this HDAC inhibitor selectively upregulates *G6PD* transcription. By utilising the Click ChIP modality, a ligand-affinity capture technique, coupled with genetic sequencing or mass spectrometry. I sought to gain a better comprehend the impact of SAHA on the human genome.

4.2 EXPERIMENTAL DESIGN

Achieving the primary objectives of the project required the use of new, innovative methodologies to monitor the functional state of HDAC complexes in native biological systems providing an unprecedented interrogation of the human genome. This method

couples ligand-affinity capture with DNA sequencing or mass spectrometry to identify the sites bound by small chemical molecules throughout the genome. By using a tagged SAHA as a molecular probe, Click CHIP provides the ability to map the direct interactions of a chemical entity with genome-wide chromatin analysis to provide new and important insights into chemical perturbation of cellular function.

I compared the chromatin pulled down from cells treated with clickable SAHA and those treated with vehicle and unmodified SAHA. Downstream analysis with RNA seq, ChIP seq and MS proteomics was performed. Here, RNA seq aims to demonstrate similar impact of SAHA and SAHA BPyne on the transcriptome thereby reinforcing the notion that the addition of the molecular tag does not affect molecular interactions of the drug. Proteomic analysis would be particularly valuable by demonstrating how HDAC inhibitors physically associate with HDAC and HDAC-containing chromatin complexes as this is likely to be a key component of the protein-protein interactions by which these drugs exert their pharmacological effect.

4.3 IMPACT OF SAHA ON TRANSCRIPTOMIC LANDSCAPE

4.3.1 Introduction

HDACs are a group of highly conserved isoenzymes which form the catalytic core of large molecular complexes involved in chromatin modification (155). These include the CoREST, NuRD and Sin3 complexes which contain a HDAC1-HDAC2 dimer at its core, and the NCoR complex associated with HDAC3 (156, 234). SAHA is a small molecule HDAC inhibitor which consists of a Zn^{2+} -binding moiety, an alkyl chain linker spanning the length of the tubular pocket (**Figure 1.7**), and a cap that blocks the active site by interacting with the external surface to fit into the pocket and inhibit the deacetylase activity (187, 188).

Although SAHA is categorised as a pan-inhibitor it only interacts with the ‘classical’ HDACs – I, II and IV – which have zinc-dependent active sites, unlike the NAD^{+} -dependent Class III enzymes. However, the exact nature of these molecular interactions is poorly understood (189).

4.3.2 Analysis of gene expression in SAHA treated cell lines

To assess the impact of HDAC inhibitors on the transcriptome, RNA seq was performed on SAHA, SAHA BPyne and methanol treated K562 cells. The threshold for statistical significance was set at a $\text{padj} < 0.05$ and $\log_2 \text{FC} \geq 1.5$. Of the 1414 genes identified in the SAHA BPyne samples, 563 genes were upregulated and 853 downregulated compared to vehicle. Of the 2056 genes expressed in the SAHA samples, 951 genes were upregulated and 1104 downregulated of a total of 2056 genes compared to vehicle (**Figure 4.1**).

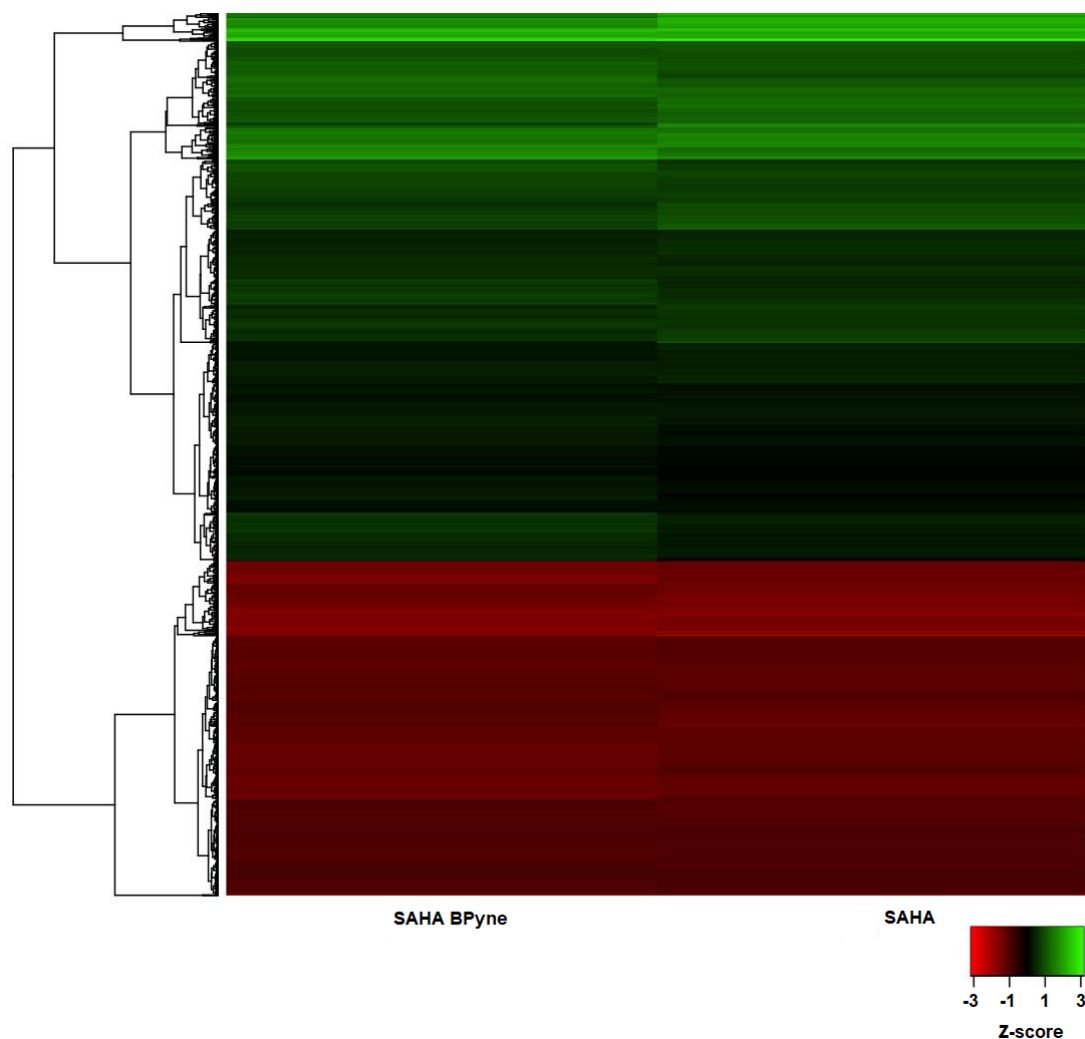


Figure 4.1 Hierarchical clustering of differentially expressed genes in SAHA and SAHA BPyne treated K562 cells.

Heatmap of differential gene expression in K562 cells treated with SAHA compared to SAHA BPyne normalised to vehicle, n=2.

Z score is a statistical measure of a value's standard deviation from the mean.

Dendrogram connecting data points in a hierarchical tree. Hierarchical cluster analysis was performed using Ward's linkage a method that specifies the distance between two clusters as the increase in the sum of squares. When applied with Euclidean distance, the calculated incremental sum of squares is square rooted. This data shows similarity between the SAHA and SAHA BPyne treated datasets.

The heatmap demonstrates the remarkable similarity between the two datasets. The proportion of shared genes differentially expressed between the two groups was 87% in the SAHA BPyne treated samples (**Figure 4.3**).

PCA plot (**Figure 4.2**) also illustrates similarity between the SAHA- and SAHA BPyne – treated samples, and significant dissimilarity between vehicle and SAHA/ SAHA BPyne treated samples.

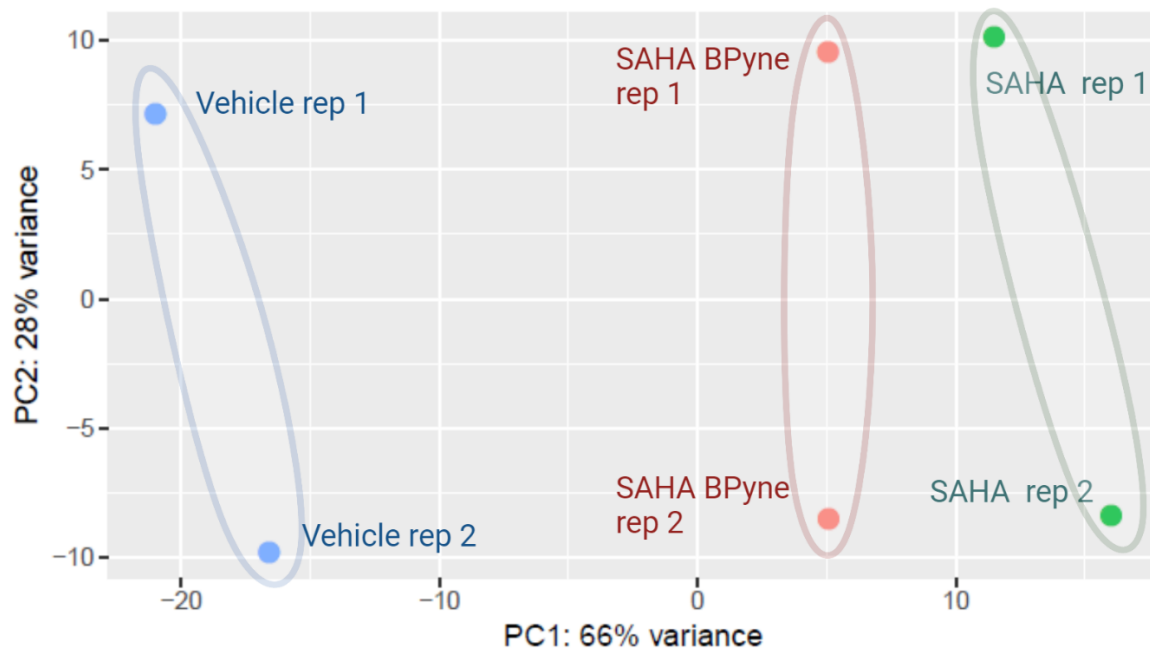


Figure 4.2 Correlation between Click RNA seq histone proteomics conditions in K562 cells treated with SAHA BPyne/SAHA /DMSO

Principal component analysis plot depicting the variance between K562 cells treated with either vehicle, SAHA BPyne or SAHA. Performed in duplicate, n=2. The graph shows the correlation between variably treated cells by plotting the principal component (PC) 1, the attribute with the greatest variability, against PC2, the 2nd greatest variable to allow similarities to be identified. Therefore, the difference along the x axis is of greatest significance.

The SAHA and SAHA BPyne treated cells show greater similarity than those cells treated with vehicle. The variation between replicates likely represents the differing *G6PD* induction from baseline

Of note, there is variation in the degree of induction between biological replicates, although all samples achieved ≥ 2 times induction of *G6PD* mRNA expression. This difference in induction affected the expression profiles for each replicate.

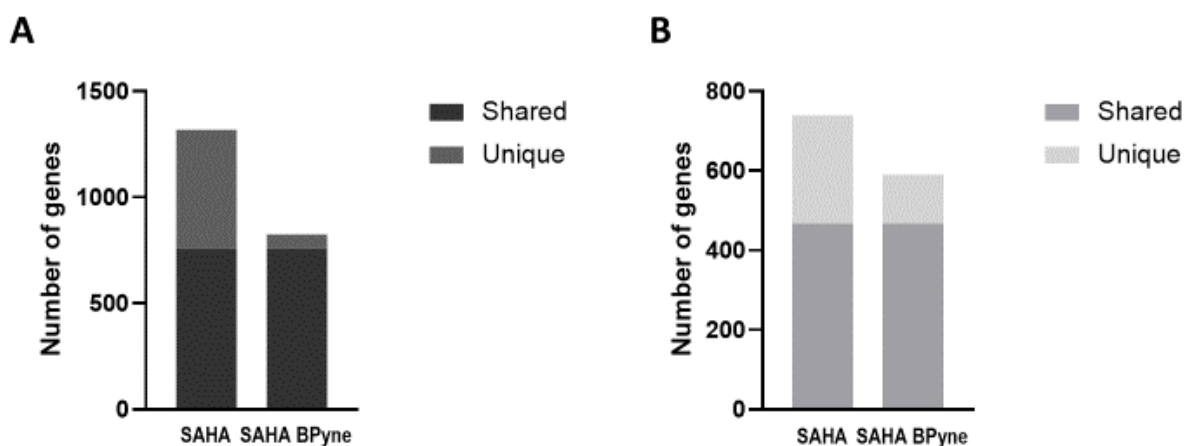


Figure 4.3 illustrates the proportion of shared genes within the SAHA and SAHA BPyne datasets.

(A) upregulated genes (padj < 0.05, log₂ fold change > 0) and (B) downregulated genes (padj < 0.05, log₂ fold change < 0). The majority of the differentially expressed genes in the SAHA BPyne dataset as also seen in the SAHA dataset.

4.3.3 Differentially expressed genes in SAHA and SAHA BPyne samples

As there are no reference datasets for SAHA treated K562 cells, I used the SAHA treated dataset as a control for the SAHA BPyne treated cells. At this stage in the protocol, the tag has yet to be employed so I would expect the RNA seq data to be similar across the two groups.

The data from SAHA and the tagged SAHA experiments served several purposes. The first, was to illustrate, on a global scale, the similarities in the effect of the two drugs on the transcriptome. This indicated that the affinity tag had little or no impact on the action of the drug, unlike the biotinylated SAHA. But also to indicate possible mechanisms by which SAHA exerts its epigenetic effect.

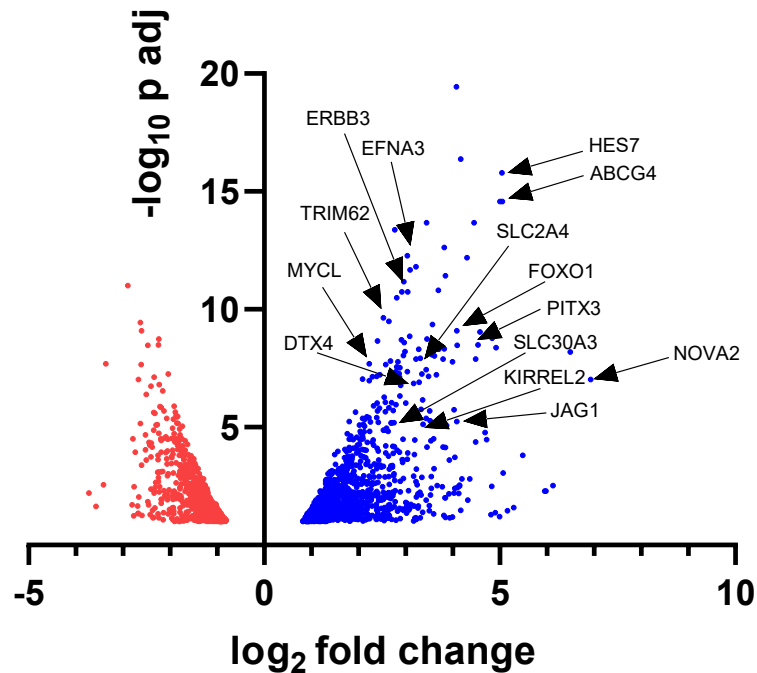


Figure 4.4 Volcano plot highlighting the expression profile of SAHA BPyne treated samples.

Plot of differential gene expression in K562 cells treated with SAHA compared to SAHA BPyne normalised to vehicle ($p_{adj} < 0.05$, $\log_2$ fold change ≥ 1.5), $n=2$. Upregulated genes are denoted by red dots and downregulated genes by blue dots. Selected genes are denoted by black arrowheads ($p_{adj} > 5$, $\log_2$ fold change > 2)

Using the PantherDB gene ontology knowledgebase, I reviewed the 50 most-highly enriched genes (**figure 4.4**). This showed that several interesting classifications dominated, particularly gene-specific transcriptional regulators, ten hits were noted in this category which comprises DNA-binding transcription factors and transcription cofactors. This group includes Forkhead box protein O1 and Pituitary homeobox 3.

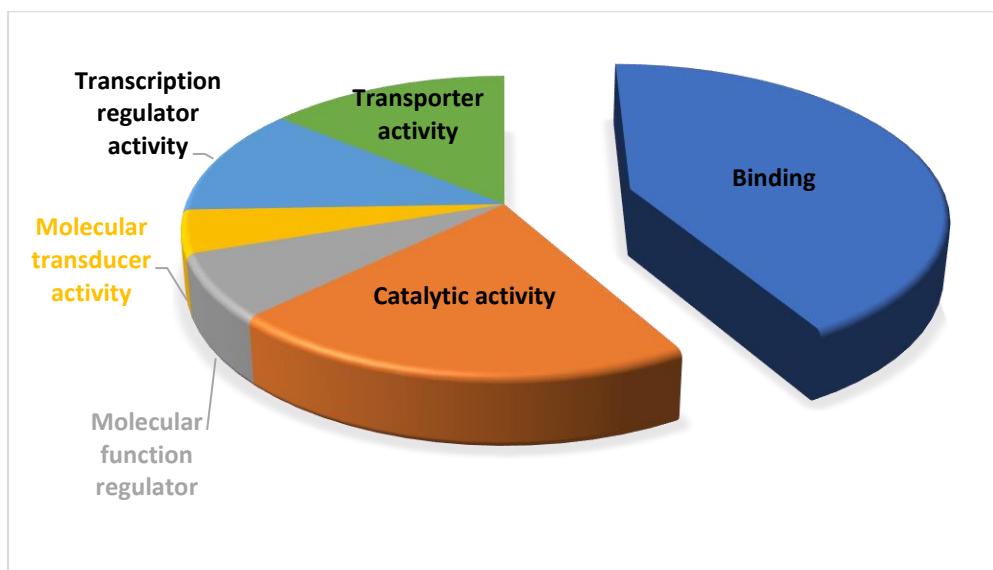


Figure 4.5 Top 50 overrepresented genes in the SAHA BPyne treated samples.

Chart shows GO slim molecular function classification for the top 50 genes in SAHA BPyne RNA seq dataset, normalised to vehicle ($p_{adj} < 0.01$, $\log_2$ fold change ≥ 3.75), $n=2$.

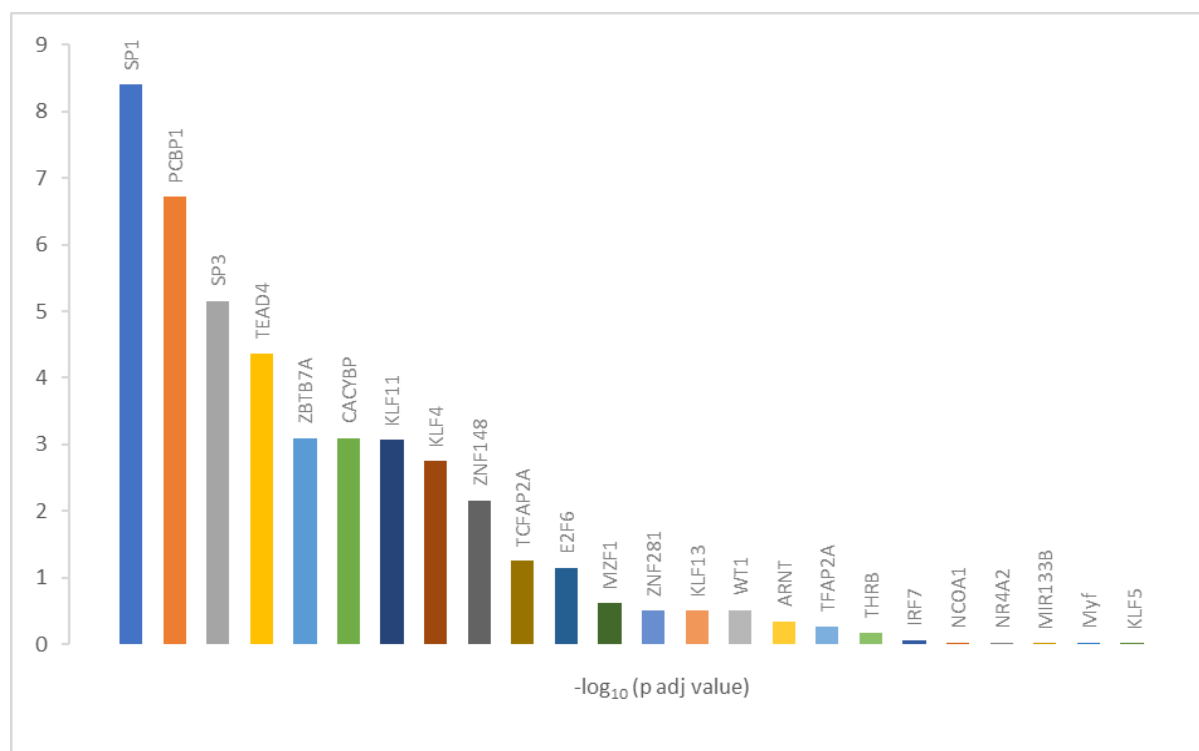
18 of these genes were classed in the binding category which denotes selective, non-covalent molecular interactions. These includes Forkhead box protein O1 and Pituitary homeobox 3. These genes play a role in stress response pathways

Interestingly, the top 50 downregulated genes show a similar profile to the upregulated genes with similar molecular function categories dominating. Also a review of the differentially expressed genes in the SAHA treated samples revealed similar patterns to SAHA BPyne.

4.4 ENRICHMENT ANALYSIS OF HDAC INHIBITOR TREATED CELL TRANSCRIPTOME

4.4.1 Transcriptional and histone modification gene set analysis

Enrichment analysis was performed on the differentially expressed genes of these datasets using the open access enrichment database Enrichr (235).



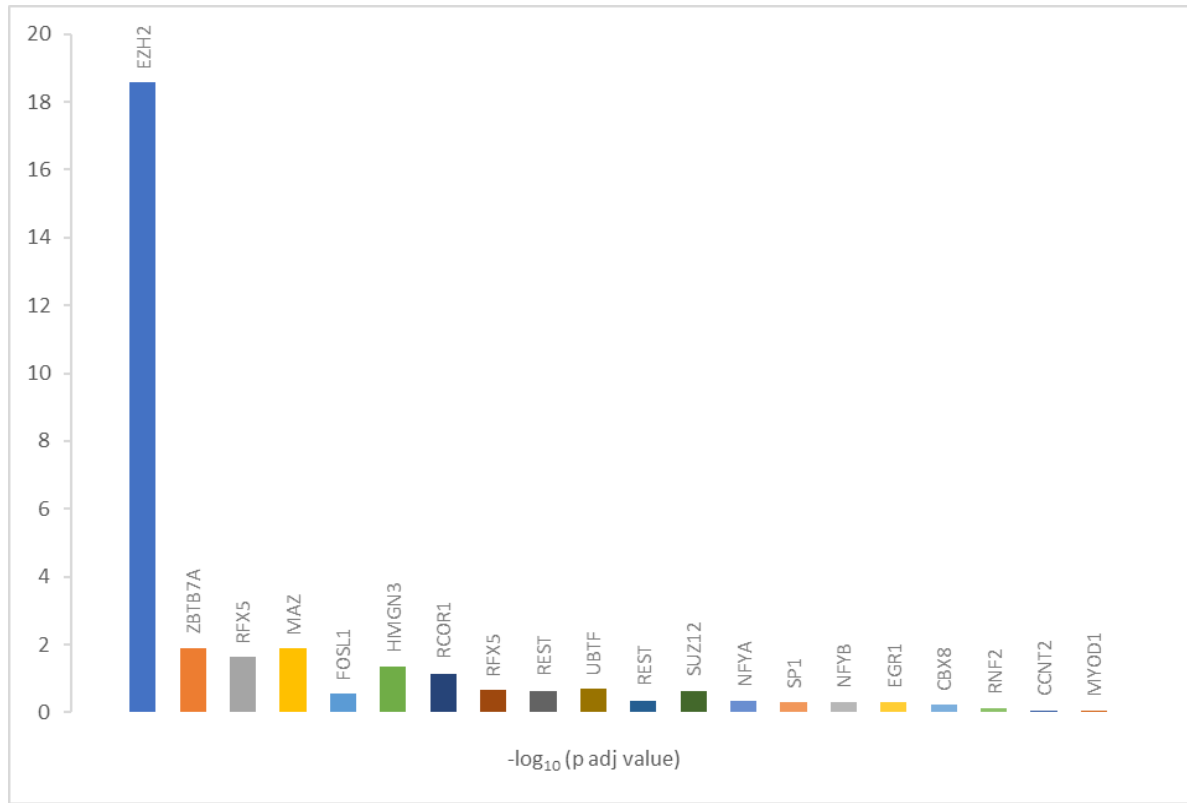
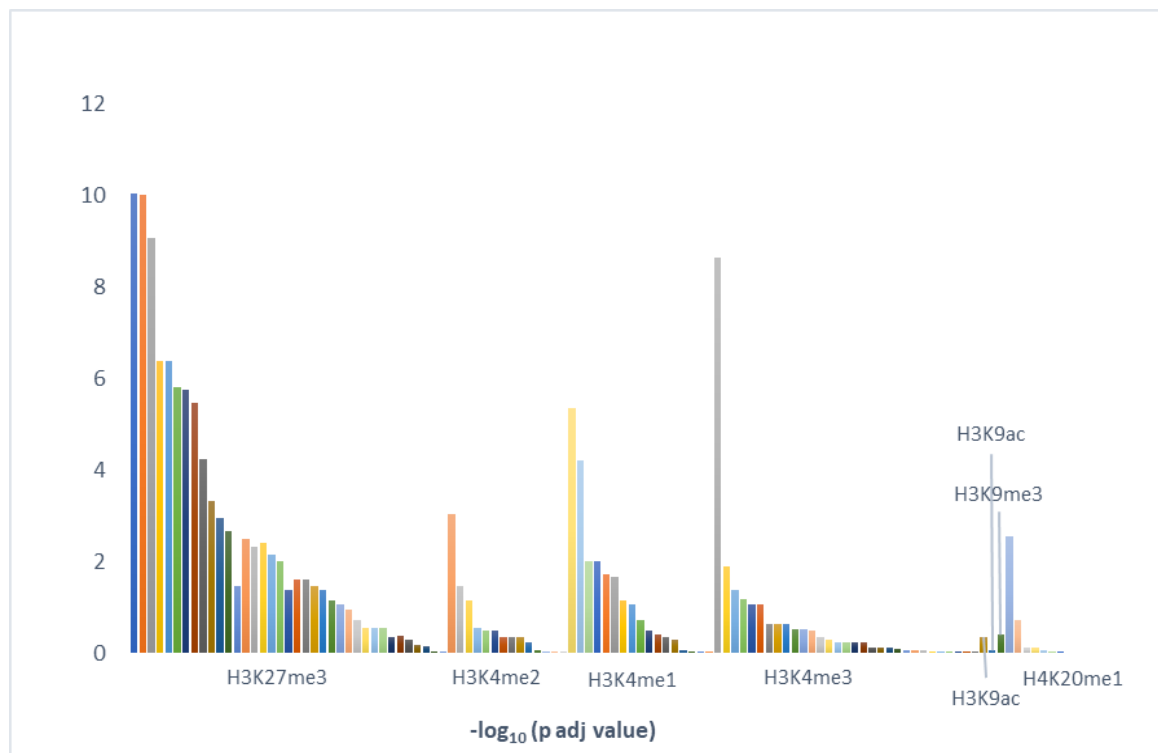
B**C**

Figure 4.6 Enrichment analysis for transcription gene set libraries of the differentially expressed genes in SAHA BPyne dataset using Enrichr (p adj <0.05 and log₂ >0). This represents 812 genes.

(A) TRANSFAC and JASPAR positional weighting matrix used to identify DNA binding motifs for transcription factors. Sp family motifs were disproportionately represented. (B) ENCODE transcription factor ChIP seq EZH2 was overrepresented in the dataset. ZBTB7A, RCOR1, REST, SUZ12 and SP1 are highlighted with less statistical significance. (C) Histone modifications in different human cell lines. ENCODE histone modifications shows the most significant histone modifications were all H3K27me3

Enrichment analysis for transcription factors and histone modifications, profiling the 517 differentially expressed genes, highlighted several interesting features. First, as previously described, when investigating the mechanism by which G6PD activity is induced by HDAC inhibitors, Sp1 plays a critical role (**Figure 4.6A**) (74). Second, searching the ENCODE transcription factor database linked the SAHA BPyne differentially expressed genes with EZH2, the functional enzymatic component of polycomb repressive complex 2. This is intriguing as PRC2-binding sites are enriched in CPG islands (236), the *G6PD* promoter is embedded within a CPG island (**Figure 1.6**) (175).

Finally, the ENCODE histone modifications enrichment datasets (**Figure 4.6C**) indicate that repressive marks, such as H3K27me3, are also important in how SAHA exerts its effect on transcription and pairs well with the transcription factor binding data, as EZH2, is an essential component of PRC2 (237). It will be interesting to compare these data with those gathered from proteomic analysis.

4.4.2 Pathway analysis

To further examine HDAC inhibitor induced change to the transcriptome, evaluation using Reactome, an open source curated and peer-reviewed knowledgebase of biomolecular pathways, was performed on the SAHA BPyne RNA seq dataset. This showed a single

pathway achieved statistical significance, FOXO-mediated transcription ($-\log_{10}$ FDR: 1) with overrepresented genes including cell cycle control genes such as FOXO1.

4.5 THE IMPACT OF SAHA ON CIS REGULATORY ELEMENTS

4.5.1 Introduction

In contrast to protein coding genes, our understanding of cis regulatory elements is limited. Work by Thurman *et al.* established that there is an abundance of proximal regulatory DNA (238, 239). However, the composition of these elements is largely unknown as the technology to study them has not been easily accessible.

Click CHIP seq was developed to examine CREs at points of SAHA binding to determine whether interactions with these DNA sequence elements contribute to the differential expression pattern seen following treatment with HDAC inhibitor. Initial optimisation of the technique with CHIP-PCR was encouraging, showing enrichment for *G6PD* (**Figure 4.6**).

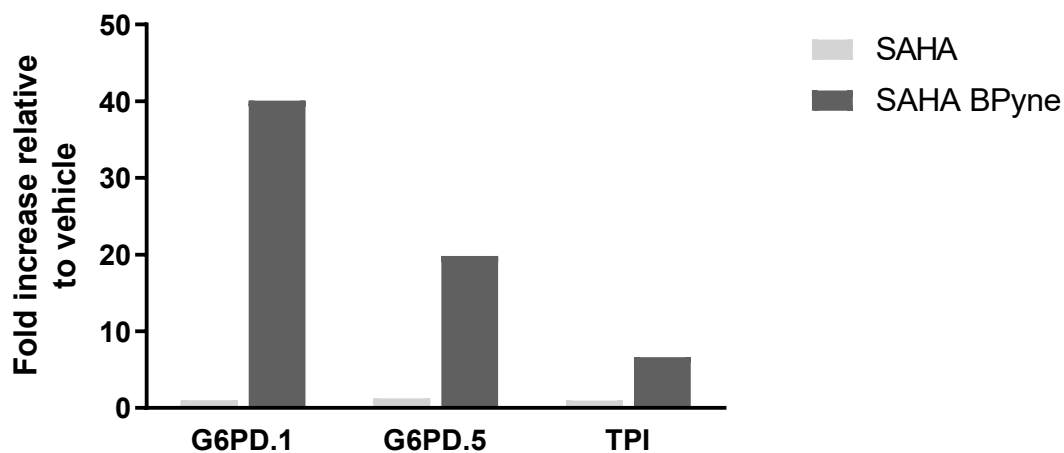


Figure 4.7 RT-PCR to confirm enrichment of *G6PD* expression following Click ChIP pull down.

Illustrating up to 40-fold *G6PD* enrichment, using two amplicons, compared with *TPI* expression, n=1

Enrichment for *G6PD* is seen only with SAHA BPyne which is able to induce *G6PD* expression and pull down *G6PD* gene thereby enrichment is seen up to 40-fold compared with SAHA that is able to induce but not enrich due to lack of pull down tag.

Also noted is that G6PD.1 amplicon shows greater enrichment than the G6PD.5 amplicon as it is closer to the *G6PD* gene

Subsequent experiments were unable to replicate early success. I postulated that a change in the SAHA BPyne formulation was the cause of the inconsistent *G6PD* enrichment. And due to factors beyond my control, namely the COVID 19 pandemic, I was unable to redress the problem, so this element of the project was abandoned.

4.6 POST TRANSLATIONAL MODIFICATION OF NON-HISTONE PROTEIN FOLLOWING TREATMENT WITH SAHA

4.6.1 Introduction

HDAC inhibitors can result not only in the acetylation of histone PTMs but also in the acetylation of non-histone proteins in transcription factor complexes that alter gene expression. Click ChIP general and histone proteomics adopt a novel approach to better understand these interactions.

4.6.2 Experimental design

To identify candidate partners to which SAHA binds, such as chromatin binding proteins and transcription factors, I performed Click ChIP MS with a general proteomics protocol. Here, I was able to examine the protein component of the chromatin pull down.

1×10^8 K562 cells were incubated for four hours with either clickable SAHA – SAHA BPyne, unmodified SAHA or vehicle at a concentration of $10 \mu\text{M}$. This was performed in two biological replicates, employing vehicle and unmodified SAHA as controls. Adequate *G6PD* induction was confirmed using RT qPCR.

The target chromatin was pulled down using streptavidin-mediated affinity purification and washed. After trypsin digestion, the derived peptides were loaded onto a fused silica capillary for reverse-phase high-performance liquid chromatography (240). These isolated peptides were then analysed by tandem mass spectrometry and assigned by database searching against the SwissProt database with the Mascot search engine.

This final processing, mass spectrometry and database search were performed by a collaborator – Dr Harry Whitwell at the National Phenome Centre, Imperial College London – with expertise in proteomics. We adopted a bottom-up proteomic approach in which the proteins were characterised and quantified through analysis of the peptides generated, following trypsinolysis of the chromatin associated proteins (241).

All reported identifications were validated by manual inspection of the LC-MS/MS data.

The Mascot server (version 2.2.04, Matrix Science, London, UK) was used to identify, characterise, and quantify proteins from mass spectral data. The software calculates a peptide score based on the probability of whether an observed match between experimental data and peptide sequences found in a decoy database has occurred by chance. Mascot also estimates the false discovery rate (FDR) by searching these data against a decoy database (242). The data was then normalised, and peaks of interest were identified for further investigation. These peaks were validated by Dr Whitwell.

Following peptide spectra match at an FDR of <4%, a total of 29 proteins were identified.

The proteins were then filtered using the Contaminant Repository for Affinity Purification (CRAPome) database. The CRAPome database is a web-based data resource for the analysis of interaction proteomics data which contains a repository of negative control affinity purifications. This resource stores and annotates negative controls generated by the proteomics research community to filter contaminants, therefore identifying genuine interaction partners (218).

Within the Click ChIP general proteomics dataset ten of thirty proteins were identified as non-nuclear localising proteins and excluded. Seven were identified as likely contaminants

using the CRAPome database and also removed. This list was further refined using the two experimental controls which identified an additional two contaminants which were eliminated, leaving fourteen proteins of interest (**Appendix E Figure S1**).

4.6.3 Pathway analysis

Despite the limited dataset, both protein classification and pathway analysis were highly informative. PantherDB is an open access curated database able to class proteins based on factors such as molecular function, biological processes, protein class and pathway.

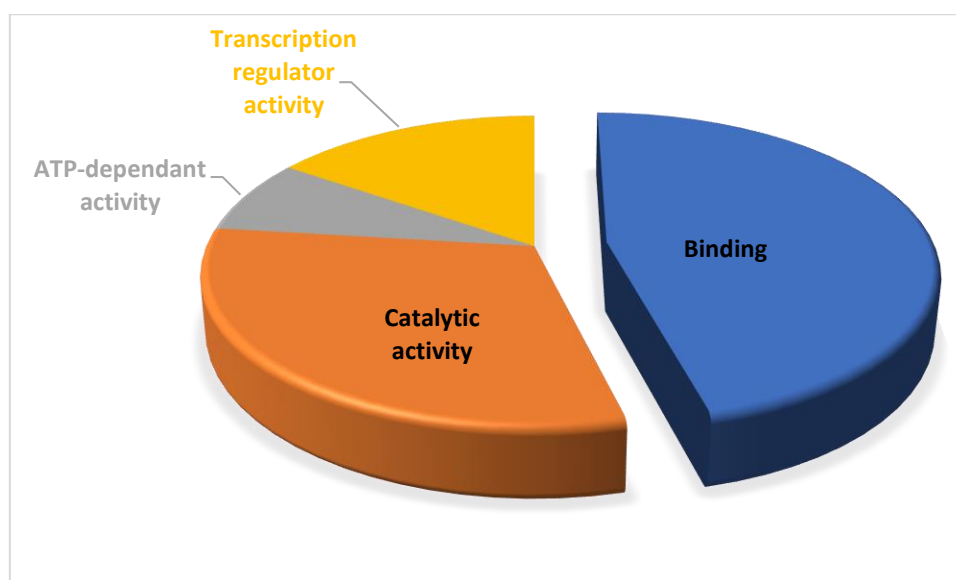


Figure 4.8 selectively expressed proteins in Click ChIP general proteomic experiment - molecular function classification.

Chart shows GO slim molecular function classification of the 25 selectively expressed SAHA BPyne treated Click ChIP general proteomics samples, n=2.

45% of the identified proteins were classed in the 'binding' molecular function classification, this was consistent with RNA seq data. Proteins of interest included STOX3, ZNF438 and ARID5B. These are proteins involved in cellular stress response pathways

Figure 4.9 shows the classification of proteins identified by Click ChIP general proteomics experiment by protein class in the Panther database. Of the nine proteins identified three,

the largest proportion, were grouped within the gene-specific transcriptional regulators protein classes: these included Zinc finger protein 438, Glucocorticoid receptor and AT-rich interactive domain-containing protein 5B.

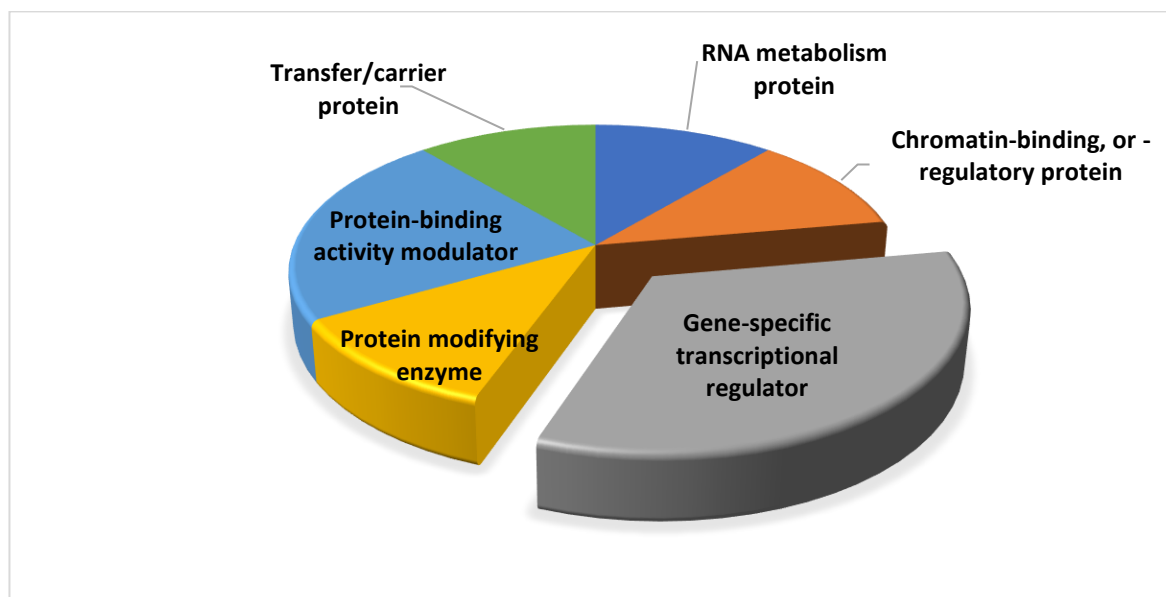


Figure 4.9 Selectively expressed proteins in Click ChIP general proteomic experiment – protein class.

Chart shows classification of proteins identified by Click ChIP general proteomics dataset by protein class. 34% of these proteins selectively expressed by the SAHA BPyne treated Click ChIP general proteomics samples, n=2, were categorised as gene-specific transcriptional regulators.

Of these proteins, three were of particular interest: Glucocorticoid receptor, PCNA-interacting partner and AT-rich interactive domain-containing protein 5B.

Glucocorticoid receptor, a transcription factor which interacts with FOXO1 and has an essential role in cellular stress response pathways (243). FOXO transcription factors have also been implicated in PI3K/AKT activation (244).

PCNA-interacting partner is a protein which plays a vital role in DNA repair pathways (245).

AT-rich interactive domain-containing protein 5B, ARID5B, a transcription coactivator which complexes with PHF2 to mediate demethylation of histone modification H3K9me2 (246).

And finally, there were several RNA interacting proteins – Storkhead-box protein 2 (STOX2), Zinc finger protein 438 (ZNF438), Probable ATP-dependent RNA helicase (DDX31), and again ARID5B.

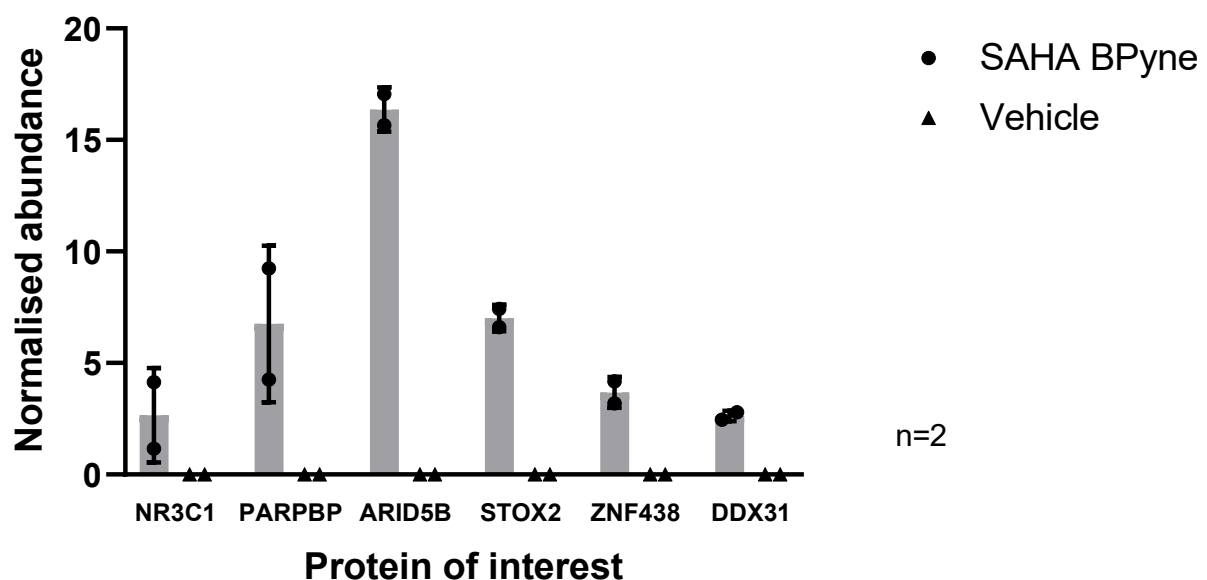


Figure 4.10 Click ChIP general proteomics data comparing SAHA BPyne with vehicle. Graph shows normalised abundance of several protein representatives of oxidative stress pathways substantially enriched compared to control. n=2

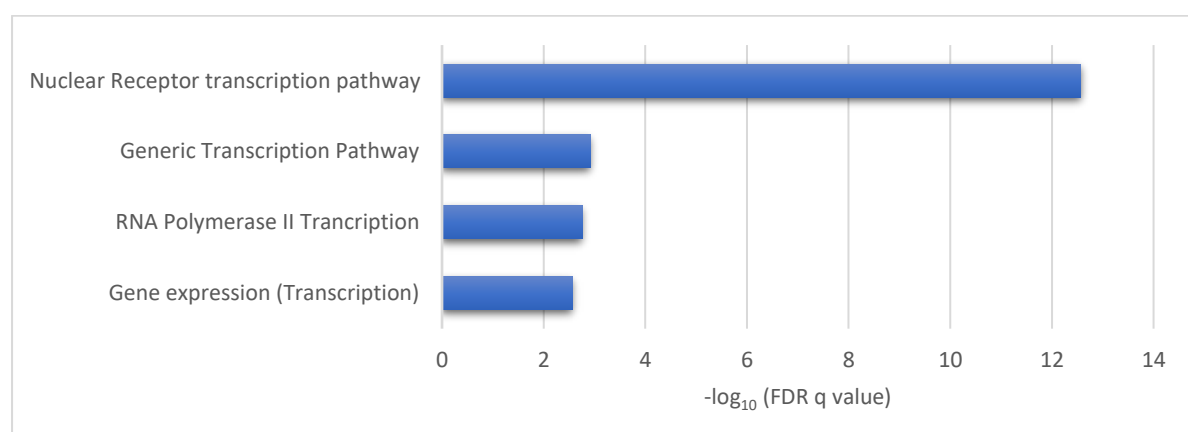


Figure 4.11 Pathway analysis performed on Click ChIP general proteomics dataset. Showing enrichment especially for proteins involved in Nuclear Receptor transcription pathways

Clearly, the Nuclear Receptor transcription pathway is significantly overrepresented in the dataset derived from the Click ChIP for general proteomics experiment. What is of particular interest is that the glucocorticoid receptor, the central protein in this pathway, is a well-established target for HDAC inhibitors.

4.7 EPIGENETIC CHANGES ASSOCIATED WITH SAHA TREATMENT

4.7.1 Introduction

The reversible acetylation of lysine residues on histone tails plays a critical role in transcriptional activation and repression (230) by altering chromatin structure and influencing the binding of remodelling factors, transcription factors, co-activators and co-repressors (247).

Genome-wide studies have revealed that various combinations of histone modifications can lead to the activation or repression of gene expression (248). The most important of which include methylation and acetylation of lysine groups (those that are responsible for gene activation include mono-, di- and tri- methylations such as H4K20me, H2BK5me, and H3K4me3, H3K36me3 and H3K79me3) and acetylation PTMs such as H3K9ac, H3K4ac, H3K14ac and H3K27ac. While methylation of histone H3 at various lysine residues are repressive histone marks, these include H3K9me2, H3K9me3 and H3K27me3 (248-250). Our current comprehension of the interactions responsible for the pharmacological effect of

SAHA are restricted mainly to those that impact malignant disease and even here our understanding of the anti-cancer effect of SAHA is limited.

4.7.2 Experimental design

Here, I used Click ChIP coupled with label-free protein quantitation in a mass spectrometry based method to identify and quantify relative changes in three different conditions. By linking liquid chromatography to the mass spectrometer quantitative values can be determined through extraction of the area of the precursor ions chromatographic peaks or using signal intensity methods by spectral counting of identified tandem mass spectra. For a given protein the spectral counts or protein abundance is directly correlated to the burden of peptide and the greater the number of digests the better the protein sequence coverage.

In brief, 1×10^8 K562 cells were treated with either SAHA BPyne, SAHA or vehicle. After incubation for 4 hours the cellular and nuclear membranes were lysed. Following streptavidin-biotin pull down the proteins underwent stringent washes. The purified histones were derivatised using propionic anhydride to prevent digestion of the unmodified and mono-methylated lysines, the proteins were then trypsinised (251). The digested peptides were quantified using label-free mass spectrometry.

The mass spectrometry and database search were performed through a collaboration with Dr Harry Whitwell. Mascot server (version 2.2.04, Matrix Science, London, UK) was used to identify, characterise and quantify proteins from mass spectral data. I then normalised the data and identified peaks of interest for further investigation.

As described in section 4.6.2, the fundamental steps of label-free quantitative proteomics using LC-MS/MS include sample preparation, sample separation by liquid chromatography and analysis of peptides by tandem mass spectrometry followed by data analysis.

This approach offers several advantages – it is versatile, no need for chemical labelling that could potentially interfere with sample analysis. It is relatively inexpensive and sample preparation time is significantly reduced. However, as samples are not multiplexed run time is increased, sequence coverage is limited, and more starting material is often required.

The pulled down histones were examined with a histone specific protocol design to identify the following histone modifications:

1. Acetylation at lysine and arginine residues.
2. Mono- and dimethylation at arginine residues.
3. Mono-, di- and trimethylation at lysine residues.
4. Oxidation of methionine residues.
5. Fixed propionylation at N-termini, unmodified lysines and monomethyl lysines.

Unlike the general proteomics protocol, the additional derivatisation step, pre-trypsin digestion neutralises the charge at the N-termini, unmodified and monomethylated lysine residues, protecting them from trypsinisation, generating fixed propionylations at these residues.

4.7.3 H3K9 dimethylation

As described in section 4.6, there were surprisingly few peptide modifications identified by this approach. It is difficult to estimate an expected number of peptide peaks for such a

novel technique. However, on reviewing the literature, I would estimate the peptide yield was approximately 10% of expected. The reduced starting material and subsequent reduced number of peaks likely represents an incomplete profile, probably as a consequence of sample loss during streptavidin bead washes.

Of the seven remaining peptides (one of which was a contaminant, excluded by the CRAPome database), two were N-terminal propionylations, processing artefacts. Three more peptides were excluded using the control experimental data.

A single histone modification remained, H3K9me2. There was a greater abundance of this modified peptide in SAHA BPyne samples compared to the methanol control indicating an increase in this PTM following SAHA treatment. This histone PTM is associated with the reversible activation of the mTOR pathway which initiates oxidative stress control pathways (252).

Although, the limited dataset suggests these data must be viewed with caution. The Mascot score of >20 is indicative of a genuine result, suggesting methylation at the 9th lysine residue of histone H3 does have a role in the action of SAHA.

Chapter Five

CRISPR affinity purification *in situ* of regulatory
elements (CAPTURE)

5 CAPTURE

5.1 INTRODUCTION

5.1.1 Background

G6PD deficiency is often characterised by a hypomorphic mutation that results in a loss of function (253, 254). The ‘Karadimitris hypothesis’ states that increased transcriptional output of a mutant gene to produce a defective protein beyond a 20% threshold can restore activity to normal/near normal levels, enough to restore red cell function and homeostasis (255, 256). Experimental work in the Karadimitris lab, initially by Almeida *et al.* and later by Makarona *et al.* established the proof of concept for amplification of abnormal proteins with HDAC inhibitors (74, 257). Interestingly, Makarona’s work demonstrated that HDAC inhibitors result in the selective induction of G6PD activity.

In this chapter, I describe how I adapted the CAPTURE protocol (195) to develop an unbiased, locus-specific approach to decipher the mechanism by which HDAC inhibition selectively upregulates G6PD activity. Using this novel approach I was able to analyse the protein complexes, intricate protein–protein interactions and protein–DNA transcriptional regulatory networks that are key to understanding these complex biological systems.

5.1.2 Current understanding of the epigenetic control of G6PD induction

Pioneering work in the Karadimitris lab established the framework for this project which aims to develop the understanding of SAHA induced *G6PD* induction. This early work noted that the basal transcriptional landscape is similar for all GPPP genes. This is characterised by comparable levels of HAT, HDAC and Sp1 binding, levels of H3 and H4 acetylation, RNA Pol II binding and activity (74). Despite all these similarities, selective induction is observed.

It was postulated that the epigenetic events that underpin the selective responsiveness to HDAC inhibitors could include an increase in H1 and H6, HATs (p300, CBP and GCN5), Sp1 binding, as well as enhanced H3/H4 acetylation and Pol II recruitment to *G6PD* promoter (74). Further work in the lab indicated that H4 acetylation at the core promoter is up to 3-fold greater than H3. Although, this may be misleading as basal H4 acetylation is lower than H3 and, therefore, the fold change is significantly increased compared to H3. It is also possible that H4 acetylation is regulated by an alternative mechanism to that which dictates the acetylation of H3. Interestingly, HDAC inhibition of *TPI* and *GPI* had no impact on H3 and H4 acetylation at the core promoters of either of these genes, **(Figure 5.2)** (74).

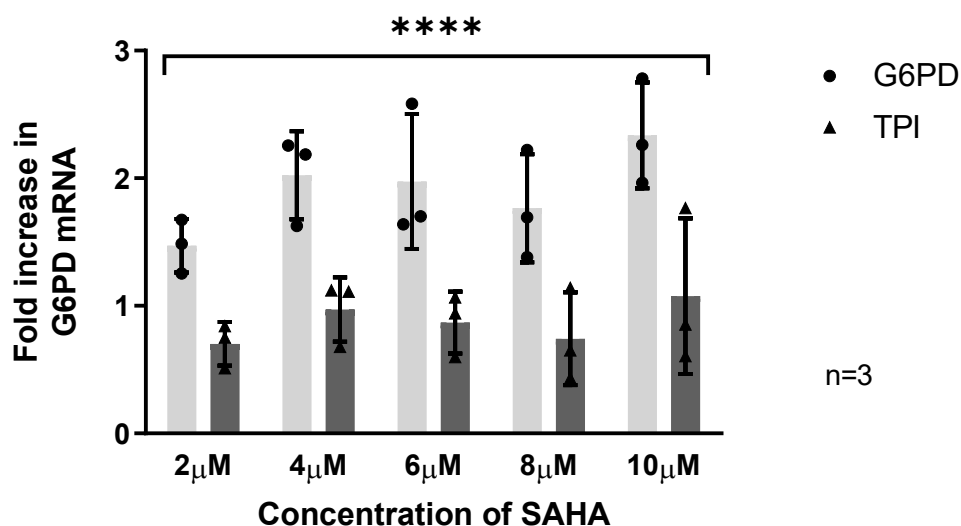


Figure 5.1 Dose response showing selective induction of *G6PD* expression following treatment with varying doses of SAHA at 4 hours. Optimal selective induction identified at 10μM SAHA after 4 hours, greater than 2-fold induction compared to *TPI*. n=2. Two way ANOVA comparing difference between *G6PD* and *TPI* expression, and concentration of SAHA has a p value <0.0001 indicating significant difference between expression of *G6PD* and *TPI* at different SAHA concentrations. Error bars show mean with standard deviation.

5.2 CHANGE IN THE TRANSCRIPTOMIC LANDSCAPE FOLLOWING SAHA TREATMENT

5.2.1 Experimental design

1x10⁶ K562 cells were collected following 4 hour incubation with either SAHA or vehicle.

Two biological replicates were purified, and RNA was processed as per the RNA seq pipeline described in section 2.5.1.

Principle component analysis (**Figure 5.2**) showed significant dissimilarity between the SAHA and control samples.

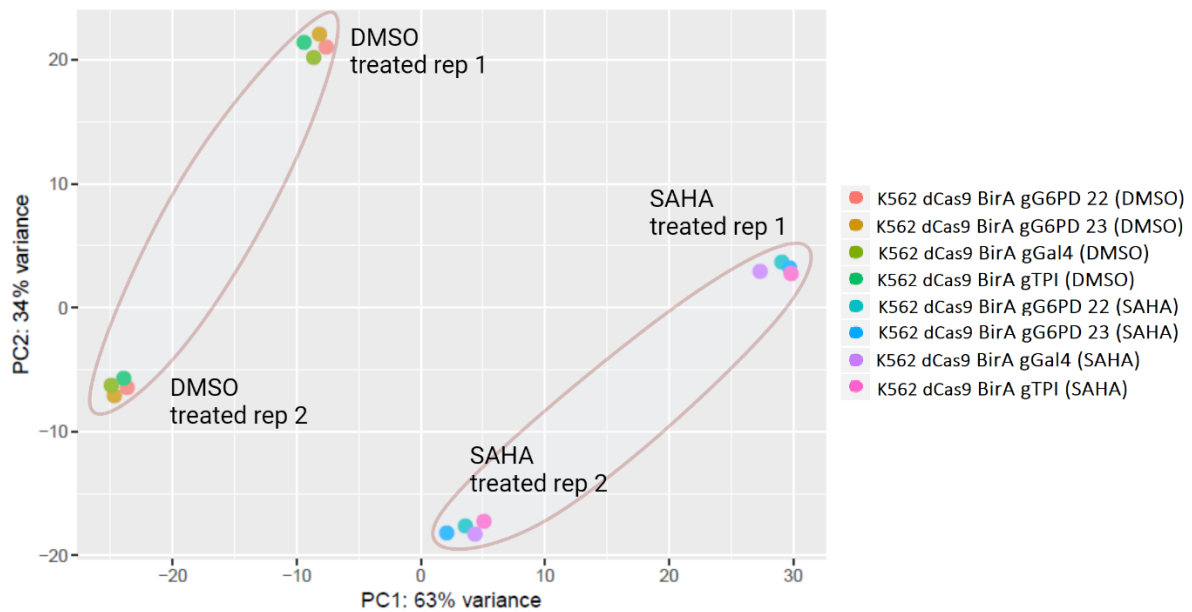


Figure 5.2 Correlation between CAPTURE RNA seq in K562 cells treated with DMSO/SAHA

Principal component analysis plot shows the variance between the mean abundance of the four different K562 CAPURE cell lines treated with DMSO (vehicle) or SAHA. Experiment performed in duplicate, $n=2$. The graph shows the correlation between the different conditions by plotting the principal component (PC) 1, the attribute with the greatest variability, against PC2, the 2nd greatest variable to allow similarities to be identified. Therefore, the difference along the x axis is of greatest significance. There is dissimilarity between SAHA treated and vehicle treated cells. Variation between replicates occurs due to differing *G6PD* induction, though both represent ≥ 2 fold from baseline

The RNA seq data revealed that, of the 1275 differentially expressed genes, 1013 were upregulated and 262 downregulated genes, based on a significance threshold set at an adjusted p value <0.05 and $\log_2$ fold change ≥ 1.5 .

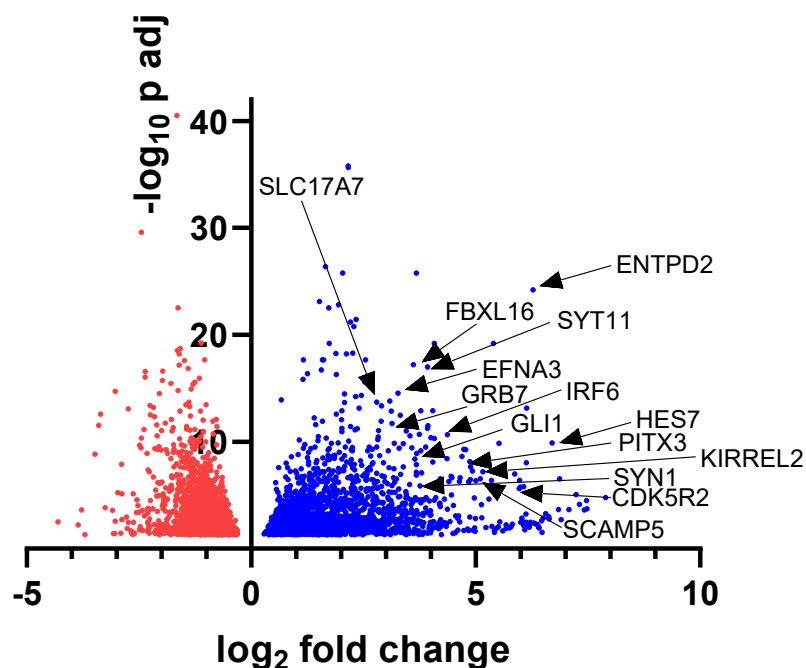


Figure 5.3. Volcano plot shows the expression profiles of SAHA treated cells normalised to DMSO.

Differentially expressed genes of SAHA treated cells when compared to vehicle (p adj <0.05, $\log_2$ fold change ≥ 1.5). Upregulated genes are denoted by red dots and downregulated genes by blue dots. Selected genes are denoted by black arrowheads (p adj >5, $\log_2$ fold change >2)

The RNA seq data shown in **figures 5.2-5.4** illustrates several important points. The first is that transcriptional activity changes significantly following treatment with SAHA. These data also show differential expression of numerous genes involved in cellular stress pathways such as *FOXO1*, *PITX3*, *MYC* and *CTCF*.

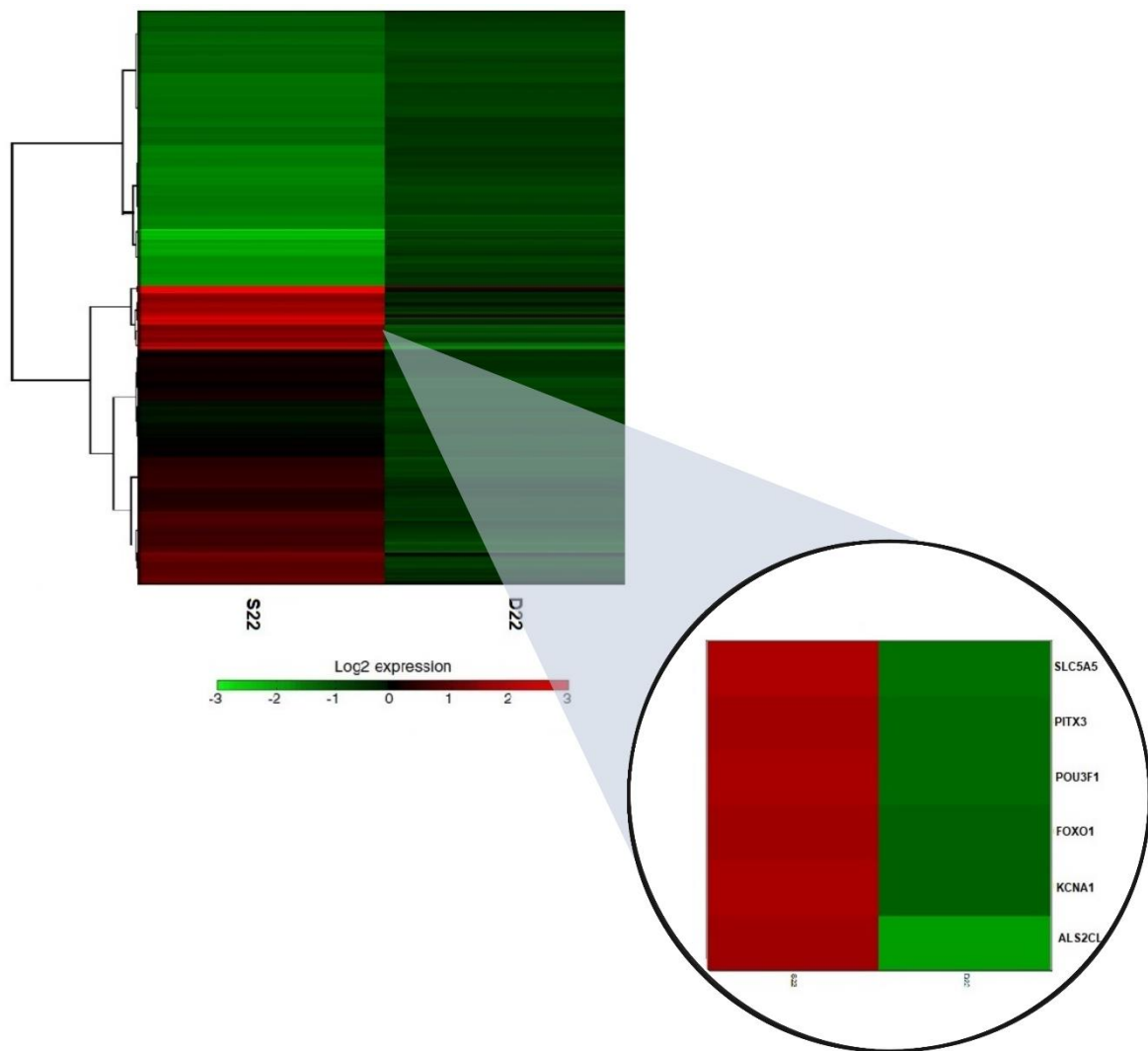


Figure 5.4 (A) Hierarchical clustering of differentially expressed genes in SAHA and vehicle treated K562 cells.

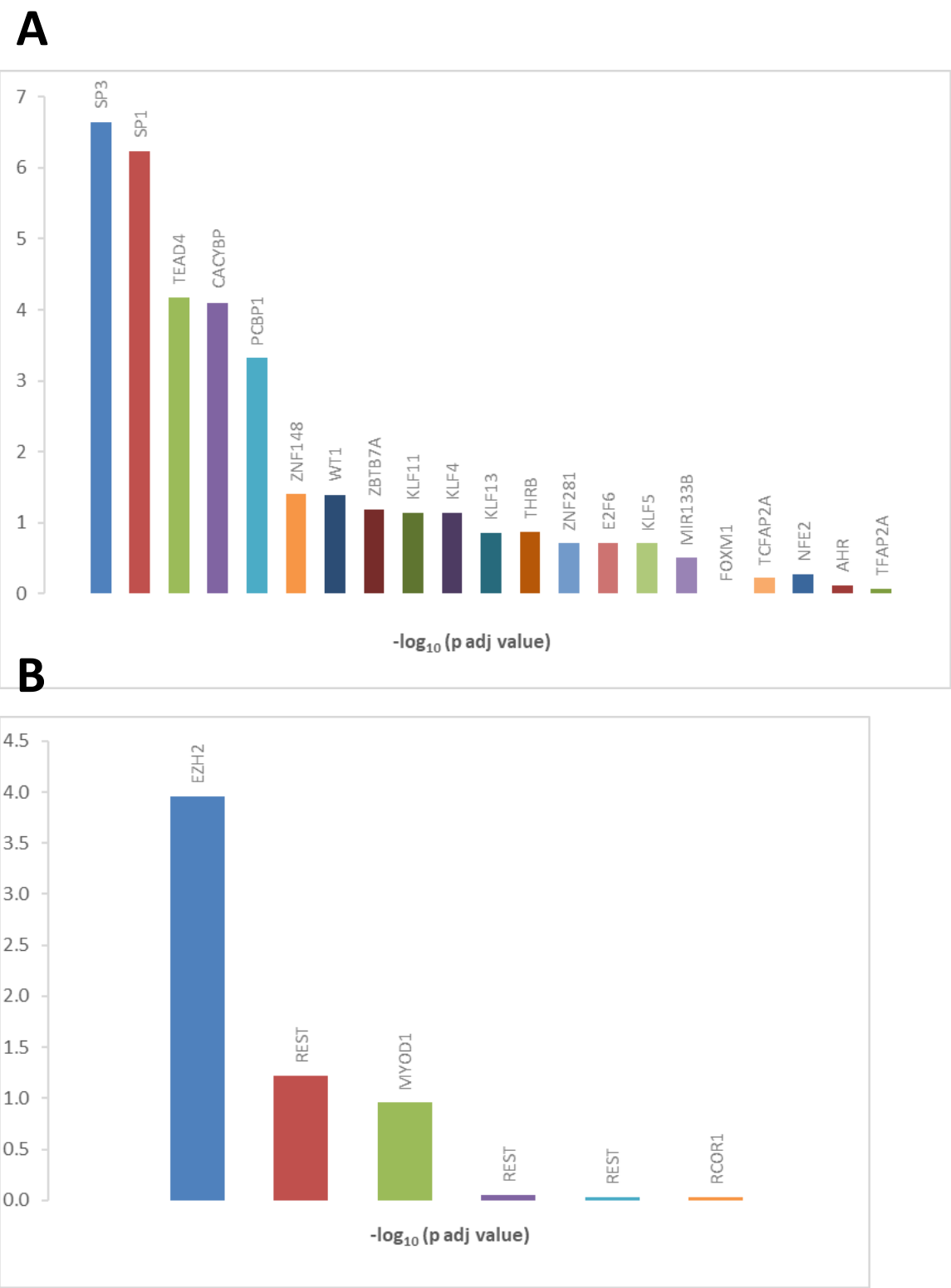
Show significant dissimilarity between the SAHA treated K562 dCas9 BirA gRNA *G6PD* cell line (S22) and DMSO treated K562 dCas9 BirA gRNA *G6PD* cells (D22). The SAHA treated cells show numerous clusters of genes which are overexpressed (red) and under expressed (green) in the control sample.

Dendrogram connecting data points in a hierarchical tree. Hierarchical cluster analysis was performed using Ward's linkage a method that specifies the distance between two clusters as the increase in the sum of squares. When applied with Euclidean distance, the calculated incremental sum of squares is square rooted.

(B) Detailed Heat map showing genes of greatest dissimilarity. The expanded region within centre of the heat map highlights two genes (*FOXO1* and *PITX3*) which are overexpressed across several modalities assessing gene expression

5.2.2 Effect of SAHA on K562 transcriptome

Enrichment analysis of the differentially expressed genes showed an overrepresentation of Sp1 and Sp3 binding motifs, **Figure 5.5**.



C

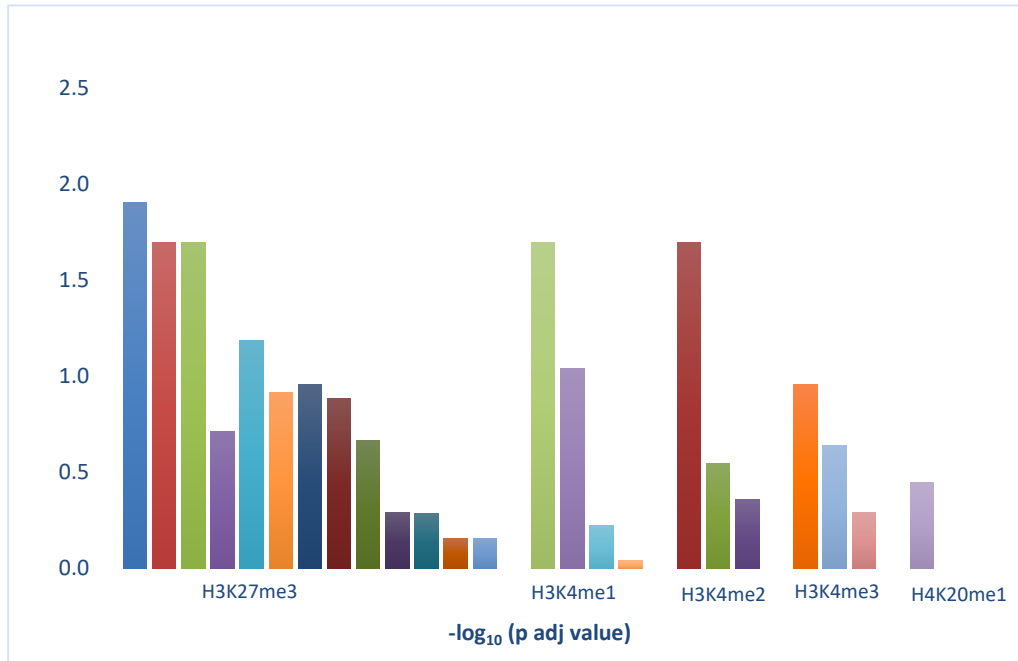


Figure 5.5. Enrichment analysis of differentially expressed genes using Enrichr.

(A) TRANSFAC and JASPAR positional weighting matrix identified overrepresentation of Sp1 and Sp3 transcription factor binding motifs. (B) ENCODE transcription factor ChIP seq shows considerable enrichment for EZH2 compared to the other five genes that overexpression was of statistical significance ($-\log_{10} p \text{ adj} > 1$). REST (RE1 Silencing Transcription Factor) is a Protein Coding gene. Among its related pathways are PIP3 activates AKT signalling.

(C) ENCODE histone modifications shows the most significant histone modifications were all H3K27me3. EZH2 and H3K27me3 are integral components of PRC2 which is involved in control of Sp1 dependant genes

Figure 5.5 A highlights Sp1 binding motifs within the *G6PD* promoter region. Sp1 is a nuclear transcription factor responsible for regulating gene expression through binding motifs within GC-rich promoters (258, 259). Following acetylation, Sp1 may act as a transactivator (260) or a superactivator (261). The transcription factor can form multimers which either

synergistically interact with multiple binding sites (262, 263) or superactivate a single binding site (261). These actions significantly alter transcription of the interacting gene.

These analyses also indicate that the polycomb repressive complex 2 plays a key role in the mechanism by which HDAC inhibitors exert their epigenetic effect. This is illustrated in **Figures 5.5 B and C** as these show that this dataset enriches for EZH2, the catalytic subunit of PRC2, and H3K27me3 (a repressive histone PTM and the primary product of EZH2).

5.2.3 Pathway analysis

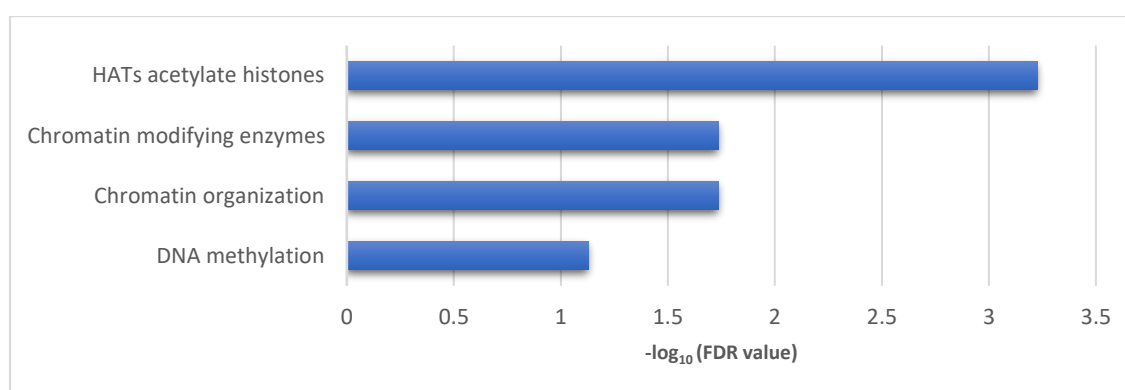


Figure 5.6 Pathway analysis performed on differentially expressed genes following treatment with SAHA shows overrepresentation of chromatin organising, DNA methylation and gene expression pathways

The pathway analysis illustrated in **Figure 5.6** highlights several important pathways.

Chromatin organising pathways were overrepresented within this dataset particularly chromatin modifying enzymes such as HATs. Within these pathways numerous entities were featured - histone proteins, methyltransferases, HATs, RNA polymerase II complexes and polycomb associated proteins.

5.3 THE CONTRIBUTION OF CIS REGULATORY ELEMENTS TO SELECTIVE G6PD

INDUCTION

5.3.1 Introduction

As described in section 4.5.1, ChIP-seq is a powerful tool for the analysis of chromatin modifiers that contribute to epigenetic effects. *De novo* discovery requires an unbiased study of the regulatory regions of genes of interest. Very few studies have addressed this in a locus-specific manner. Currently, these epigenetic control mechanisms, such as transcription binding sites, are not well studied. CAPTURE is a novel and exciting tool to expand our limited understanding of these transcriptional controls.

5.3.2 Experimental design

1×10^8 K562 cells were incubated for 4 hours with either SAHA or vehicle at a concentration of $10 \mu\text{M}$. Following incubation, the samples were cross-linked, and sufficient *G6PD* induction was confirmed by RT-PCR. Streptavidin enrichment purification was performed and after confirmation of enrichment for *G6PD*, sequencing libraries were prepared, and samples were sequenced on the Illumina HiSeq 2500 platform.

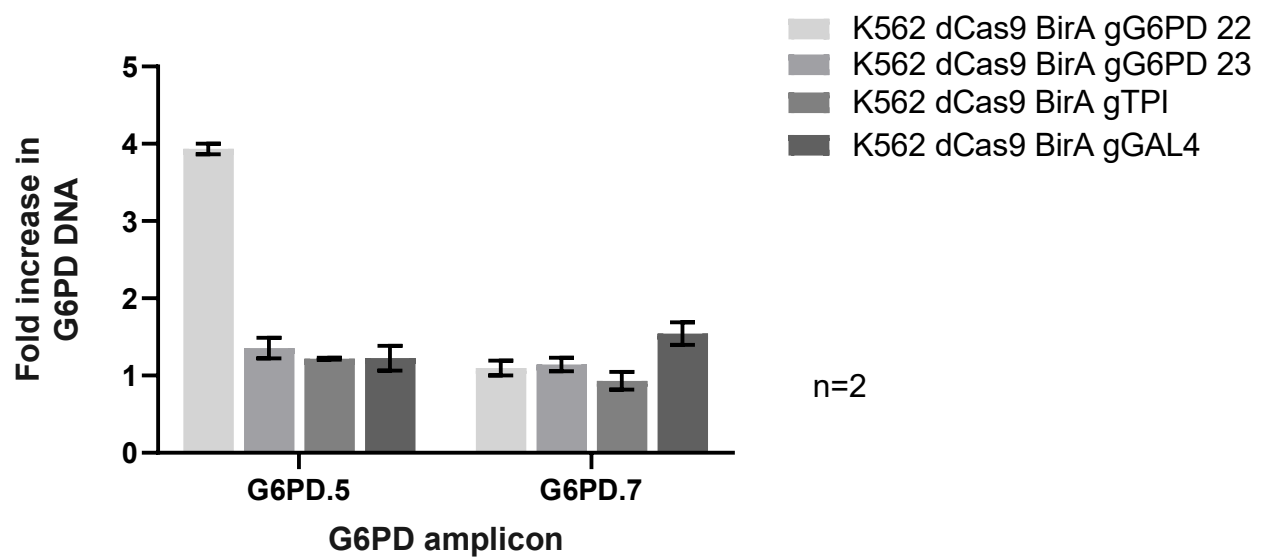


Figure 5.7 RT-PCR performed on purified DNA using CAPTURE seq protocol indicating four times enrichment of *G6PD* DNA compared to controls using the amplicon proximate to the responsive region. n=2

Although, the technical challenges were met, further work to characterise these samples by CAPTURE seq were not performed due to the laboratory restrictions put in place during the COVID-19 pandemic.

5.4 GENERAL PROTEOMIC CHANGES AT G6PD LOCUS FOLLOWING SAHA

TREATMENT

5.4.1 Introduction

To identify non-histone protein interactions, such as transcription factors and chromatin binding proteins that form part of the epigenetic machinery responsible for the selective induction of G6PD by SAHA, I employed CAPTURE proteomics. This technique enables unbiased examination of the protein component of the purified chromatin by mass spectrometry.

5.4.2 Experimental design

For this purpose, large numbers of cells were required (1×10^9 cells per condition) to generate sufficient starting material to ensure quantitative analysis. This involved the mass culture of four CAPTURE cell lines, each expressing a different guide RNA – *G6PD* guide 22, *G6PD* guide 23, *TPI* guide and *Gal4* guide. Each cell line also expressed dCas9 and biotin ligase (**Figure 2.4**). Following incubation with either SAHA or DMSO, the cells were processed as outlined in section 2.5.3. Each condition was performed in duplicate. The biological replicates were TMT labelled and then pooled. The samples were subjected to LC-MS at the Proteomics Facility at the Centre of Excellence for Mass Spectrometry, King's College London (KCL). Analysis of the mass spectrometry output was also performed by the Proteomics Facility at KCL.

The resulting dataset was filtered, as follows: non-nuclear and non-specific proteins were identified using control experiments and CRAPome database, and excluded.

A total of 1614 proteins were returned. After applying the selection criteria described above, 68 proteins of interest remained - 57 upregulated and 11 downregulated.

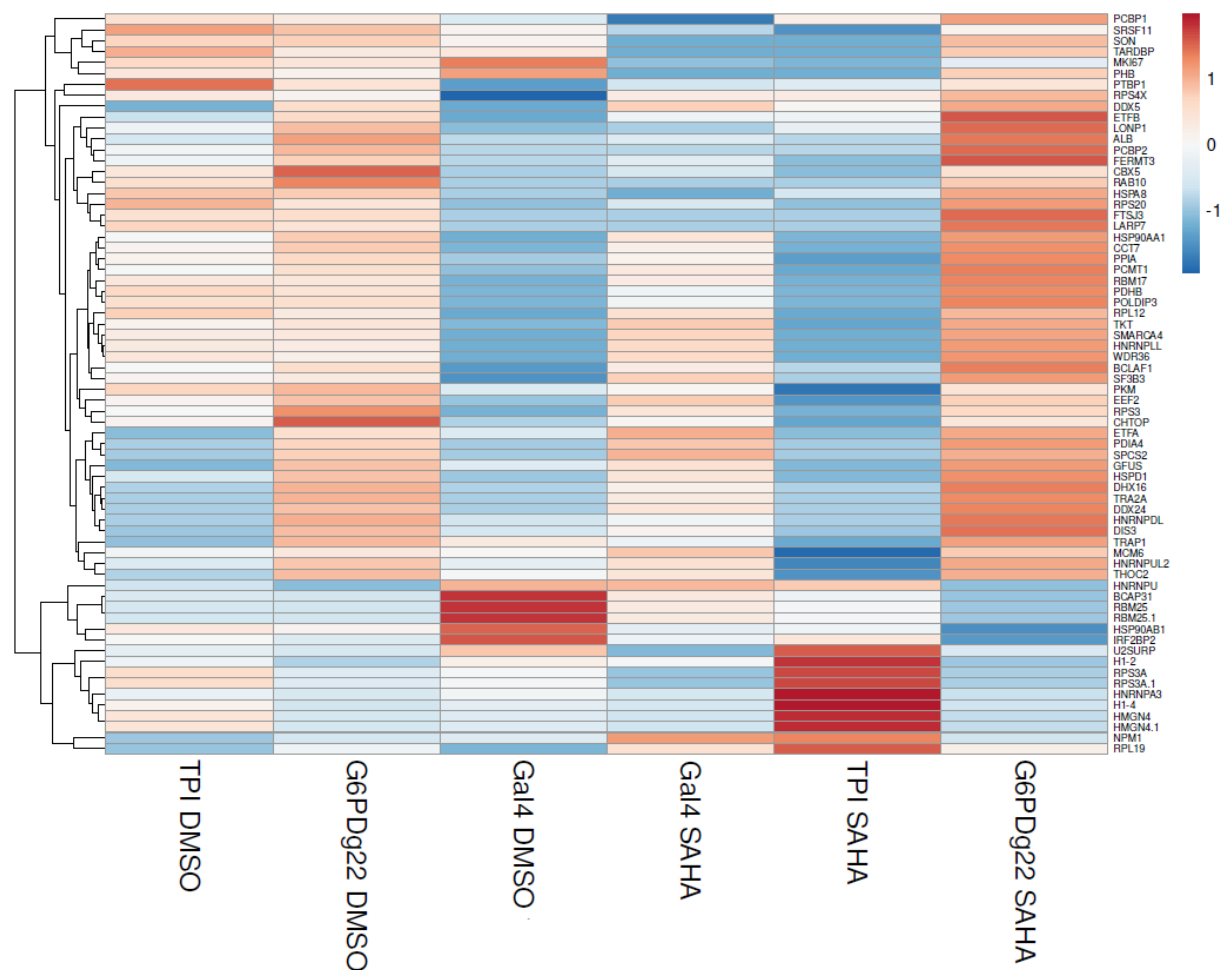


Figure 5.8 Heat map illustrating differential expression of proteins in CAPTURE proteomics dataset

There is a marked difference between the protein abundance in the SAHA treated *TPI* guide cell line and SAHA treated *G6PD* guide cell line. These include histone proteins, histone modifying proteins, heat shock proteins, dead box proteins and RNA modifying proteins (Ward's Linkage using Euclidean Distances of differentially abundant proteins).

These data also illustrate the variability in response to treatment with SAHA in GPPP enzymes. **Figure 5.9** shows the dissimilarity between the SAHA treated K562 dCas9 BirA *G6PD* gRNA expressing cells and SAHA treated *TPI* gRNA expressing cells.

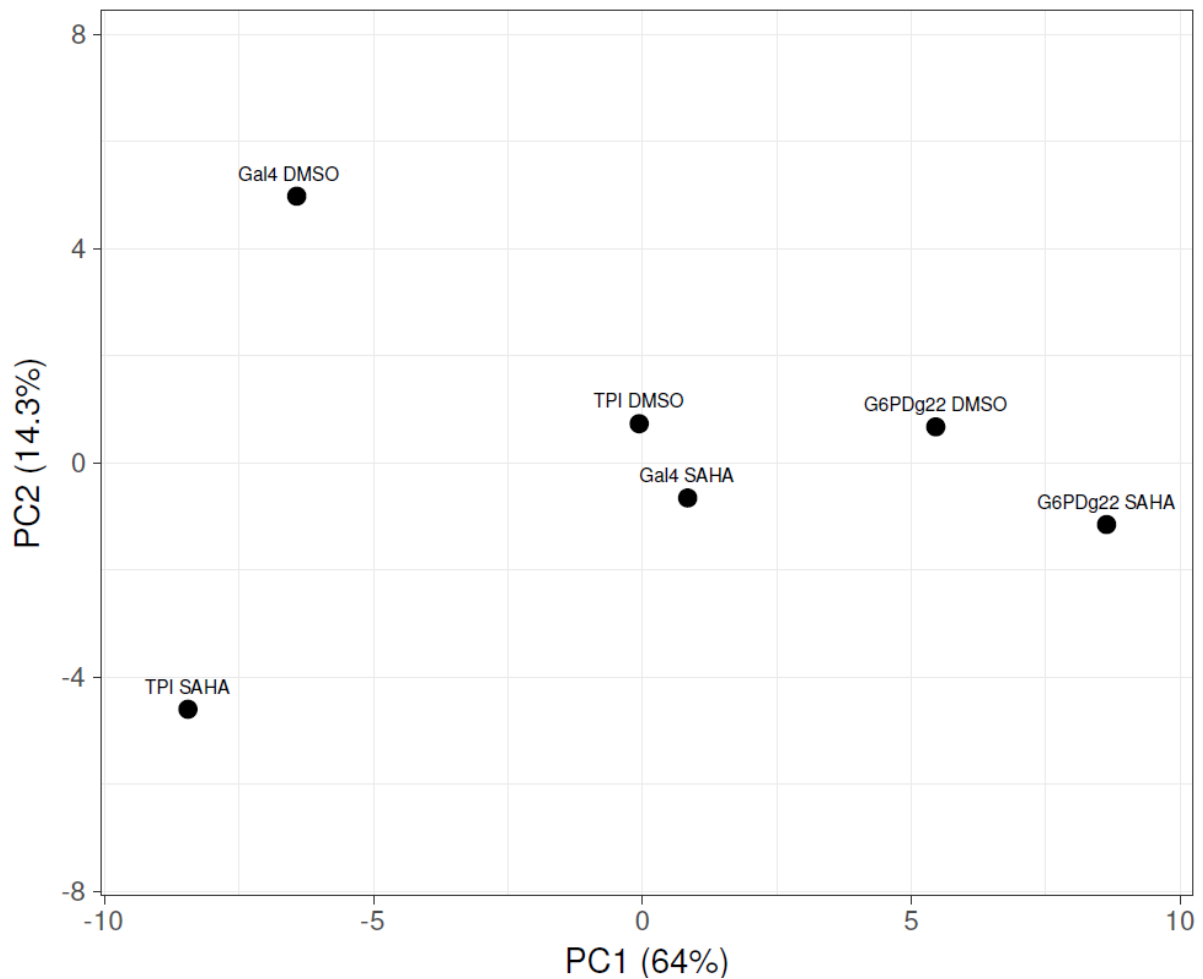


Figure 5.9 Correlation between CAPTURE proteomics conditions in K562 cells treated with DMSO/SAHA

Principal component analysis plot depicting the variance between the mean abundance of three different K562 CAPURE cell lines treated with DMSO (vehicle) or SAHA, and subjected to CAPTURE proteomics protocol. Each data point denotes guide RNA of the cell line and treatment with SAHA/vehicle. The graph shows the correlation between the different conditions by plotting the PC1, the attribute with the greatest variability, against PC2, the 2nd greatest variable to allow similarities to be identified. Therefore, the difference along the x axis is of greatest significance.

Graph illustrates greatest dissimilarity between the SAHA treated *G6PD* gRNA expressing cells and SAHA treated *TPI* gRNA cell line

This was further examined by pathway analysis of the CAPTURE proteomics dataset.

5.4.3 Pathway analysis

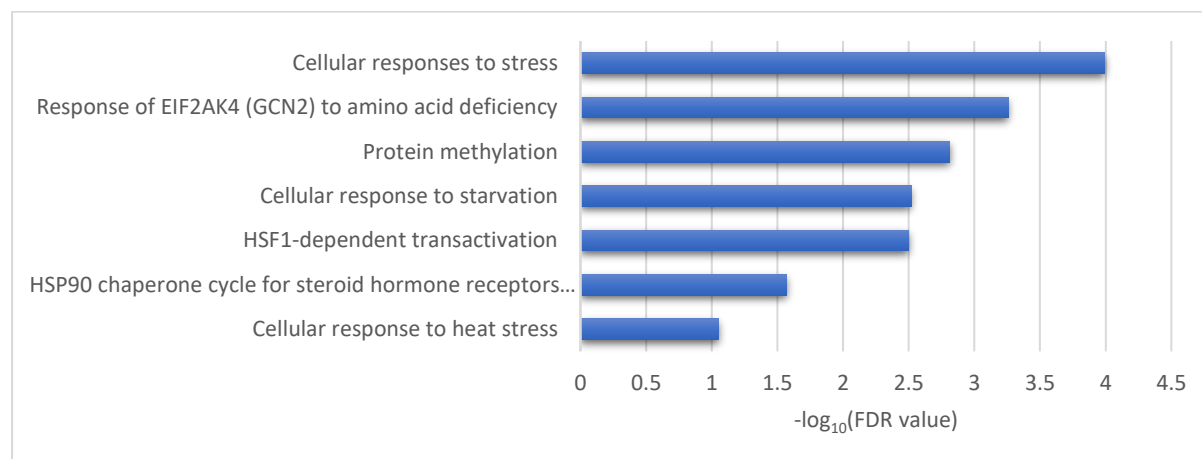


Figure 5.10 shows details of selected pathway analysis of the differentially abundant proteins following CAPTURE ChIP general proteomics. Multiple cellular stress response pathways were overrepresented suggesting that these pathways may play a role in the mechanism by which SAHA induces G6PD expression

Again, cellular stress response pathways were overrepresented, specifically the HSP90 chaperone cycle for steroid hormone receptors (SHR) in the presence of ligand, cellular response to starvation and HSF1-dependant transactivation. Proteins enriched from the CAPTURE proteomic experiment which are represented in these pathways include heat shock proteins such as HSP90 α , HSP60 and HSPA8, ribosomal proteins and histone proteins. Protein methylation pathways were also overrepresented.

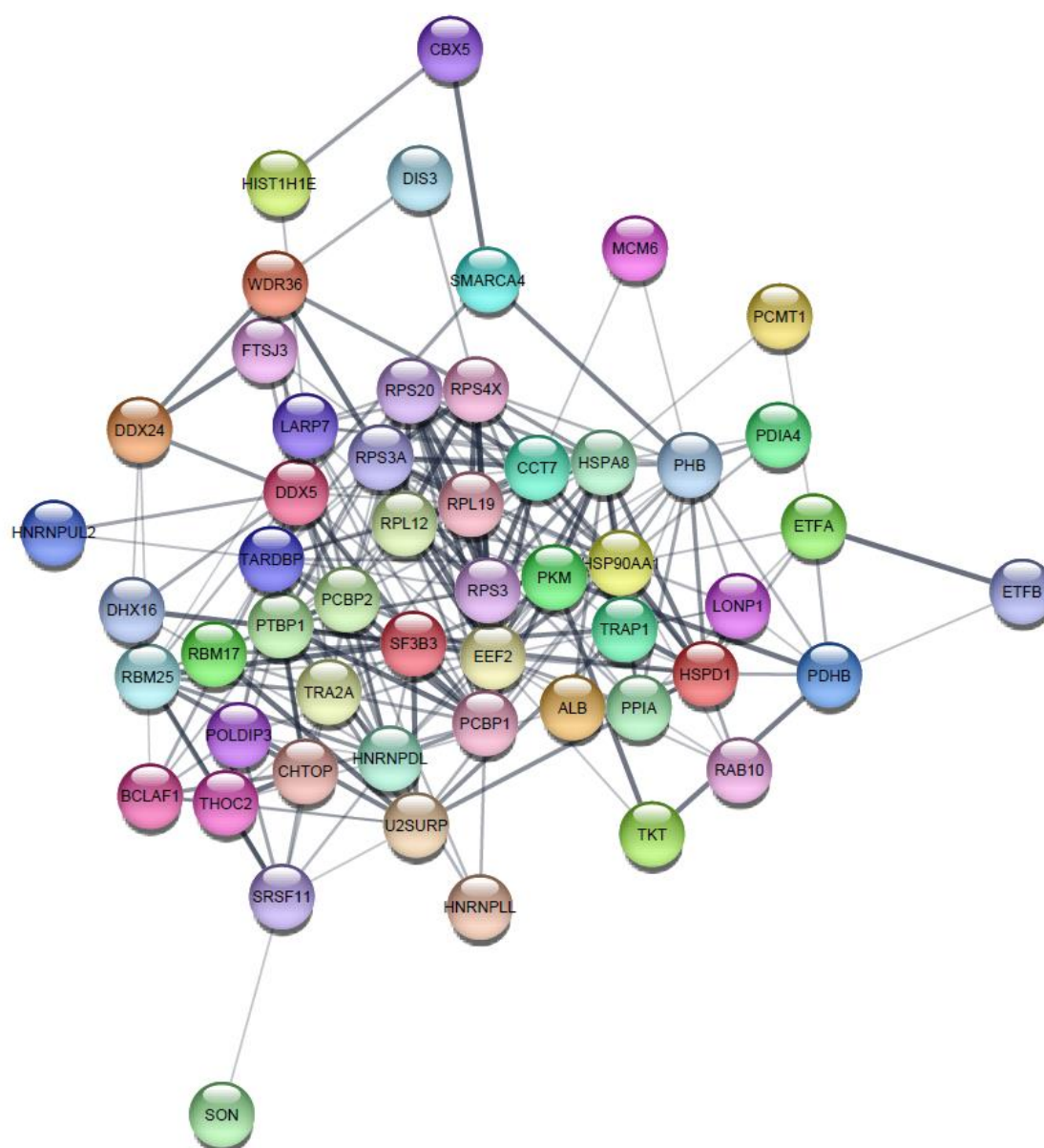


Figure 5.11 String network illustrating the protein interactions that occur following CAPTURE proteomics enrichment. These include interactions between heat shock proteins, DEAD box proteins, histone associated proteins, transcription activator BRG1 (SMARCA4) – a component of SWI/SNF chromatin remodelling complexes, ribosomal proteins and members of the DEAD box family of RNA helicases (DDX5 and DDX24).

5.5 HISTONE POST TRANSLATIONAL CHANGES AT G6PD LOCUS FOLLOWING SAHA TREATMENT

5.5.1 Introduction

As described in section 5.1.3, histone post translational modifications are an essential contributor to the selective response of the GPP pathways when exposed to HDAC inhibitors. Recent publications have suggested that G6PD activity may be induced by specific histone marks, H3K9 di- and trimethylation (252, 264).

In this project, I combined two pioneering molecular techniques, CAPTURE proteomics to purify *G6PD* chromatin in an unbiased manner, followed by label-free quantitative MS analysis of >30 global histone PTMs (240).

5.5.2 Experimental design

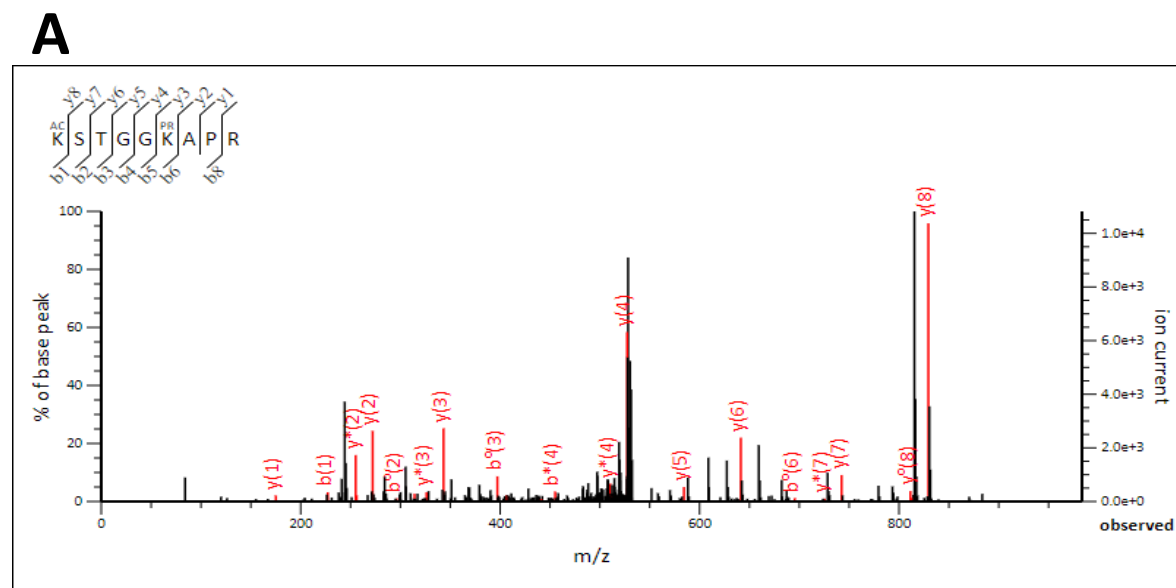
In summary, 1×10^8 of each of the CAPTURE cell lines was treated with either SAHA or vehicle, incubated for 4 hours and cross linked. Following streptavidin-biotin pull down, the proteins underwent stringent washes. The purified proteins were derivatised using propionic anhydride to prevent digestion of the unmodified and mono-methylated lysines, then trypsinised. The digested peptides were quantified using a histone specific, label-free mass spectrometry protocol.

Non-nuclear and non-specific peptides were identified using control experiments and excluded. The remaining 958 peptides were subdivided based on expression profile, identifying 85 over- and 285 under-expressed. These lists were further scrutinised, excluding

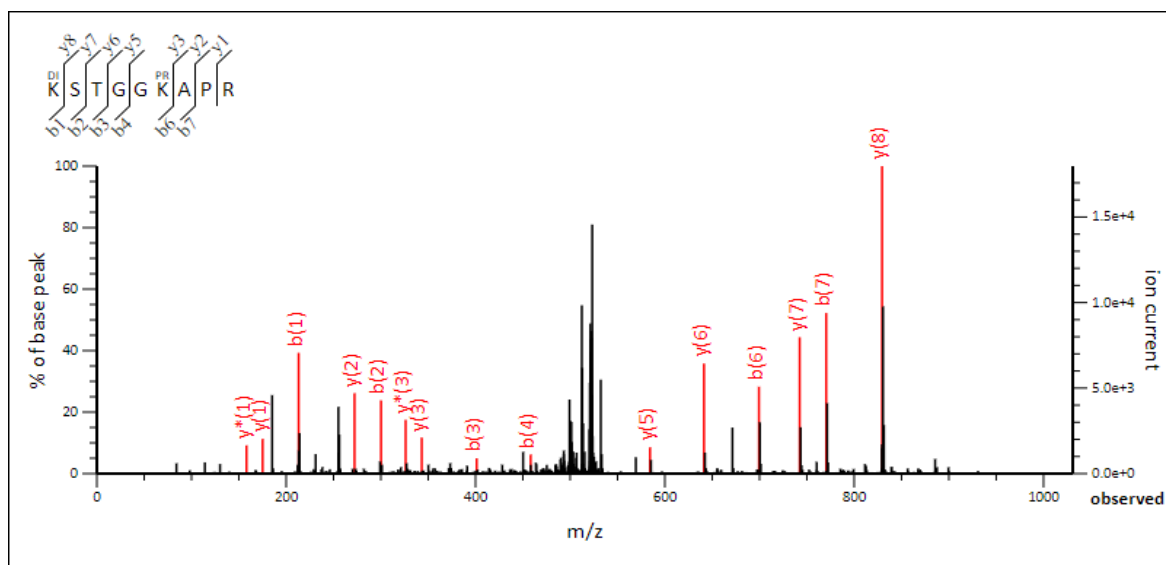
those peptides with only fixed propionylation at N-termini, unmodified lysines and monomethyl lysines. This left 26 upregulated (6 PTMs) and 44 downregulated peptides (20 PTMs), and their histone modifications were annotated manually.

5.5.3 Relative abundance of histone modifications

Here, the approach adopted differed from that used with the general proteomics dataset. Analysis was performed at a peptide level, manually identifying histone PTMs using the protein reconstruction data only to identify which histone the peptide derived.



B



C

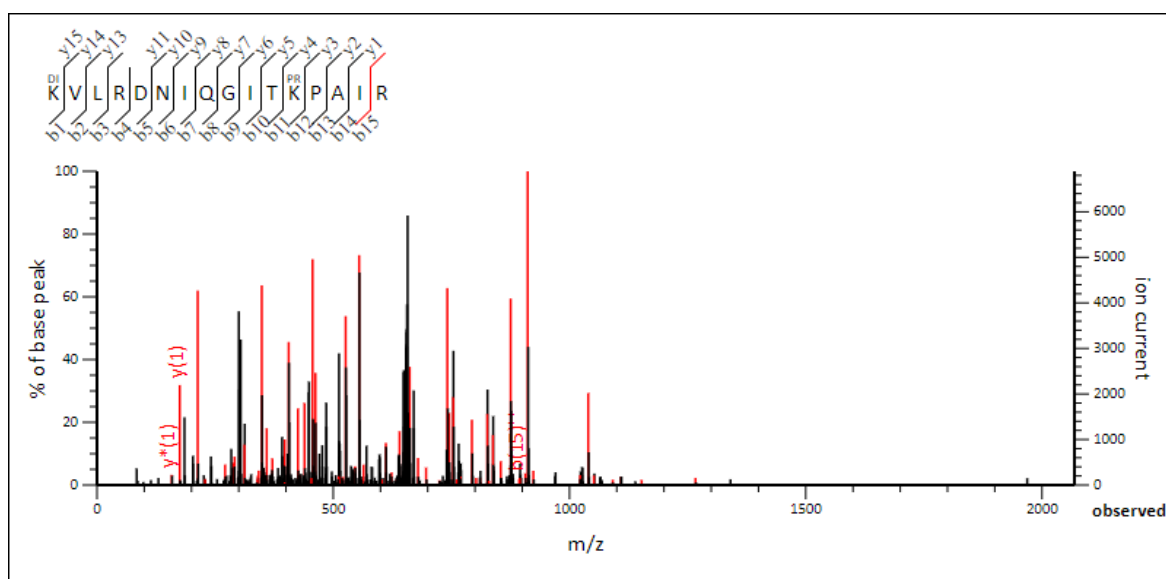


Figure 5.12 Mass spectra of most frequent histones modified affecting histones H3 and H4 in modified K562 cells following treatment with SAHA and purified using the CAPTURE histone proteomics protocol.

(A) H3K9 Acetyl K14 Propionyl. (B) H3K9 Dimethyl K14 Propionyl. (C) H4K20 Dimethyl K32 Propionyl

Fixed modifications and non-histone PTMs were excluded. MS spectra, **figure 5.12**, for the remaining histone modifications were manually validated and the site of the modifications

were identified using UniProt (<https://www.uniprot.org/>); a free to access, manually curated and peer-reviewed database of protein sequence and functional information.

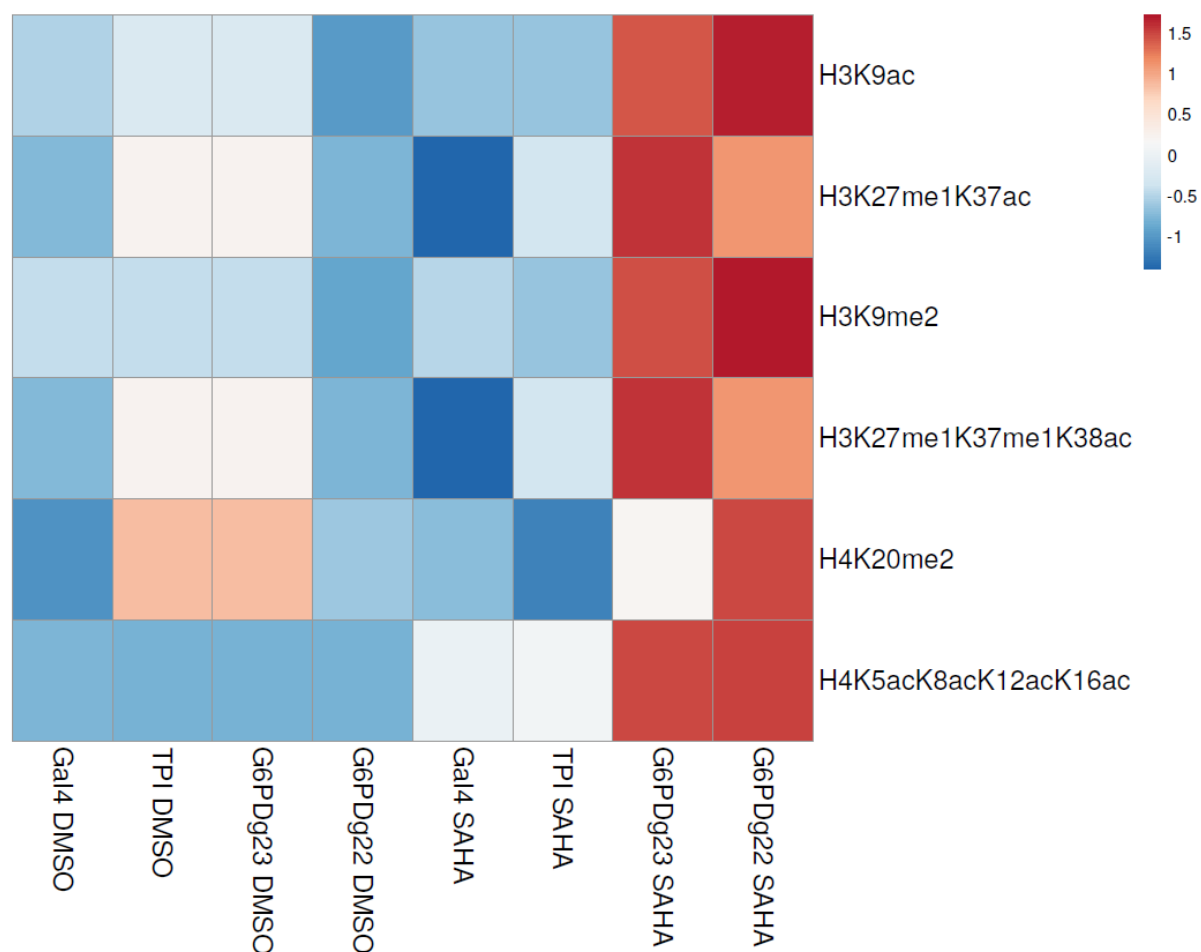


Figure 5.13 Heat maps showing relative abundance of histone modifications identified in CAPTURE histone proteomics experiment. The two cell lines expressing the *G6PD* guides showed increased expression of single and combinatorial PTMs compared to controls

These data highlighted three histone modifications for further examination, in fact these modifications - H3K9ac, H3K9me2 and H4K20me2 accounted for most of the modifications identified in this dataset.

Interestingly, the PCA plot of this dataset (**Figure 5.14**) showed the greatest dissimilarity between *G6PD* gRNA expressing cell lines and those expressing *TPI* gRNA after treatment

with SAHA. Indicating that these data provide some explanation why *G6PD*, unlike other GPPP genes, is induced by SAHA.

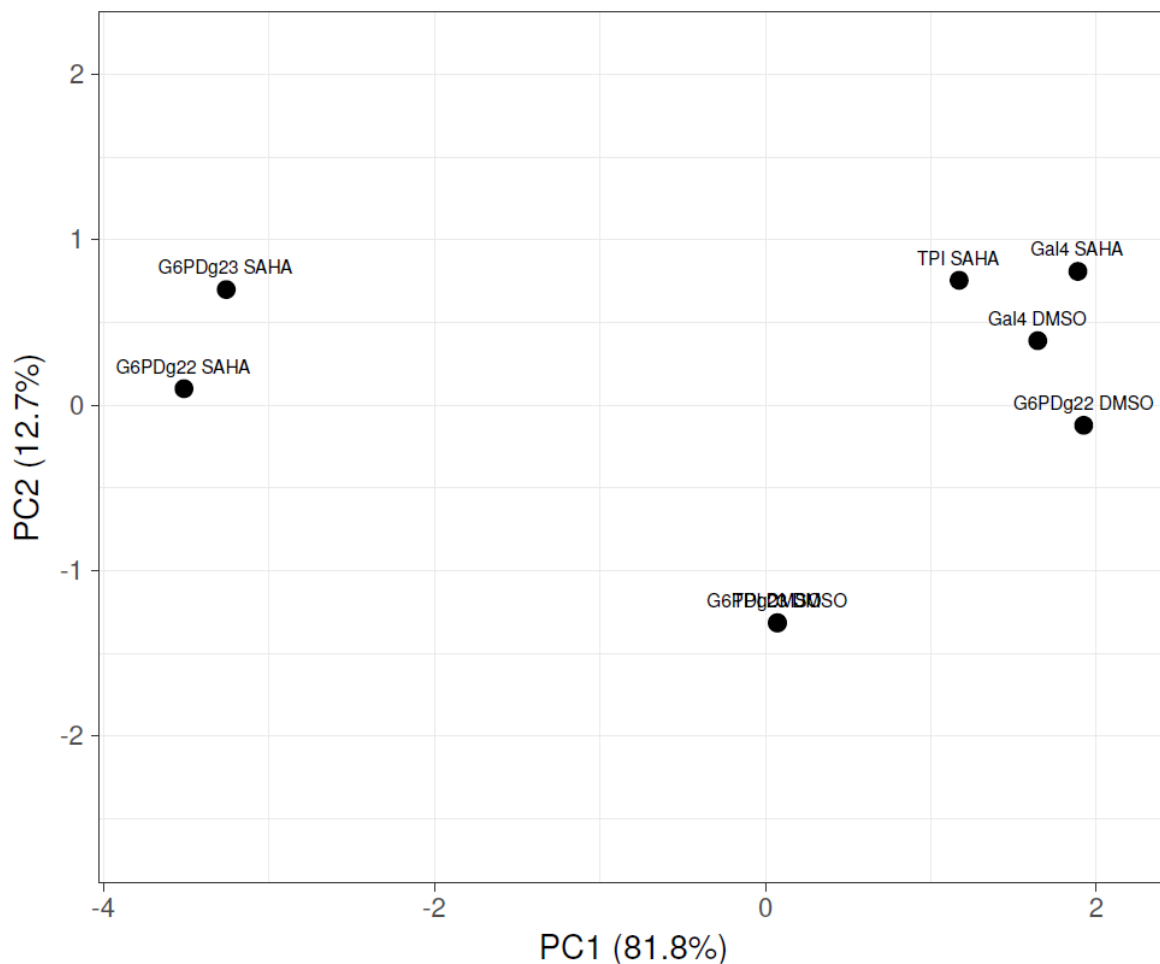


Figure 5.14 Correlation between CAPTURE histone proteomics conditions in K562 cells treated with DMSO/SAHA

Principal component analysis plot depicting the variance between the mean abundance of the four different K562 constructs each treated with either DMSO and SAHA, and subjected to CAPTURE histone proteomics protocol. SAHA treated *G6PD* gRNA expressing cells are dissimilar to other cell lines in all conditions

This was further assessed in **Figure 5.11** which compared modifications of histone marks, H3K9 and H4K20 across different genes.

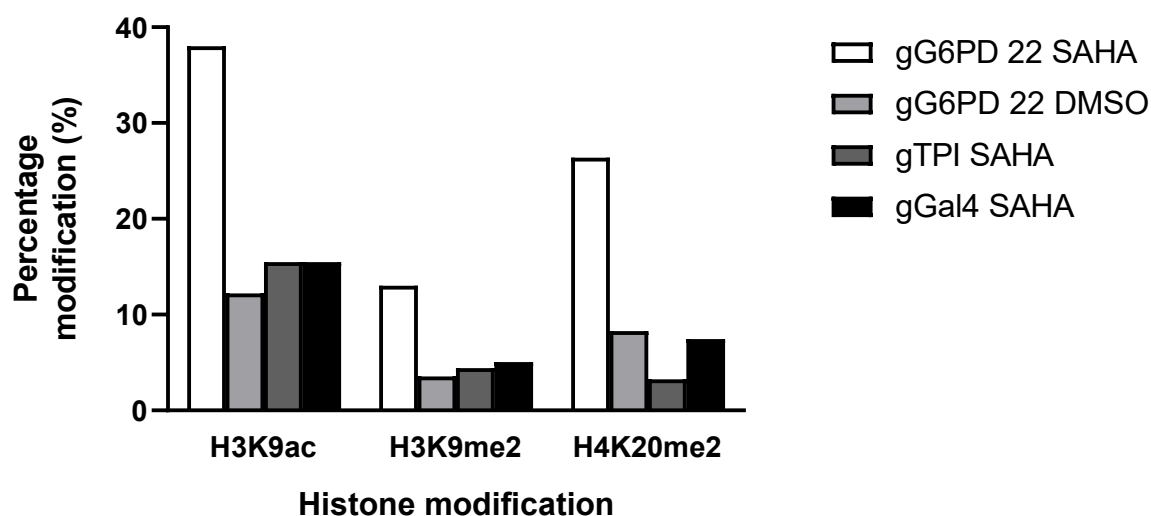


Figure 5.15 shows the proportion of lysine residues modified in different CAPTURE cell lines following treatment with SAHA

After treatment with SAHA over 38% of histone H3 lysine residues at the 9th lysine residue were acetylated and 13% were dimethylated in K562 dCas9 BirA *gG6PD* 22 cells. These percentages fall to 12% and 3.5% respectively when the same cell line is treated with vehicle. And when the two control cell lines, K562 dCas9 BirA *gTPI* and K562 dCas9 BirA *gGal4*, were treated with SAHA there was a similar proportion of post translational modifications at this lysine residue

At histone H4 K20 there were greater than 3x more dimethylation in the K562 dCas9 BirA *gG6PD* 22 cells following treatment with drug compared with the controls.

Figure 5.15 shows several key features. Firstly, in the *G6PD* targeting cell line, acetylation at the 9th lysine residue of histone H3 increases almost 4-fold to nearly 40% following HDAC inhibition. Interestingly, acetylation of the non-inducible GPPP enzyme, *TPI*, remained at baseline. Secondly, dimethylation of the same histone mark also increases about 4-fold, albeit from a lower baseline to 12%. This demonstrates that this residue may exist in a poised state with both activating and repressive PTMs. Finally, H4K20 dimethylation also

increases at more than 3 times baseline following treatment with SAHA and that this change is more than 2 times the abundance than that of *TPI*.

These observations at the *G6PD* locus but not at *TPI* indicates that these PTMs likely contribute to the selective induction of *G6PD* activity following HDAC inhibition.

Chapter Six

Discussion

6 DISCUSSION

The success of this project was dependant on establishing two complementary techniques which I proposed would provide independent yet complementary data to explain the selective induction of *G6PD* activity with SAHA.

As the project progressed, the two primary techniques evolved from Chem ChIP and enChIP to Click ChIP and CAPTURE.

The fundamental challenge with Chem ChIP was the difficulty phenocopying the selective induction of *G6PD* activity seen in SAHA treated cells using biotinylated SAHA.

To explore how the modifications to SAHA impacted activity *in vivo*, a collaborator evaluated the HDAC inhibitory effect of Bio SAHA compared with the parent drug. *In vitro* IC₅₀ studies showed Bio SAHA was a highly effective inhibitor of HDAC activity, twice the potency of the parental compound (**Figure 3.14A**). However, despite this encouraging result, *in vivo* induction experiments did not show adequate *G6PD* induction. Has the modification impacted the ability of the drug to enter the nucleus and/or its ability to enter the tubular pocket of the active site? SAHA is a small molecule which exerts its inhibitory effect by interacting with the active site of Zn²⁺-dependent HDAC. The drug sits within a tubular pocket at one end, and at the other a cap blocks the active site by interacting with the external surface inhibiting deacetylase activity, **Figure 1.7**, (187, 188).

The seminal paper by Anders *et al.* which describes the Chem ChIP protocol, also cautioned that modifications necessary to facilitate the technique could impair important interactions, particularly when the receptor interaction is dependent on size and orientation of ligand

(193). This group suggested that the design of the biotinylated probe must be informed by structural data such as X-ray structure. Although these modalities were not available to me to interrogate the interaction, information about the interaction of the parent molecule and target receptor did offer some insight. The addition of the large biotin moiety likely altered the pharmacodynamics of the drug. As a consequence of this, I was unable to increase *G6PD* activity by more than two times baseline, optimal induction, **Figure 3.14B**.

To limit the impact of the affinity tag, I abandoned Chem CHIP and adopted a slightly different approach which utilised Click chemistry. Here, the affinity tag is considerably smaller than its predecessor, reducing the likelihood of steric hindrance, **Figure 3.15B**.

This technique, Click CHIP (198), requires the majority of the pull down machinery to be added *ex-vivo*, i.e., after cells are lysed thus limiting the impact of the tag on pharmacodynamics. An added benefit is that, unlike Bio SAHA, the 'clickable SAHA' – SAHA BPyne is a proprietary product with an abundance of previously published functional data (199, 265). Although this change did enable a greater than two times induction of *G6PD* expression, this was only achieved with a concomitant increase in other GPPP enzymes, **Figure 3.14A**.

Given the transduction issues highlighted in section 3.2.1, I decided to abandon the GM12878 cell line and replace it with another tier 1 ENCODE cell line, K562. K562 is an erythroleukaemia cell line produced from a female patient with CML. At this point, I had already successfully transduced the GM12878 cell line with the CAPTURE constructs, so I had to replicate this in K562. The use of a single biological model ensured that data from both arms of the project were directly comparable.

Fortunately, the implementation of these measures – the switch from GM12878 to K562 cell line and from Chem ChIP to Click ChIP protocol – enabled the phenocopying of parental SAHA with SAHA BPyne. This was achieved using a dose of 10 μ M of SAHA BPyne for 4 hours, conditions identical to the parental drug, **Figure 3.16**.

The implementation of the Click ChIP protocol enabled sophisticated interrogation of the interaction between SAHA and the transcriptional machinery. However, it was vital to confirm that the addition of the affinity tag did not impact on gene expression.

Chapter 4 examined the Click ChIP experiments. Here SAHA BPyne treated cells are compared with those treated with SAHA followed by transcriptomic and proteomic analysis to determine the impact of SAHA on cellular function to better understand how this drug exerts its epigenetic effect.

The RNA seq experiments were essential to establishing that the addition of a benzophenone crosslinker and an alkyne tag did not impact the drug's effect on gene expression.

RNA seq demonstrated, on a global scale, the similarities in the transcriptomic profile between SAHA and SAHA BPyne. Both the heat map (**Figure 4.2**) and PCA plot (**Figure 4.3**) confirmed similar transcriptomic effects of the two drugs which was encouraging for further work using the tagged SAHA as a surrogate for the parent drug.

Analysis of the top 100 upregulated genes highlighted *FOXO1* and *PITX3* as genes of interest. These transcription factors are both involved in oxidative stress response pathways.

FOXO1 is one of a family of FOXO transcription factors that have a diversity of function. These include the regulation of differentiation, metabolism, proliferation and survival in response to environmental changes, such as growth factor deprivation, starvation and oxidative stress (266). Their transcriptional output is regulated by a two-tiered mechanism of acetylation and phosphorylation. Acetylation occurs via histone acetyltransferase which regulates the function of FOXO1 by altering the affinity with the target DNA and the sensitivity for phosphorylation (267). CBP/P300 acetylates FOXO1 followed by phosphorylation initiated by AKT which, in turn, activates the transcription factor (268). The activated FOXO1 binds to the promoter, mediating transcription of oxidative stress response genes (269). Interestingly, the aforementioned AKT also forms part of the Ras driven signalling pathway that stimulates the pentose phosphate pathway to neutralise ROS (270). Similarly, *PITX3* encodes for a transcription regulator Pituitary homeobox 3 which is involved in neurological development and maintenance (271). It achieves this with a transcriptional co-repressor NCOR2/SMRT that recruits HDAC to the promoter region of the target gene, maintaining it in a repressive state (272). There is also increasing evidence that *PITX3* is a key regulator of intracellular redox state (273).

These examples inform the notion that HDAC/HAT play a vital role in maintaining the oxidative equilibrium.

The enrichment analysis performed on this dataset revealed interesting and pertinent results. The first is that Sp1 and Sp3 are overrepresented. These are zinc finger transcription factors that bind GC-rich motifs of gene promoters. It is well established that promoters of butyrate-responsive genes have butyrate response elements and these elements are often controlled by Sp1/Sp3 transcription factors (274). These transcription factors recruit HATs

and HDACs to alter the acetylation state of the associated gene acting as activator or repressor. The enrichment data reinforces the fact that the action of SAHA involves Sp1, and to a lesser extent, Sp3.

Next, the enrichment analysis illustrated the role of PRC2 in the action of SAHA. Several essential features of polycomb activity, EZH2 and H3K27 trimethylation, feature heavily in the enrichment analysis. Enhancer of zeste homolog 2 (EZH2) is a histone methyltransferase responsible for the methylation of histones H3K27, and to a lesser extent H3K9 (275). These histone modifications are responsible for the repressive effect of PRC2 on transcription (237).

Finally, despite the limited dataset the Click CHIP histone proteomics indicated that H3K9 methylation is a potential mechanism by which SAHA induces cellular stress response mechanisms. Recent publications by Lu *et al.* and Chang *et al.* suggest that this histone modification can induce cellular stress control mechanisms (252, 264). Clearly, more work is required to fully understand these mechanisms.

Taken together, these data indicate that SAHA, at least in part, works by stimulating cellular stress response pathways.

Chapter 5 highlights the importance of non-coding regions in the pathology of Mendelian disorders, such as G6PD deficiency, and how increased insight into their mechanisms will aid the development of more effective therapies.

As anticipated, the data from the CAPTURE proteomics complemented the data obtained from the Click ChIP experiments. Several key features have been highlighted across both modalities. The first is the hypothesis that SAHA stimulates cellular response pathways. This is evidenced by the overrepresentation in the RNA sequencing dataset genes involved in cellular stress response pathways.

These pathways respond to internal and external molecular, and physical stresses. The ability of cells to modulate these stresses is critical to maintaining cellular homeostasis (276, 277).

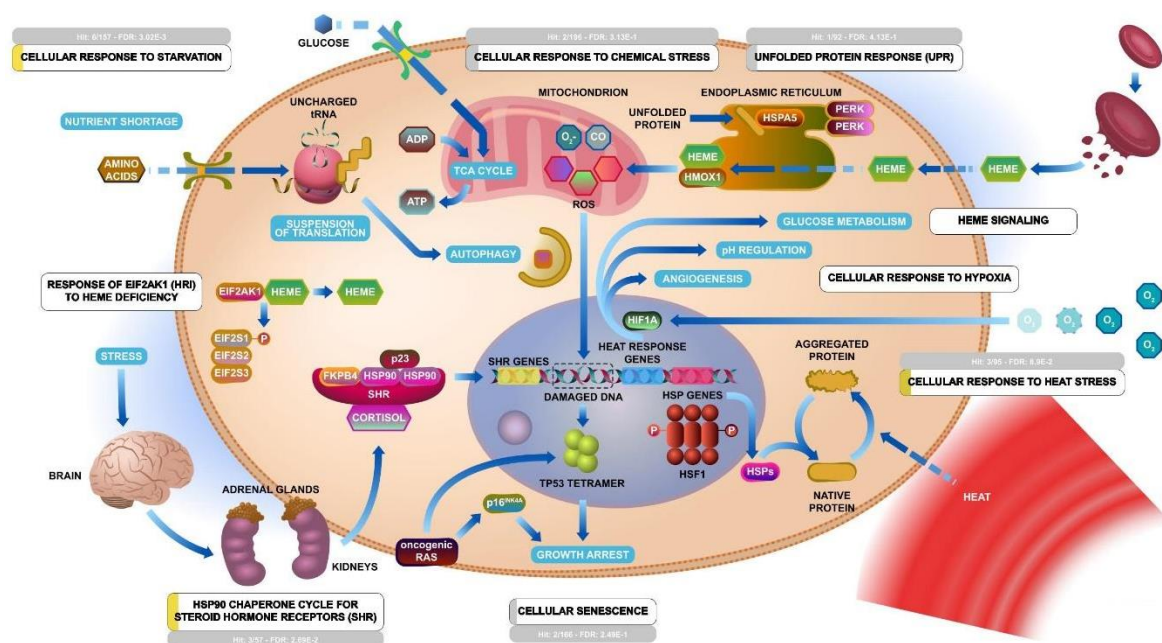


Figure 6.1 details the relationship between various cellular stress response pathways and how important interactors such as HSP90 and mTOR appear within them (Reactome database)

Steroid hormone receptors (SHR) are transcription factors activated by steroid hormones such as glucocorticoids, mineralocorticoids and sex hormones (278, 279). Subcellular localisation is dependent on the presenting receptor type and the ligand. The receptors include oestrogen receptors (ER α and ER β) which are predominantly intranuclear, glucocorticoid (GR) and androgen receptors (AR) which move between the cytoplasmic and nuclear compartments depending on whether they have a bound ligand (280-282). Both transcriptomic and proteomic datasets were enriched for ER α and glucocorticoid receptors. Importantly, in the absence of a suitable ligand these receptors can form complexes with heat shock proteins, mainly, HSP70/HSP90 chaperone machinery (283). The highly dynamic ATP-dependent interaction of SHRs with HSP90 complexes regulate SHR cellular location, protein stability, competency to bind steroid hormones and transcriptional activity (284). Glucocorticoid receptors and heat shock proteins appear prominently in the CAPTURE and Click ChIP datasets. Interestingly, Forneris *et al.* proposed that a mechanism involving H3K9me2 demethylation and ER α is responsible for activating the pentose phosphate pathway via a mechanism distinct from heat shock proteins (285). Unfortunately, this hypothesis has yet to be substantiated.

Heat shock proteins - HSP90, HSP60 and HSPD1 - are a recurring feature of the transcriptomic and proteomic profiles of both Click ChIP and CAPTURE datasets. Most significant amongst these is HSP90 which performs several crucial functions.

Firstly, HSP90 acts as a molecular chaperone that stabilises specific target proteins, protecting against heat stress (286). Secondly, HSP90 interacts with HDAC6, the deacetylation of this chaperone protein ensures the stability and function of client proteins such as AKT (287). Finally, HSPs participate in the removal of histones from the promoter

region of certain genes, allowing access of the transcriptional machinery for gene expression (288).

The overrepresentation of these proteins on a genome-wide and locus-specific level implies HDAC inhibition impacts these pathways. HSP dependant pathways are influenced by stressors and the acetylation status of these same proteins which modulates the pathway response.

Another important consideration when interrogating these processes is the biological model used and how this can be influenced by other factors. The Warburg effect describes the metabolic reprogramming that occurs in cancer cells resulting in the cell preferentially performing glycolysis in lieu of the more energy-efficient oxidative phosphorylation which generates ROS, promoting oxidative stress (289). However, there is increasing evidence that different malignant cell lines shunt through the GPP pathway in different ways utilising oxidative stress in promoting or challenging neoplastic growth (290).

Therefore, when studying the GPPP one must consider the Warburg effect when deciding the biological model to adopt, as the baseline redox state and microenvironmental response of some cell lines could potentially affect results and introduce additional variables to the interpretation of data. The use of a G6PD deficient cell line, an environment that mimics the *in-vivo* state most closely would overcome this problem.

Overrepresentation of PRC2 was evident in all datasets. Polycomb is a highly conserved transcriptional regulatory region which plays a vital role in maintaining gene expression patterns (291). PRC2 has a critical role in controlling gene expression during cellular differentiation, stem cell pluripotency, and beyond (292). PRC2 is a functional multimer

comprising three major components - EZH2 (histone methyltransferase), Suz12 (zinc finger domain), and EED (coordinates activity between PRC1 and PRC2) (293).

EZH2 is the catalytic core of the complex by which PRC2 exerts its regulatory effect, principally through methylation of H3K23, and to a lesser extent H3K9, resulting in transcriptional repression (294). EZH2 and H3K27me3 were overrepresented in the enrichment analysis of the transcriptome of SAHA treated K562 cells (**figure 5.5**), while pathway analysis significantly overrepresented DNA methylation (**figure 5.6**).

Work by Caputo *et al.* explored the impact of polycomb on a housekeeping gene, *glycosylphosphatidylinositol* (*GPI*) expression, and how GPI deficiency can be ameliorated using HDAC inhibitors to activate Sp1 (295). Polycomb repressed genes can maintain a poised or bivalent state, characterised by antagonistic histone modifications appearing simultaneously on the same histone tail (292). CAPTURE histone proteomics demonstrated that H3K9me2, a repressive histone PTM, and two activating PTMs – H3K9ac and H4K20me2 co-occupy the same histone tail (**figure 5.15**). It has been postulated that this poised state prepares the promoter for rapid activation (296). More recent studies have, however, refuted this theory, rather suggesting that this state protects the promoter from permanent methylation, resulting in gene silencing (297). I suspect this bivalent state performs both roles.

Caputo also describes the pivotal role Sp1 plays in restoring GPI activity (254, 298). It was proposed that Sp1 acetylation is an essential component of most housekeeping genes.

Further work in the Karadimitris lab by Makarona, *et al.* established Sp1 hyperacetylation as a critical component of selective *G6PD* induction by HDAC inhibitors, section 5.1.2 (74).

The Click CHIP RNA seq dataset showed overrepresentation of Sp1 and Sp3 DNA binding motifs and the concurrent PRC2 overrepresentation suggests (similar to the GPI deficiency scenario) that HDAC inhibition hyperacetylates Sp transcription factors. This overcomes the repressive effect of polycomb destabilising the nucleosome by reducing H3K27me3 binding (299), increasing chromatin openness at the promoter region thus increasing transcription (295).

These data highlight the importance of Sp1 in the function of housekeeping genes such as *G6PD* and the influence of the polycomb complex.

My work has shown that indeed Sp1 is essential for HDAC inhibitor induction of *G6PD* activity, given the overrepresentation of polycomb, its repressive PTMs, and bivalent histone methylation marks (**figure 5.15**). It is highly likely that *G6PD* is held in a poised state and exposure to a HDAC inhibitor hyperacetylates Sp1, suppressing polycomb, and allowing *G6PD* induction.

The 9th lysine residue of histone H3, depending on the modification, is either activating or repressive (300). Acetylation of this histone mark has a strong correlation with transcriptional activation while methylation gives rise to gene repression (Karmodiya et al., 2012). There is good evidence of the existence of bivalent domains characterised by activating and repressive histone PTMs on different amino acid residues associated with a promoter (301). Perillo *et al.* reported the combinatorial model in which antagonistic PTMs occur simultaneously on the same nucleosome, placing it in a poised state (302).

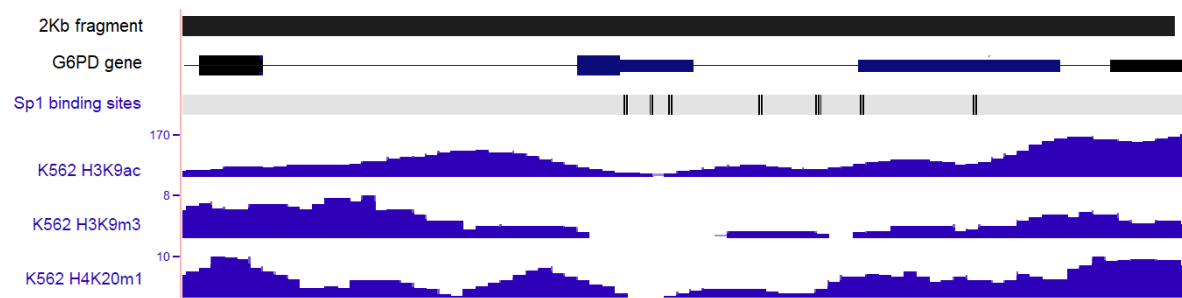


Figure 6.2 shows density graphs for key features that influence *G6PD* induction by SAHA.

Density graph of signal enrichment depicting key features that impact *G6PD* promoter activity of the 2Kb fragment isolated for CAPTURE protocols, Sp1 binding and histone modifications, as identified by ChIP-seq in K562 cells. Sp1 has multiple binding sites in the *G6PD* promoter region. The regulatory peaks for H3K9 acetylation and methylation, and H4K20 methylation show similar patterns though relative abundance is greatest in the H3K9ac profile. They are also concordant with data summarised in Figure 5.11.

ENCODE Consortium Datasets obtained UCSC gene browser (<http://genome.ucsc.edu>)

CAPTURE histone proteomics demonstrated the basal occupancy of the H3K9 histone mark is bivalent, comprising antagonistic modifications. These features are consistent with the phenotype observed in poised genes, although these results require validation to confirm bivalency. **Figure 5.15** shows H3K9 acetylation, dimethylation and H4K20 dimethylation are all present in the *G6PD* promoter region.

The CAPTURE histone proteomic data also revealed that following HDAC inhibition the H3K9 histone mark is hyperacetylated to greater than 38% but methylation only increases to 13%. This would be consistent with the hypothesis that poised genes are able to be rapidly activated (303), as the promoter is pre-loaded with poised polymerase II to initiate transcription (304, 305). The bivalent nature of the *G6PD* promoter region, unique to GPPP genes, would explain the selective induction of its activity following hyperacetylation.

Recent publications suggest that oxidative stress may be associated with specific epigenetic changes. Chang *et al.* described the activation of mammalian target of rapamycin complex 2

(mTORC2), a stress-responsive complex, which regulates *G6PD* activity (306) and plays a role in propagating epigenetic changes such as H3K9me2 to maintain oxidative homeostasis (252).

Another group observed that the inhibition of H3K9 methylation enhances the pentose phosphate pathway in human lung cancer cells (264). This, and many other groups, have highlighted the importance of the pentose phosphate pathway to control ROS in malignant disease and how the manipulation of this could form a novel strategy in the treatment of certain cancers.

Finally, it has been proposed that H3K9 demethylation by LSD1 (Lysine-specific demethylase 1) generates hydrogen peroxide and that this contributes to the assembly of RNA polymerase II complex (285). The ROS localises the transcriptional machinery to the promoter and thereby selectively activates transcription. Therefore, this oxidative burst triggered by the demethylation of H3K9 activates a DNA repair pathway that once again links the pentose phosphate pathway and this histone mark (307).

My work supports this hypothesis; the data obtained from the Click ChIP histone proteomic experiment enriched for a single histone modification, H3K9me2. This is further supported by the CAPTURE histone proteomic experiment, which showed that there was a preponderance of histone modifications at the H3K9 position, and that, at this histone mark, there was a characteristic feature observed in poised genes (i.e. opposing modifications at the same lysine residue). As this project adopted a bulk assay approach it is impossible to confirm that the repressive and activating modifications are occurring at the same lysine group without additional work.

H4K20 dimethylation is the most common methylation state on histone H4 (308). This modification is associated with transcriptional activation (309) and critical for genome integrity especially DNA damage repair (308). Like H3K9, H4K20me2 plays a key role in the DNA damage response (309).

The role of histone PTMs in the DNA repair process is increasingly recognised. Importantly, this is a pathway strongly associated with managing ROS formation, although we have yet to establish a complete understanding of the role of H4K20 methylation in this cellular stress pathway. My work demonstrates that H4K20 dimethylation is more abundant following treatment with SAHA, emphasising the importance of these pathways in the activity of this drug and the role these histone marks have in its mode of action.

More work is required to establish, firstly, whether increased *G6PD* activity due to HDAC inhibition is a direct epigenetic effect or a toxic effect of this drug activating cellular stress pathways. Secondly, there is evidence to support the notion that H3K9 and H4K20 are important histone marks in the activation of the pentose phosphate pathway. Further interrogation of this mechanism is required to better characterise this process.

6.1 CONCLUSIONS

Glucose-6-phosphate dehydrogenase is critical to maintaining oxidative homeostasis (1), particularly in the red blood cell, which is uniquely susceptible to changes in redox balance (11). This is clearly illustrated in the deficient state where haemolysis can cause significant morbidity and mortality. Targeted treatment options are desperately needed for this patient group and, although there is currently no treatment for G6PD deficiency, previous work in the Karadimitris lab identified that HDAC inhibitors selectively increase *G6PD* expression (74). This option represents a potential therapeutic approach. The aim of this project has been to decipher the mechanism by which this selective induction occurs. Ultimately, to develop a more targeted agent with better pharmacokinetic and pharmacodynamic profiles.

By adapting two novel techniques that had never been used together but worked in a complementary manner. I was able to progress our understanding of the mechanisms central to the primary question. Given the adaptability of the techniques, I was able to successfully integrate histone proteomics to both protocols so that I could explore the epigenetic changes on a genome-wide and locus-specific level.

A critical and fundamental step of the Click ChIP process was to confirm the phenocopying of parental SAHA by the clickable SAHA. I was able to demonstrate the equivalent transcriptomic impact of a clickable HDAC inhibitor to its tag-free counterpart. RNA seq illustrated the similarity of the two datasets (**Figures 4.1, 4.2 and 5.4**). This is the first time the two drugs have been compared on the transcriptomic level which was crucial to adopting SAHA BPyne as a SAHA surrogate.

Next, cellular stress response pathways were overrepresented in all SAHA treated datasets, particularly heat shock protein and steroid receptor dependant pathways. The prominence of these two families of proteins throughout the analysis highlights their importance.

Studies examining the association between steroid receptor signalling PI3K/AKT pathway activation and H3K9 demethylation have proposed mechanisms by which these factors interact, and their influence on GPP pathways (244, 310).

Wu *et al.* describe a strong correlation between AKT pathway activity, histone acetylation and glycolysis in a malignant model (244). My project supports much of those findings, and further work is required in a more appropriate cellular model to evaluate the complex interactions of these pathways and how HDAC inhibitors modify them.

I also observed that transcriptomic and proteomic datasets identified Sp1 binding motifs. This suggests a crucial role for the Sp1 transcription factor in *G6PD* expression. Early work by Almeida *et al.* investigating GPI deficiency was able to restore Sp1 binding at a mutant-binding site through Sp1 hyperacetylation by HDAC inhibition (257). This increased transcriptional output ameliorates disease. Further work by Makarona *et al.* showed that HDAC inhibitors selectively increase the binding of Sp1 on the *G6PD* promoter (74) which suggests a potential role of Sp1 in the selective upregulation of *G6PD* expression. Taken together, this data indicates that Sp1 plays a vital role in the selective induction of *G6PD*. Indeed, the post translational changes described below may influence Sp1 activity, contributing to the hyperacetylation and increased activity.

Another significant finding of this project was the co-expression of H3K9 dimethylation and acetylation, and H4K20me2. H3K9 PTMs have been described in association with multiple stress response pathways (252, 311), and in response to ROS (307). H4K20me2 has a strong

correlation with the activation of DNA repair pathways (309); these pathways are associated with the generation of ROS (312, 313). The presence of activating (H3K9ac and H4K20me2) and repressive (H3K9me2) PTMs associated with a single gene is indicative of a bivalent or poised state (292, 302, 314). It has been postulated that poised status prepares a gene for rapid activation (315). This would explain why hyperacetylation by SAHA activates *G6PD* transcription and no other GPPP genes, such as *TPI*, which does not exist in a similar state. This is a significant finding of the project that warrants further exploration. Sequential ChIP qPCR, where H3K9me2 is initially pulled down, followed by H3K9ac or H4K20me2 would be an effective method of confirming bivalency. Further work to assess the impact of acetylation and methylation at histone H3K9 is needed. Using short hairpin RNA, CRISPR-Cas9 or CRISPR-dCas9-KRAB to knockdown CBP/p300 reducing E2A-induced acetylation, and knockdown of LSD1 to reduce methylation, the effect of these histone modifications could be evaluated. This would enrich our understanding of the contribution these PTMs have to the action of SAHA on *G6PD* expression.

The biological model must be an important consideration for future work. The difficulty transducing GM12878 (a B cell line with a normal karyotype) necessitated the use of a malignant cell line, K562. The theoretical risk of excessive ROS accumulation in malignant cell lines, as described in the Warburg effect (289), must be considered in cell line selection for further work. In addition, the bulk cell omics approach used in my work provided crude population data but was difficult to interpret due to cellular heterogeneity. This is particularly relevant when examining the CAPTURE histone PTM data. Alone, it is challenging to determine whether the histone H3K9 PTMs are occurring at the same site or within different cells. Single cell omics should provide clarity with greater resolution,

revealing cellular heterogeneity and subpopulation epigenetic changes to answer these outstanding questions.

Based on these data, I propose that the *G6PD* promoter hosts an asymmetric nucleosome that maintains the *G6PD* gene in a poised state which prepares the gene for rapid activation. This poised state also represents a barrier to full trimethylation of both PTM sites which would place this housekeeping gene in a silenced state.

CAPTURE is a highly adaptable, innovative technology that can be used to understand epigenetically modifiable disease and therapies. By developing such a technique, and advancing to include histone PTM analysis, I was able to gain a greater understanding of the epigenetic landscape that differentiates *G6PD* from all other GPPP genes. I was also able to establish the sound hypothesis that *G6PD* activity can be induced by HDAC inhibitors because of activation of cellular stress responses.

Click ChIP is another versatile and pioneering technique in which the tagged drug is used as a molecular probe to gain insight into drug action and disease processes. But, when combined with histone PTM analysis, as with CAPTURE, this widens the scope of the technique.

Despite significant chromatin loss during the processing of the Click ChIP samples, data from this arm of the project provided insight that certainly aided understanding and helped answer the primary question. In future work, I would refine the pull down process (**Figures 1.9A and 1.9C**) and I anticipate that future proteomic assessment of the Click ChIP pull down would provide further comprehension of the mechanisms that contribute to SAHA epigenetic activity and the pathways therein.

The question is no longer whether HDAC inhibitors activate these pathways. That fact has been clearly established by my work and others (73, 127, 270). The more pertinent question is the mechanism by which this occurs. Is it a direct effect in which the HDAC inhibitor induces cytotoxicity? The evidence for this is strong as multiple components of these pathways are found to be altered by SAHA. Alternatively, are these the terminal events of a cascade initiated by an epigenetic change to H3K9 and H4K20 histone PTMs, as described in section 5.5. Or are these two processes simultaneously activated and occur concurrently with minimal interaction? To determine whether stimulation of cellular stress response pathways is due to direct cytotoxicity of SAHA - SAHA-induced cytotoxicity can be inhibited by 3-methyladenine (3- MA), an inhibitor of phosphatidylinositol 3-kinases (PI3K). PI3K plays a significant role in many biological processes, including controlling the activation of mTOR which I have established is crucial to the cellular stress response stimulated by SAHA, section 5.6 (316, 317).

The clinical implications of this work are potentially substantial. Once validated, these data will advance our understanding of the epigenetic controls of *G6PD* expression so that drugs with better pharmacological profiles can be developed. Current HDAC inhibitors have limitations. SAHA has a poor side-effect profile, mainly due to its broad pharmacological action. As a pan inhibitor, SAHA affects multiple systems, especially platelet activation, signalling, aggregation and degranulation, resulting in thrombocytopenia, platelet dysfunction and haemorrhage (318). Sodium butyrate, as a naturally occurring substance, has a much better side-effect profile. However, aliphatic acids such as sodium butyrate are relatively weak HDAC inhibitors. These drugs also undergo extensive first-pass metabolism so require doses in the millimolar range compared to other HDAC inhibitors which are dosed

in nM- μ M range to achieve therapeutic concentrations (233, 319, 320). This makes sodium butyrate a less attractive therapeutic agent.

A new therapeutic approach in the treatment of G6PD deficiency is desperately needed, particularly in treating the most severe forms of the disease, but also to ameliorate haemolysis for those with class II/III disease who have been exposed to oxidative stress.

The global distribution of G6PD deficiency strongly correlates with malaria, as G6PD deficiency has a protective effect against plasmodium infection, the so-called malaria protection hypothesis (321). There has been great difficulty in eradicating chronic plasmodium vivax or ovale infections. The treatment of choice is Primaquine which generates intermediate oxidative metabolites which can result in haemolysis in G6PD-deficient individuals. Attempts at overcoming this problem with novel agents have been unsuccessful. The development of a preparation containing a targeted G6PD inducer could be taken with the anti-malarial and would mitigate risk of haemolysis, thus helping to resolve this global dilemma.

Mitapivat is a first-in-class allosteric activator of red cell pyruvate kinase (322), which is used to treat pyruvate kinase deficiency. The success of this GPPP enzyme inducer illustrates the benefit of this drug in treating this disease and other red cell disorders such as, sickle cell disease (323) and thalassaemia (324). The versatility demonstrated by Mitapivat shows the potential benefit that a G6PD enzyme inducer could have in the treatment of G6PD deficiency and other disorders. An example of this would be the use of a G6PD activator as an adjuvant therapy in treating those malignant disorders reliant on the accumulation of ROS for malignant expansion. As our understanding of tumour metabolism expands, the utilisation of this approach will become increasingly relevant.

I am confident that the outstanding questions can be answered using the tools I have developed during this project, in addition to those already available in our lab. A class I G6PD-deficient B cell line would be an ideal biological system to explore these questions. This would be complemented by a recent development in the CAPTURE protocol which utilises a two-plasmid system, significantly reducing the transduction payload and increasing the likelihood of successful transduction in a B cell line (325). Used in conjunction with the advances made in the Click ChIP and CAPTURE protocols, we now have the tools to further our understanding of the mechanisms responsible for selective *G6PD* induction, and this insight will drive the innovation necessary to develop treatments for this and other disorders.

Appendices

Appendix A

Reagents and kits

Product	Company	Catalogue Number
2-Propanol	Sigma-Aldrich	I9516
Acetonitrile (LC-MS Grade)	ThermoFisher Scientific	51101
Agarose gel extraction kit	Jena Bioscience	PP-202
Agencourt AMPure XP Beads	Beckman Coulter	A63881
Agilent Bioanalyzer High Sensitivity DNA kit	Agilent	5067-4626
Agilent RNA 6000 Pico Kit	Agilent	5067-1513
β-mercaptoethanol	Sigma-Aldrich	M6250
BCA Protein Assay Kit	Thermo Scientific	22235
Biotin-PEG ₃ -azide	Cayman Chemicals	23419
Biotinylated SAHA	Sai Life Science	-
Blue Prestained Protein Standard	New England BioLabs	P7718
Copper (II) sulphate	Sigma-Aldrich	C2284
Dithiothreitol	Sigma-Aldrich	D9779
DMEM	Sigma-Aldrich	5796
DMSO	Sigma-Aldrich	D2650
Dynabeads MyOne streptavidin T1	Thermo-Fisher	65601
Dynabeads Protein A	ThermoFisher Scientific	10002D
Dynabeads protein G	Thermo-Fisher	10004D
FastDigest Eco72I (PmlI)	ThermoFisher Scientific	FD0364
FastDigest XbaI	ThermoFisher Scientific	FD0684
FastDigest Green Buffer (10X)	ThermoFisher Scientific	B72
G418	Sigma-Aldrich	A1720
Gibson Assembly	New England BioLabs	E5510
IGEPAL	Sigma-Aldrich	I8896
Iodoacetamide	Sigma-Aldrich	A3221
LC-MS Grade water	ThermoFisher Scientific	51140
L-Glutamine	Sigma-Aldrich	G7513
Methanol free formaldehyde (16%)	ThermoFisher Scientific	28908
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 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
PageRuler Plus Prestained Protein Ladder	ThermoFisher Scientific	26619
PCRBIO HiFi Polymerase	PCR Biosystems	PB10.41-02
Penicillin / Streptomycin	Sigma-Aldrich	P0781
Peptide Desalting Spin Columns	ThermoFisher Scientific	89852
Phenol Chloroform	Sigma-Aldrich	77617
Phosphate Buffer Saline	Sigma-Aldrich	D8537
Pierce™ Spin Cups - Cellulose Acetate Filter	ThermoFisher Scientific	69702
PMSF	Sigma-Aldrich	93482
ProLong® Silver Antifade Reagent with DAPI	Life Technologies	P36931
Proteinase K	ThermoFisher Scientific	26160
Puromycin	Thermo-Fisher	A1113802
Qubit High Sensitivity DNA kit	Life Technologies	Q32851
Qubit RNA assay kit	Life Technologies	Q32852
ReliaPrep RNA Cell Miniprep System kit	Promega	Z6012
Restriction enzyme Bpil (BbsI)	ThermoFisher Scientific	ER1001
Restriction enzyme ClaI (Bsu15I)	ThermoFisher Scientific	ER0141
Restriction enzyme Kpn2I (BspEI)	ThermoFisher Scientific	ER0531
Restriction enzyme Mph1103I (NsiI)	ThermoFisher Scientific	ER0731
RevertAid RT kit	ThermoFisher Scientific	K1691
RNAse A	ThermoFisher Scientific	EN0531
RPMI1640	Sigma-Aldrich	R8758
SAHA	Cayman Chemical	10009929
SAHA BPyne	Cayman Chemical	10675
SDS proteomic grade (20%)	Thermo Scientific	M112
SIGMAFAST Protease inhibitor	Sigma-Aldrich	S8820
Sodium butyrate	Sigma-Aldrich	156-54-7
Sodium deoxycholate	Sigma-Aldrich	302-95-4
Sodium L-ascorbate	Sigma-Aldrich	D5097
Streptavidin agarose beads	Thermo-Fisher	SA10004
SYBR Select Master Mix	Applied Biosystems	4472908
T4 DNA ligase	New England Biolabs	M0202L
T4 DNA Ligase Reaction Buffer	New England Biolabs	B0202S
TMT10plex Isobaric Mass Tag Labelling Kit	Thermo Scientific	90111
Trifluoroacetic acid (0.1%) LC-MS grade	ThermoFisher Scientific	85172
Tris EDTA	Fisher BioReagents	165404
Tris-Glycine gel (4-12%)	ThermoFisher Scientific	XP04122BOX
Tris(2-carboxyethyl) phosphine hydrochloride	Thermo Scientific	PG82090
Tris-HCl 1M	ThermoFisher Scientific	15568025
Triton X-100 (20%)	Cayman Chemicals	600217

Appendix B

Plasmids



Figure S1 pHR-SFFV-dCas9-BFP plasmid

Plasmid used to modify pEF1a-BirA-V5-neo plasmid for viral LTRs and BFP expression

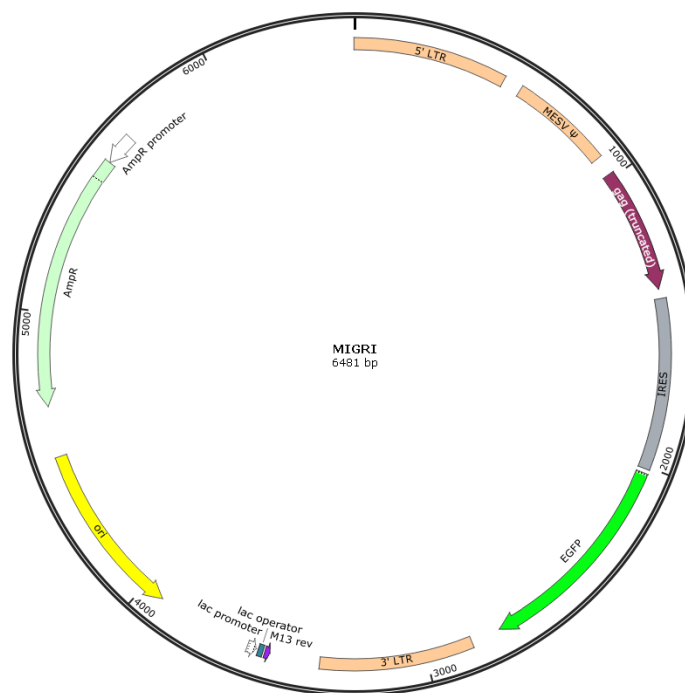


Figure S2 MIGRI plasmid

Plasmid used to modify pEF1a-FB-dCas9-puro plasmid for eGFP expression

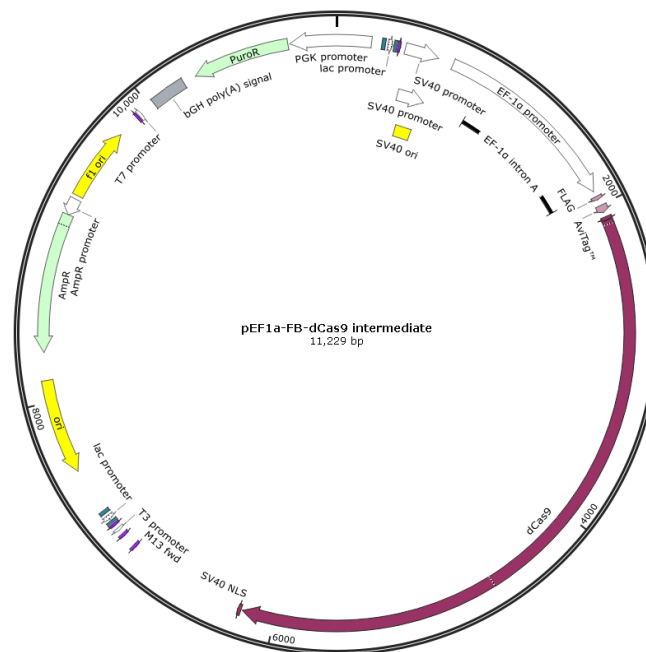


Figure S3 pEF1a-FB-dCas9 intermediate plasmid

Intermediate construct used in second step in development of pEF1a-FB-dCas9-GFP

Appendix C

Workflows

Figure S1 RNA seq analysis workflow

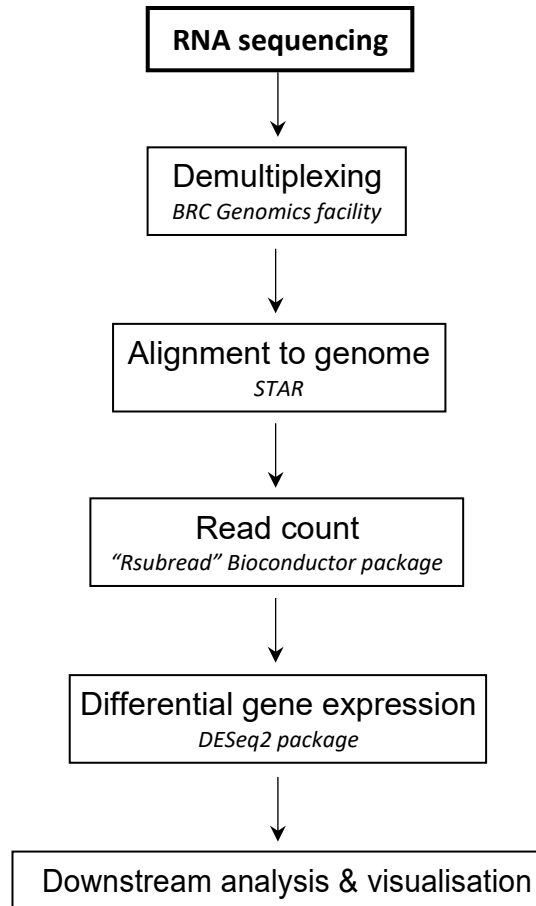
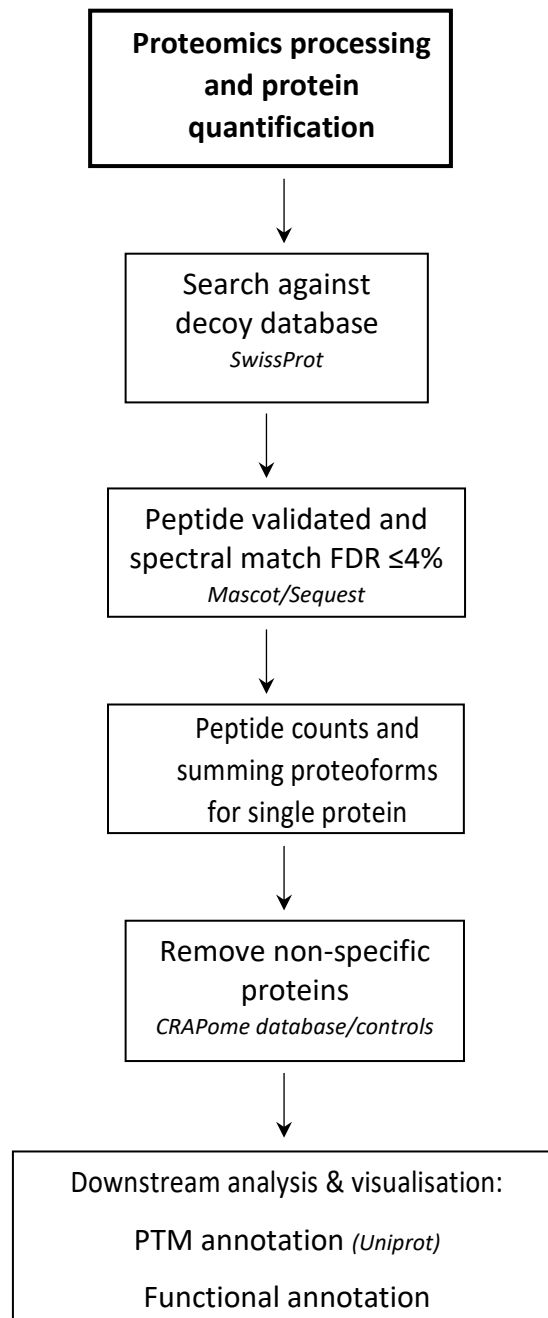


Figure S2 Proteomics analysis workflow



Appendix D

Sanger sequencing chromatograms

Figure S1-2: Sanger sequencing chromatograms for two *G6PD* guide RNA illustrating integration to pU6-sgX-NT1 plasmid and alignment with guide RNA sequences.

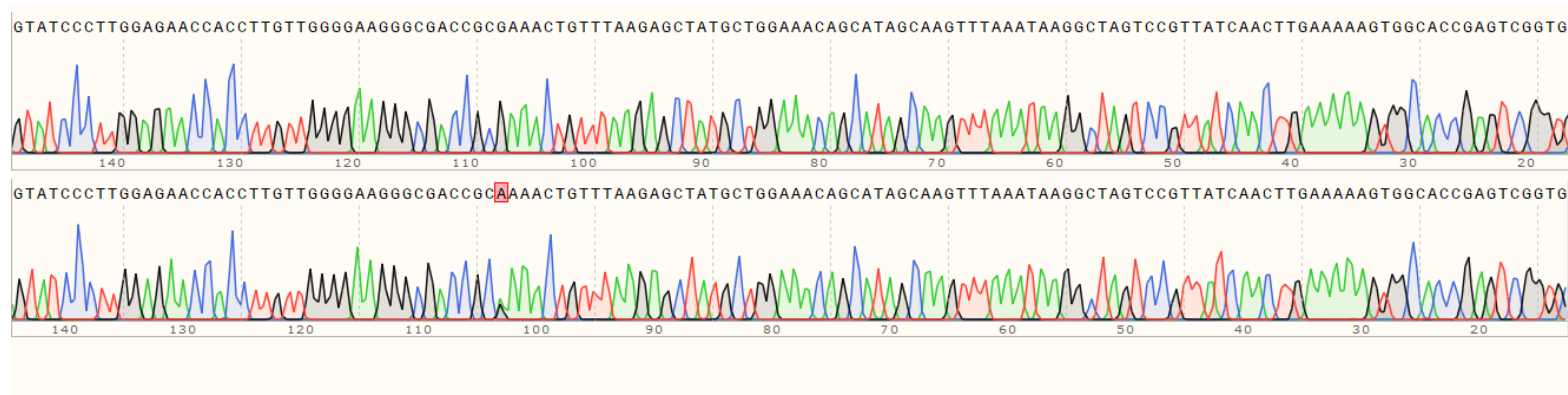


Figure S1 sequencing chromatograms for pU6-sgG6PD 22-NT1.

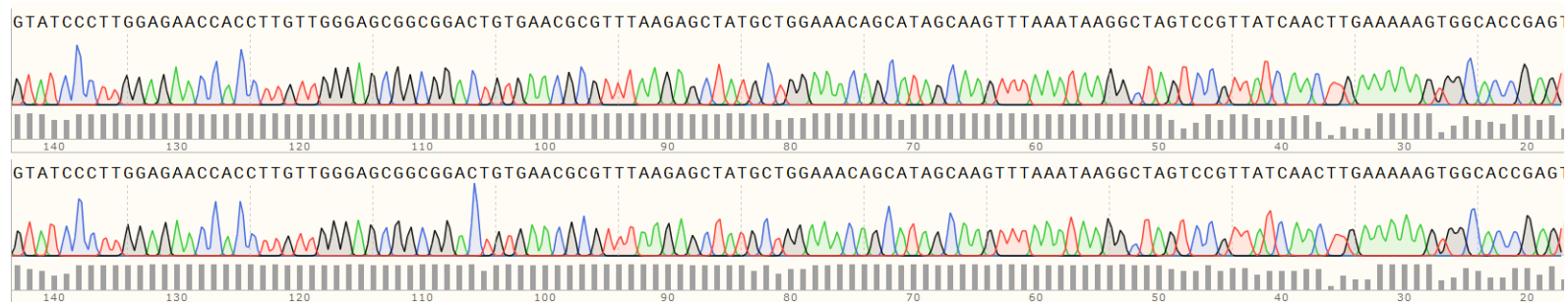


Figure S2 sequencing chromatograms for pU6-sgG6PD 23-NT1.

Appendix E

Table S1 CAPTURE fragment sequence (5'-3')

Sequence includes *G6PD* promoter region and exon 1

AGGGTCAGACCCCGCAGTGGAGGAGGAGGGAGGAGTAGAAACAGTCCAGCTCGCCGCCCAAGTAACCTGG
GTCCTGAATCGGCCCGCCTTGCCAGTGCTCCAGAAGCGCGGAGCAGGAACGGGCTGGGGCCCAAAAAGA
GGGGGAGCCTGAACGTCCGGGGGAAGTTTCGGAGGCGGCGGAACGCCACGGATGGAACCTGTCTTTGG
GGAAAAGGACCACACCTGTCAGCAGAGTCCGTGACAGCTGAGAAGGGTGGGAGCGGCGGACTGTGAACGCT
GGTAGGGCCCCGGCGCTCCGAGAAAGTCCCAGTTTCGGGTGCGCCTTCCCTACCACGCTTCCGGCTTCCGGT
GTCATAGCTGTGGGATCCGGAAGTAAAAACACAAGCCCCGCCCGAGAACTCGGGAAGCCGGCGAGAAGT
GTGAGGCCGCGGTAGGGCCGCATCCCGCTCCGGAGAGAAAGTCTGAGTCCGCCAGGCTCTGCAGGCCCGCGG
AAGCTCGGTAATGATAAGCACGCCGGCCACTTTGCAGGGCGTCACCGCCTACACGCCCTCGTCTCTCGGAC
GGCGGCGTCTAGCCTCGGGGCGCTCGGCCGCCCGCCCTCTCCGGGGGAGGAATCAAGAAGAGACTGCCCA
ATAGGGCCGGCTTGACCCGCGAACAGGCGAGGGTTCCCGGGGGAGTGCGCGGCGAGAAGGCCCGCCAG
GAGCCGAGGGACAGCCAGAGGAGGCGTGGCCACGCTGCCGGCGGAAGTGGAGCCCTCCGCGAGCGCGCG
AGGCCGCCGGGGCAGGCGGGGAAACCGGACAGTAGGGGCGGGGCCTGGCCGGCGATGGGGATTCTGGGAG
CACTACGCGGAGCTGCACCCGTGCCCCGCCGAATTGGGGATGCAGAGCAGCGGCAGCGGGTATGGCAGGCA
GCCGGCGGGCCGGCCTCCAGCGCAGGTGCCGAGAGGCAGGGGCTGGCCTGGGATGCGCGCGCACCTGCC
TCGCCCCGCCCGCCCGCACGAGGGGTGGTGGCCGAGGCCCCGCCCGCACGCCTCGCTGAGGCGGGTCCG
CTCAGCCCAGGCGCCCGCCCCGCCCGCCGATTAATGGGCCGGCGGGGCTCAGCCCCGGAAACGGTCG
TACACTTCGGGGCTGCGAGCGCGGAGGGCGACGACGACGAAGCGCAGGTAACCGGCCGGGCGGGCGCCGC
GCAGGCGGAGGAGCGTACTGTCCCGCGCTGCGCCGCGCGGCGGTAAAATACACGCTGTTTGTGTGCTTGAG
AACCGAGCAGAATCGAGAGGGTCTTAACCAATCCCTTTATACCCCGCACCTCCTCTTTGAGCCCCTGAGACCC
CGAGAGCGAAGGGGACTTGCCGACCGGGGTACCCAGCTTGGAAGGGGAGGGCTGGAGCTGAACCTCAGC
ATCTGCACCATCTCCATGCTCCAGGTCATTGTGGAGTTCGCGCTACAGTCGGAATGAGATGGTCCTGGGCA
CGCAGTTCATGCCCCACAAGGATTTTACTCGTTGTCCAGAATTGATGCTGTAGTCGGAATACACCAATGCTT
TGAGTAATTTTGAATGTACACCCTGAATGAAGGCTGCCTAGGAGAGAGTGGCTGGAGCCCAGAGCCAGCAG
TTTCTAACCCATCAACCACTCCCAATGCCAGCCGTTACAAGGAGTGATTTGGGCAATCAGGTGTCACCCTG
GTGTGAGACCCAGAGGAACTCTCAAGAAAGGGGCTAACTTCTCAATGCTCTCCTGTTCTTCTGCCTTGTTAAC
GAGCCTTTCTTCCACCAGACAGCGTCATGGCAGAGCAGGTGGCCCTGAGCCGGACCCAGGTGTGCGGGATCC
TGCGGGAAGAGCTTTTCCAGGGCGATGCCTTCCATCAGTCGGATACACACATATTCATCATCATGGGTGCATC
GGTGAGTATCTCCAGGCCCAATCTTAAAGCCAGGAAGTGCCTGCTCCATGCCTCAGCTTTT

Total: 2,000bp

Appendix F

Proteins of interest

Figure S1 Refinement of Click ChIP general proteomics list

Total list of proteins returned	
Accession	Protein
A0A075B6S5	Immunoglobulin kappa variable 1-27
A2RRD8	Zinc finger protein 320
A6ND36	Family with sequence similarity 83 member g
O14727	Apoptotic protease-activating factor 1
O75970	Multiple PDZ domain protein
P00924	Enolase 1
P04150	Glucocorticoid receptor
P05413	Fatty acid-binding protein, heart
P13929	Beta-enolase
P35998	26S proteasome regulatory subunit 7
P39900	Macrophage metalloelastase-12
P51157	Ras-related protein Rab-28
P56277	Cx9C motif-containing protein 4
Q5T0J7	Testis-expressed protein 35
Q6UXS9	Inactive caspase-12
Q7L576	Cytoplasmic FMR1-interacting protein 1
Q7Z4V0	Zinc finger protein 438
Q8WXH0	Nesprin-2
Q9BV73	Centrosome-associated protein CEP250
Q9H8H2	Probable ATP-dependent RNA helicase (DDX31)
Q9NRD0	F-box only protein 8
Q9NWS1	PCNA-interacting partner
Q9P2F5	Storkhead-box protein 2
Q9UL54	Serine/threonine-protein kinase (TAO2)
Q9ULL8	Protein Shroom4
Q86UK5	Limbin
Q96AY4	Tetratricopeptide repeat protein 28
Q562F6	Shugoshin 2
Q14865	AT-rich interactive domain-containing protein 5B
Q15526	Surfeit locus protein 1

Non localising proteins	
Accession	Protein
A6ND36	Immunoglobulin kappa variable 1-27
O14727	Zinc finger protein 320
P35998	Family with sequence similarity 83 member g
P39900	Apoptotic protease-activating factor 1
P56277	Multiple PDZ domain protein
Q6UXS9	Enolase 1
Q7L576	Glucocorticoid receptor
Q9BV73	Fatty acid-binding protein, heart
Q9ULL8	Beta-enolase
Q96AY4	26S proteasome regulatory subunit 7

Proteins removed by controls	
Accession	Protein
Q8WXH0	Nesprin-2
Q9UL54	Serine/threonine-protein kinase (TAO2)

CRAPome	
Accession	Protein
A0A075B6S5	Immunoglobulin kappa variable 1-27
A2RRD8	Zinc finger protein 320
O75970	Multiple PDZ domain protein
P00924	Enolase 1
P13929	Beta-enolase
Q96AY4	Tetratricopeptide repeat protein 28
Q15526	Surfeit locus protein 1

Proteins of interest	
Accession	Protein
P04150	Glucocorticoid receptor
P05413	Fatty acid-binding protein, heart
P51157	Ras-related protein Rab-28
Q5T0J7	Testis-expressed protein 35
Q7Z4V0	Zinc finger protein 438
Q8WXH0	Nesprin-2
Q9H8H2	Probable ATP-dependent RNA helicase (DDX31)
Q9NRD0	F-box only protein 8
Q9NWS1	PCNA-interacting partner
Q9P2F5	Storkhead-box protein 2
Q9UL54	Serine/threonine-protein kinase (TAO2)
Q86UK5	Limbin
Q562F6	Shugoshin 2
Q14865	AT-rich interactive domain-containing protein 5B

References

1. Thomas D, Cherest H, Surdin-Kerjan Y. Identification of the structural gene for glucose-6-phosphate dehydrogenase in yeast. Inactivation leads to a nutritional requirement for organic sulfur. *The EMBO journal*. 1991;10(3):547-53.
2. Domingo GJ, Satyagraha AW, Anvikar A, Baird K, Bancone G, Bansil P, et al. G6PD testing in support of treatment and elimination of malaria: recommendations for evaluation of G6PD tests. *Malar J*. 2013;12:391.
3. Alfarouk KO, Ahmed S, Elliott RL, Benoit A, Alqahtani SS, Ibrahim ME, et al. The Pentose Phosphate Pathway Dynamics in Cancer and Its Dependency on Intracellular pH. *Metabolites*. 2020;10(7):285.
4. Berg JM, Tymoczko JL, Stryer L. *Biochemistry*. 6th ed. Freeman WH, editor. New York 2006.
5. Luzzatto L, Mehta A, Vulliamy T. Glucose-6-phosphate Dehydrogenase Deficiency. *Metabolic and Molecular Basics of Inherited Disease*. 2001.
6. Luzzatto L, Ally M, Notaro R. Glucose-6-phosphate dehydrogenase deficiency. *Blood*. 2020;136(11):1225-40.
7. Kruger NJ, von Schaewen A. The oxidative pentose phosphate pathway: structure and organisation. *Current opinion in plant biology*. 2003;6(3):236-46.
8. Kim JW, Dang CV. Multifaceted roles of glycolytic enzymes. *Trends Biochem Sci*. 2005;30(3):142-50.
9. Peretó J. Embden-Meyerhof-Parnas Pathway. In: Gargaud M, Amils R, Quintanilla JC, Cleaves HJ, Irvine WM, Pinti DL, et al., editors. *Encyclopedia of Astrobiology*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 485-.
10. Weitzman JK-P, editor. *Krebs citric acid cycle: half a century and still turning*. Biochem Soc Symp.
11. Schmidt-Rohr K. Oxygen Is the High-Energy Molecule Powering Complex Multicellular Life: Fundamental Corrections to Traditional Bioenergetics. *ACS omega*. 2020;5(5):2221-33.
12. Ramos-Martinez JI. The regulation of the pentose phosphate pathway: remember Krebs. *Archives of biochemistry and biophysics*. 2017;614:50-2.
13. Corrons J-LV. Red blood cell enzyme defects. *ESH handbook on disorders of erythropoiesis, erythrocytes and iron metabolism*. 2009;17:436-53.
14. Luzzatto L, Arese P. Favism and Glucose-6-Phosphate Dehydrogenase Deficiency. *The New England journal of medicine*. 2018;378(11):1068-9.
15. Howes RE, Battle KE, Satyagraha AW, Baird JK, Hay SI. G6PD deficiency: global distribution, genetic variants and primaquine therapy. *Adv Parasitol*. 2013;81:133-201.
16. Pai G, Sprenkle JA, Do TT, Marenzi CE, Migeon BR. Localization of loci for hypoxanthine phosphoribosyltransferase and glucose-6-phosphate dehydrogenase and biochemical evidence of nonrandom X chromosome expression from studies of a human X-autosome translocation. *Proceedings of the National Academy of Sciences*. 1980;77(5):2810-3.
17. Luzzatto L, Battistuzzi G. Glucose-6-phosphate dehydrogenase. *Advances in Human Genetics* 14: Springer; 1985. p. 217-329.
18. Wrigley N, Heather J, Bonsignore A, De Flora A. Human erythrocyte glucose 6-phosphate dehydrogenase: electron microscope studies on structure and interconversion of tetramers, dimers and monomers. *Journal of molecular biology*. 1972;68(3):483-99.
19. Cohen P, Rosemeyer M. Subunit interactions of glucose-6-phosphate dehydrogenase from human erythrocytes. *European journal of biochemistry*. 1969;8(1):8-15.

20. Howes RE, Battle KE, Satyagraha AW, Baird JK, Hay SI. G6PD deficiency: global distribution, genetic variants and primaquine therapy. *Advances in parasitology*. 81: Elsevier; 2013. p. 133-201.
21. Kotaka M, Gover S, Vandeputte-Rutten L, Au SW, Lam VM, Adams MJ. Structural studies of glucose-6-phosphate and NADP⁺ binding to human glucose-6-phosphate dehydrogenase. *Acta Crystallographica Section D: Biological Crystallography*. 2005;61(5):495-504.
22. Wang XT, Chan TF, Lam VM, Engel PC. What is the role of the second “structural” NADP⁺-binding site in human glucose 6-phosphate dehydrogenase? *Protein Science*. 2008;17(8):1403-11.
23. Oberlé I, Camerino G, Wrogemann K, Arveiler B, Hanauer A, Raimondi E, et al. Multipoint genetic mapping of the Xq26-q28 region in families with fragile X mental retardation and in normal families reveals tight linkage of markers in q26–q27. *Human genetics*. 1987;77(1):60-5.
24. Filosa S, Calabro V, Lania G, Vulliamy T, Brancati C, Tagarelli A, et al. G6PD haplotypes spanning Xq28 from F8C to red/green color vision. *Genomics*. 1993;17(1):6-14.
25. Boyer SH, Graham JB. Linkage between the X chromosome loci for glucose-6-phosphate dehydrogenase electrophoretic variation and hemophilia A. *American journal of human genetics*. 1965;17(4):320.
26. Arngimsson R, Dokal I, Luzzatto L, Connor J. Dyskeratosis congenita: three additional families show linkage to a locus in Xq28. *Journal of medical genetics*. 1993;30(7):618-9.
27. Persico MG, Viglietto G, Martini G, Toniolo D, Paonessa G, Moscatelli C, et al. Isolation of human glucose-6-phosphate dehydrogenase (G6PD) cDNA clones: primary structure of the protein and unusual 5' non-coding region. *Nucleic acids research*. 1986;14(6):2511-22.
28. Takizawa T, Huang I-Y, Ikuta T, Yoshida A. Human glucose-6-phosphate dehydrogenase: primary structure and cDNA cloning. *Proceedings of the National Academy of Sciences*. 1986;83(12):4157-61.
29. Martini G, Toniolo D, Vulliamy T, Luzzatto L, Dono R, Viglietto G, et al. Structural analysis of the X-linked gene encoding human glucose 6-phosphate dehydrogenase. *EMBO J*. 1986;5(8):1849-55.
30. Luzzatto L, Nannelli C, Notaro R. Glucose-6-phosphate dehydrogenase deficiency. *Hematology/Oncology Clinics*. 2016;30(2):373-93.
31. Gómez-Manzo S, Marcial-Quino J, Vanoye-Carlo A, Serrano-Posada H, Ortega-Cuellar D, González-Valdez A, et al. Glucose-6-phosphate dehydrogenase: update and analysis of new mutations around the world. *International journal of molecular sciences*. 2016;17(12):2069.
32. Jin DY, Jeang KT. Isolation of full-length cDNA and chromosomal localization of human NF- κ B modulator NEMO to Xq28. *J Biomed Sci*. 1999;6(2):115-20.
33. Fusco F, Paciolla M, Napolitano F, Pescatore A, D'Addario I, Bal E, et al. Genomic architecture at the Incontinentia Pigmenti locus favours de novo pathological alleles through different mechanisms. *Human molecular genetics*. 2012;21(6):1260-71.
34. Zinatizadeh M, Masoumalinejad Z, Nejatizadeh A, Shekari M, Parnak F, Zaree F. A review of NEMO protein and its relationship with genetic diseases. *J Genet Genome Res*. 2018;5(1):33.
35. Rothwarf DM, Zandi E, Natoli G, Karin M. IKK- γ is an essential regulatory subunit of the I κ B kinase complex. *Nature*. 1998;395(6699):297-300.
36. Hinz M, Scheidereit C. The I κ B kinase complex in NF- κ B regulation and beyond. *EMBO reports*. 2014;15(1):46-61.
37. Meletis J, Konstantopoulos K. Favism-from the “avoid fava beans” of Pythagoras to the present. *Haema*. 2004;7(1):17-21.
38. Cordes W. Experiences with plasmochin in malaria. Anonymous 15th Annual Report Boston, MA: United Fruit Co. 1926:66-71.
39. Arese P, De Flora A, editors. Pathophysiology of hemolysis in glucose-6-phosphate dehydrogenase deficiency. *Seminars in hematology*; 1990.
40. Luzzatto L, Arese P. Favism and glucose-6-phosphate dehydrogenase deficiency. *New England Journal of Medicine*. 2018;378(1):60-71.

41. Nkhoma ET, Poole C, Vannappagari V, Hall SA, Beutler E. The global prevalence of glucose-6-phosphate dehydrogenase deficiency: a systematic review and meta-analysis. *Blood Cells Mol Dis*. 2009;42(3):267-78.
42. Murray CJ, Barber RM, Foreman KJ, Ozgoren AA, Abd-Allah F, Abera SF, et al. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: quantifying the epidemiological transition. *The Lancet*. 2015;386(10009):2145-91.
43. Gelabert P, Olalde I, de-Dios T, Civit S, Lalueza-Fox C. Malaria was a weak selective force in ancient Europeans. *Scientific reports*. 2017;7(1):1-10.
44. Mehta A, Mason PJ, Vulliamy TJ. Glucose-6-phosphate dehydrogenase deficiency. *Baillieres Best Pract Res Clin Haematol*. 2000;13(1):21-38.
45. Beutler E, editor *Abnormalities of the hexose monophosphate shunt*. Seminars in hematology; 1971.
46. Grimes A. *Human red cell metabolism* Blackwell Scientific Publications. Oxford, London; 1980.
47. Wintrobe M, Lee G, Boggs D, Bithell T, Foerster J, Athens J, et al. Nutritional factors in the production and function of erythrocytes. *Clinical hematology: Lea & Febiger*, Philadelphia, PA USA; 1981. p. 136-70.
48. Bonsignore A, Fornaini G, Fantoni A, Leoncini G, Segni P. Relationship between age and enzymatic activities in human erythrocytes from normal and fava bean-sensitive subjects. *The Journal of clinical investigation*. 1964;43(5):834-42.
49. Cappellini MD, Fiorelli G. Glucose-6-phosphate dehydrogenase deficiency. *Lancet*. 2008;371(9606):64-74.
50. Gómez-Manzo S, Marcial-Quino J, Ortega-Cuellar D, Serrano-Posada H, González-Valdez A, Vanoye-Carlo A, et al. Functional and biochemical analysis of glucose-6-phosphate dehydrogenase (G6PD) variants: elucidating the molecular basis of G6PD deficiency. *Catalysts*. 2017;7(5):135.
51. Mason PJ, Vulliamy TJ. Glucose-6-phosphate Dehydrogenase (G6PD) Deficiency: Genetics. *eLS*.
52. Mason PJ, Bautista JM, Gilsanz F. G6PD deficiency: the genotype-phenotype association. *Blood Rev*. 2007;21(5):267-83.
53. Chen EY, Cheng A, Lee A, Kuang WJ, Hillier L, Green P, et al. Sequence of human glucose-6-phosphate dehydrogenase cloned in plasmids and a yeast artificial chromosome. *Genomics*. 1991;10(3):792-800.
54. Beutler E, Vulliamy TJ. Hematologically important mutations: glucose-6-phosphate dehydrogenase. *Blood Cells Mol Dis*. 2002;28(2):93-103.
55. Beutler E. The genetics of glucose-6-phosphate dehydrogenase deficiency. *Semin Hematol*. 1990;27(2):137-64.
56. Ho L, Hoefel D, Bock F, Saint CP, Newcombe G. Biodegradation rates of 2-methylisoborneol (MIB) and geosmin through sand filters and in bioreactors. *Chemosphere*. 2007;66(11):2210-8.
57. Roper D, Layton M, Rees D, Lambert C, Vulliamy T, De la Salle B, et al. Laboratory diagnosis of G6PD deficiency. A British Society for Haematology Guideline. *British Journal of Haematology*. 2020;189(1):24-38.
58. Minucci A, Moradkhani K, Hwang MJ, Zuppi C, Giardina B, Capoluongo E. Glucose-6-phosphate dehydrogenase (G6PD) mutations database: review of the “old” and update of the new mutations. *Blood Cells, Molecules, and Diseases*. 2012;48(3):154-65.
59. Vulliamy T, D'urso M, Battistuzzi G, Estrada M, Foulkes N, Martini G, et al. Diverse point mutations in the human glucose-6-phosphate dehydrogenase gene cause enzyme deficiency and mild or severe hemolytic anemia. *Proceedings of the National Academy of Sciences*. 1988;85(14):5171-5.

60. Vulliamy TJ, Kaeda JS, Ait-Chafa D, Mangerini R, Roper D, Barbot J, et al. Clinical and haematological consequences of recurrent G6PD mutations and a single new mutation causing chronic nonspherocytic haemolytic anaemia. *British journal of haematology*. 1998;101(4):670-5.
61. Filosa S, Cai W, Galanello R, Cao A, De Mattia D, Schettini F, et al. A novel single-base mutation in the glucose 6-phosphate dehydrogenase gene is associated with chronic non-spherocytic haemolytic anaemia. *Human genetics*. 1994;94(5):560-2.
62. Cunningham AD, Hwang S, Mochly-Rosen D. Glucose-6-phosphate dehydrogenase deficiency and the need for a novel treatment to prevent kernicterus. *Clinics in perinatology*. 2016;43(2):341-54.
63. Slusher TM, Zamora TG, Appiah D, Stanke JU, Strand MA, Lee BW, et al. Burden of severe neonatal jaundice: a systematic review and meta-analysis. *BMJ paediatrics open*. 2017;1(1).
64. Luzzatto L. *Oxford Textbook of Medicine*. Glucose-6-phosphate dehydrogenase (G6PD) deficiency: Oxford University Press; 2010.
65. Beutler E. Glucose-6-phosphate dehydrogenase deficiency: a historical perspective. *Blood*. 2008;111(1):16-24.
66. Maisels MJ. Neonatal jaundice. *Pediatrics in Review*. 2006;27(12):443.
67. In: rd, editor. *Guidelines for the Treatment of Malaria*. WHO Guidelines Approved by the Guidelines Review Committee. Geneva 2015.
68. Bosman A, Cunningham J. Proposal for an Evidence Review Group (ERG) on G6PD testing to support increased access to primaquine for radical cure of *Plasmodium vivax* and for malaria chemoprophylaxis. WHO Malaria Policy Advisory Committee Meeting. 2013:1-5.
69. Au SW, Gover S, Lam VM, Adams MJ. Human glucose-6-phosphate dehydrogenase: the crystal structure reveals a structural NADP⁺ molecule and provides insights into enzyme deficiency. *Structure*. 2000;8(3):293-303.
70. Mason P, Sonati M, MacDonald D, Lanza C, Busutil D, Town M, et al. New glucose-6-phosphate dehydrogenase mutations associated with chronic anemia. 1995.
71. Gomez-Manzo S, Marcial-Quino J, Vanoye-Carlo A, Serrano-Posada H, Ortega-Cuellar D, Gonzalez-Valdez A, et al. Glucose-6-Phosphate Dehydrogenase: Update and Analysis of New Mutations around the World. *International journal of molecular sciences*. 2016;17(12).
72. Naylor CE, Rowland P, Basak AK, Gover S, Mason PJ, Bautista JM, et al. Glucose 6-phosphate dehydrogenase mutations causing enzyme deficiency in a model of the tertiary structure of the human enzyme. *Blood*. 1996;87(7):2974-82.
73. Kim HJ, Bae SC. Histone deacetylase inhibitors: molecular mechanisms of action and clinical trials as anti-cancer drugs. *Am J Transl Res*. 2011;3(2):166-79.
74. Makarona K, Caputo VS, Costa JR, Liu B, O'Connor D, Iskander D, et al. Transcriptional and epigenetic basis for restoration of G6PD enzymatic activity in human G6PD-deficient cells. *Blood*. 2014;124(1):134-41.
75. Gray S, Qian C-N, Furge K, Guo X, Teh B. Microarray profiling of the effects of histone deacetylase inhibitors on gene expression in cancer cell lines. *International journal of oncology*. 2004;24(4):773-95.
76. Rada-Iglesias A, Enroth S, Ameer A, Koch CM, Clelland GK, Respuela-Alonso P, et al. Butyrate mediates decrease of histone acetylation centered on transcription start sites and down-regulation of associated genes. *Genome research*. 2007;17(6):708-19.
77. Wang Z, Zang C, Cui K, Schones DE, Barski A, Peng W, et al. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell*. 2009;138(5):1019-31.
78. Waddington CH. The epigenotype. *International journal of epidemiology*. 2012;41(1):10-3.
79. Waddington CH. Embryology, epigenetics and biogenetics. *Nature*. 1956;177(4522):1241-.
80. Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. *Science*. 1974;184(4139):868-71.
81. Roth SY, Denu JM, Allis CD. Histone acetyltransferases. *Annual review of biochemistry*. 2001;70(1):81-120.

82. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 1997;389(6648):251-60.
83. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. 2000;403(6765):41-5.
84. Jenuwein T, Allis CD. Translating the histone code. *Science*. 2001;293(5532):1074-80.
85. Musselman CA, Lalonde M-E, Côté J, Kutateladze TG. Perceiving the epigenetic landscape through histone readers. *Nature structural & molecular biology*. 2012;19(12):1218.
86. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature structural & molecular biology*. 2007;14(11):1025-40.
87. Ruthenburg AJ, Allis CD, Wysocka J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Molecular cell*. 2007;25(1):15-30.
88. Biswas S, Rao CM. Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *European journal of pharmacology*. 2018;837:8-24.
89. Shahbazian MD, Grunstein M. Functions of site-specific histone acetylation and deacetylation. *Annu Rev Biochem*. 2007;76:75-100.
90. Lo W-S, Trievel RC, Rojas JR, Duggan L, Hsu J-Y, Allis CD, et al. Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. *Molecular cell*. 2000;5(6):917-26.
91. Allfrey VG, editor Structural modifications of histones and their possible role in the regulation of ribonucleic acid synthesis. *Proceedings Canadian Cancer Conference*; 1966.
92. Hon GC, Hawkins RD, Ren B. Predictive chromatin signatures in the mammalian genome. *Human molecular genetics*. 2009;18(R2):R195-R201.
93. Mas JC, Noel ES, Ober EA. Chromatin modification in zebrafish development. *Methods in Cell Biology*. 104: Elsevier; 2011. p. 401-28.
94. Geisler SJ, Paro R. Trithorax and Polycomb group-dependent regulation: a tale of opposing activities. *Development*. 2015;142(17):2876-87.
95. Wysocka J, Allis CD, Coonrod S. Histone arginine methylation and its dynamic regulation. *Front Biosci*. 2006;11:344-55.
96. Miao F, Li S, Chavez V, Lanting L, Natarajan R. Coactivator-associated arginine methyltransferase-1 enhances nuclear factor-κB-mediated gene transcription through methylation of histone H3 at arginine 17. *Molecular Endocrinology*. 2006;20(7):1562-73.
97. Singh RK, Gunjan A. Histone tyrosine phosphorylation comes of age. *Epigenetics*. 2011;6(2):153-60.
98. Zhang Y. Transcriptional regulation by histone ubiquitination and deubiquitination. *Genes Dev*. 2003;17(22):2733-40.
99. Wilkinson KD. Ubiquitination and deubiquitination: targeting of proteins for degradation by the proteasome. *Semin Cell Dev Biol*. 2000;11(3):141-8.
100. Seigneurin-Berny D, Verdel A, Curtet S, Lemerrier C, Garin J, Rousseaux S, et al. Identification of components of the murine histone deacetylase 6 complex: link between acetylation and ubiquitination signaling pathways. *Mol Cell Biol*. 2001;21(23):8035-44.
101. Sun ZW, Allis CD. Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature*. 2002;418(6893):104-8.
102. Hunt CR, Ramnarain D, Horikoshi N, Iyengar P, Pandita RK, Shay JW, et al. Histone modifications and DNA double-strand break repair after exposure to ionizing radiations. *Radiat Res*. 2013;179(4):383-92.
103. Shiio Y, Eisenman RN. Histone sumoylation is associated with transcriptional repression. *Proceedings of the National Academy of Sciences*. 2003;100(23):13225-30.
104. Nathan D, Ingvarsdottir K, Sterner DE, Bylebyl GR, Dokmanovic M, Dorsey JA, et al. Histone sumoylation is a negative regulator in *Saccharomyces cerevisiae* and shows dynamic interplay with positive-acting histone modifications. *Genes & development*. 2006;20(8):966-76.

105. Zhang Y, Shin H, Song JS, Lei Y, Liu XS. Identifying positioned nucleosomes with epigenetic marks in human from CHIP-Seq. *BMC Genomics*. 2008;9:537.
106. Clapier CR, Cairns BR. The biology of chromatin remodeling complexes. *Annu Rev Biochem*. 2009;78:273-304.
107. Jin B, Li Y, Robertson KD. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes & cancer*. 2011;2(6):607-17.
108. Fischle W, Tseng BS, Dormann HL, Ueberheide BM, Garcia BA, Shabanowitz J, et al. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature*. 2005;438(7071):1116-22.
109. An W. Histone acetylation and methylation: combinatorial players for transcriptional regulation. *Subcell Biochem*. 2007;41:351-69.
110. Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet*. 2009;10(5):295-304.
111. Consortium EP. The ENCODE (ENCyclopedia of DNA elements) project. *Science*. 2004;306(5696):636-40.
112. Kaikkonen MU, Lam MT, Glass CK. Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular research*. 2011;90(3):430-40.
113. Prasanth KV, Spector DL. Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. *Genes & development*. 2007;21(1):11-42.
114. Iorio MV, Piovan C, Croce CM. Interplay between microRNAs and the epigenetic machinery: an intricate network. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*. 2010;1799(10-12):694-701.
115. Chen Z, Place RF, Jia Z-J, Pookot D, Dahiya R, Li L-C. Antitumor effect of dsRNA-induced p21WAF1/CIP1 gene activation in human bladder cancer cells. *Molecular cancer therapeutics*. 2008;7(3):698-703.
116. Phillips DM. The presence of acetyl groups of histones. *Biochem J*. 1963;87:258-63.
117. Allfrey VG, Faulkner R, Mirsky AE. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A*. 1964;51:786-94.
118. Allfrey VG, Faulkner R, Mirsky A. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America*. 1964;51(5):786.
119. Hebbes TR, Thorne AW, Crane-Robinson C. A direct link between core histone acetylation and transcriptionally active chromatin. *The EMBO journal*. 1988;7(5):1395-402.
120. Kimura A, Matsubara K, Horikoshi M. A decade of histone acetylation: marking eukaryotic chromosomes with specific codes. *J Biochem*. 2005;138(6):647-62.
121. Brownell JE, Allis CD. An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. *Proc Natl Acad Sci U S A*. 1995;92(14):6364-8.
122. Kleff S, Andrulis ED, Anderson CW, Sternglanz R. Identification of a gene encoding a yeast histone H4 acetyltransferase. *J Biol Chem*. 1995;270(42):24674-7.
123. Taunton J, Hassig CA, Schreiber SL. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science*. 1996;272(5260):408-11.
124. Yang X-J, Ogryzko VV, Nishikawa J-i, Howard BH, Nakatani Y. A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature*. 1996;382(6589):319-24.
125. Mariadason JM, Corner GA, Augenlicht LH. Genetic reprogramming in pathways of colonic cell maturation induced by short chain fatty acids: comparison with trichostatin A, sulindac, and curcumin and implications for chemoprevention of colon cancer. *Cancer research*. 2000;60(16):4561-72.
126. Li G, Margueron R, Hu G, Stokes D, Wang Y-H, Reinberg D. Highly compacted chromatin formed in vitro reflects the dynamics of transcription activation in vivo. *Molecular cell*. 2010;38(1):41-53.

127. Eckschlager T, Plch J, Stiborova M, Hrabeta J. Histone deacetylase inhibitors as anticancer drugs. *International journal of molecular sciences*. 2017;18(7):1414.
128. Spange S, Wagner T, Heinzel T, Krämer OH. Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *The international journal of biochemistry & cell biology*. 2009;41(1):185-98.
129. Cano A, Pestana A. Purification and properties of a histone acetyltransferase from *Artemia salina*, highly efficient with H1 histone. *Eur J Biochem*. 1979;97(1):65-72.
130. Marmorstein R. Structure of histone acetyltransferases. *J Mol Biol*. 2001;311(3):433-44.
131. Mujtaba S, Zeng L, Zhou MM. Structure and acetyl-lysine recognition of the bromodomain. *Oncogene*. 2007;26(37):5521-7.
132. Sanchez R, Zhou MM. The role of human bromodomains in chromatin biology and gene transcription. *Curr Opin Drug Discov Devel*. 2009;12(5):659-65.
133. Lee KK, Workman JL. Histone acetyltransferase complexes: one size doesn't fit all. *Nature reviews Molecular cell biology*. 2007;8(4):284-95.
134. Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY, et al. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell*. 1996;84(6):843-51.
135. Nagy Z, Tora L. Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene*. 2007;26(37):5341-57.
136. Avvakumov N, Cote J. The MYST family of histone acetyltransferases and their intimate links to cancer. *Oncogene*. 2007;26(37):5395-407.
137. Lafon A, Chang CS, Scott EM, Jacobson SJ, Pillus L. MYST opportunities for growth control: yeast genes illuminate human cancer gene functions. *Oncogene*. 2007;26(37):5373-84.
138. Doi M, Hirayama J, Sassone-Corsi P. Circadian regulator CLOCK is a histone acetyltransferase. *Cell*. 2006;125(3):497-508.
139. Shiama N. The p300/CBP family: integrating signals with transcription factors and chromatin. *Trends in cell biology*. 1997;7(6):230-6.
140. Henry KW, Wyce A, Lo W-S, Duggan LJ, Emre NT, Kao C-F, et al. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes & development*. 2003;17(21):2648-63.
141. John S, Howe L, Tafrov ST, Grant PA, Sternglanz R, Workman JL. The something about silencing protein, Sas3, is the catalytic subunit of NuA3, a γ TAFII30-containing HAT complex that interacts with the Spt16 subunit of the yeast CP (Cdc68/Pob3)–FACT complex. *Genes & development*. 2000;14(10):1196-208.
142. Wittschieben BØ, Otero G, de Bizemont T, Fellows J, Erdjument-Bromage H, Ohba R, et al. A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. *Molecular cell*. 1999;4(1):123-8.
143. Bolden JE, Peart MJ, Johnstone RW. Anticancer activities of histone deacetylase inhibitors. *Nat Rev Drug Discov*. 2006;5(9):769-84.
144. Gao L, Cueto MA, Asselbergs F, Atadja P. Cloning and functional characterization of HDAC11, a novel member of the human histone deacetylase family. *J Biol Chem*. 2002;277(28):25748-55.
145. Gregoret I, Lee Y-M, Goodson HV. Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. *Journal of molecular biology*. 2004;338(1):17-31.
146. Marks PA, Dokmanovic M. Histone deacetylase inhibitors: discovery and development as anticancer agents. *Expert opinion on investigational drugs*. 2005;14(12):1497-511.
147. Bhalla KN. Epigenetic and chromatin modifiers as targeted therapy of hematologic malignancies. *Journal of Clinical Oncology*. 2005;23(17):3971-93.
148. Haberland M, Montgomery RL, Olson EN. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nature Reviews Genetics*. 2009;10(1):32-42.

149. Haggarty SJ, Koeller KM, Wong JC, Grozinger CM, Schreiber SL. Domain-selective small-molecule inhibitor of histone deacetylase 6 (HDAC6)-mediated tubulin deacetylation. *Proceedings of the National Academy of Sciences*. 2003;100(8):4389-94.
150. Klein MA, Denu JM. Biological and catalytic functions of sirtuin 6 as targets for small-molecule modulators. *Journal of Biological Chemistry*. 2020:jbc. REV120. 011438.
151. Michishita E, Park JY, Burneskis JM, Barrett JC, Horikawa I. Evolutionarily conserved and nonconserved cellular localizations and functions of human SIRT proteins. *Molecular biology of the cell*. 2005;16(10):4623-35.
152. Jin Q, Yan T, Ge X, Sun C, Shi X, Zhai Q. Cytoplasm-localized SIRT1 enhances apoptosis. *Journal of cellular physiology*. 2007;213(1):88-97.
153. Haigis MC, Guarente LP. Mammalian sirtuins--emerging roles in physiology, aging, and calorie restriction. *Genes Dev*. 2006;20(21):2913-21.
154. Sengupta N, Seto E. Regulation of histone deacetylase activities. *Journal of cellular biochemistry*. 2004;93(1):57-67.
155. Bantscheff M, Hopf C, Savitski MM, Dittmann A, Grandi P, Michon AM, et al. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat Biotechnol*. 2011;29(3):255-65.
156. Yang XJ, Seto E. The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. *Nat Rev Mol Cell Biol*. 2008;9(3):206-18.
157. You A, Tong JK, Grozinger CM, Schreiber SL. CoREST is an integral component of the CoREST-human histone deacetylase complex. *Proc Natl Acad Sci U S A*. 2001;98(4):1454-8.
158. Tong JK, Hassig CA, Schnitzler GR, Kingston RE, Schreiber SL. Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature*. 1998;395(6705):917-21.
159. Zhang Y, Iratni R, Erdjument-Bromage H, Tempst P, Reinberg D. Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell*. 1997;89(3):357-64.
160. Karagianni P, Wong J. HDAC3: taking the SMRT-N-CoRrect road to repression. *Oncogene*. 2007;26(37):5439-49.
161. Guenther MG, Barak O, Lazar MA. The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. *Mol Cell Biol*. 2001;21(18):6091-101.
162. Zaratiegui M, Irvine DV, Martienssen RA. Noncoding RNAs and gene silencing. *Cell*. 2007;128(4):763-76.
163. Li Y, Kao GD, Garcia BA, Shabanowitz J, Hunt DF, Qin J, et al. A novel histone deacetylase pathway regulates mitosis by modulating Aurora B kinase activity. *Genes & development*. 2006;20(18):2566-79.
164. Grossman SR, Deato ME, Brignone C, Chan HM, Kung AL, Tagami H, et al. Polyubiquitination of p53 by a ubiquitin ligase activity of p300. *Science*. 2003;300(5617):342-4.
165. Linares LK, Kiernan R, Triboulet R, Chable-Bessia C, Latreille D, Cuvier O, et al. Intrinsic ubiquitination activity of PCAF controls the stability of the oncoprotein Hdm2. *Nat Cell Biol*. 2007;9(3):331-8.
166. Toleman C, Paterson AJ, Whisenhunt TR, Kudlow JE. Characterization of the histone acetyltransferase (HAT) domain of a bifunctional protein with activable O-GlcNAcase and HAT activities. *Journal of Biological Chemistry*. 2004;279(51):53665-73.
167. Boyault C, Sadoul K, Pabion M, Khochbin S. HDAC6, at the crossroads between cytoskeleton and cell signaling by acetylation and ubiquitination. *Oncogene*. 2007;26(37):5468-76.
168. Aufsatz W, Stoiber T, Rakic B, Naumann K. Arabidopsis histone deacetylase 6: a green link to RNA silencing. *Oncogene*. 2007;26(37):5477-88.
169. Chen Y, Sprung R, Tang Y, Ball H, Sangras B, Kim SC, et al. Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Molecular & cellular proteomics*. 2007;6(5):812-9.
170. Sykes SM, Mellert HS, Holbert MA, Li K, Marmorstein R, Lane WS, et al. Acetylation of the p53 DNA-binding domain regulates apoptosis induction. *Molecular cell*. 2006;24(6):841-51.

171. Philippe M, Larondelle Y, Lemaigre F, Mariamé B, Delhez H, Mason P, et al. Promoter function of the human glucose-6-phosphate dehydrogenase gene depends on two GC boxes that are cell specifically controlled. *European journal of biochemistry*. 1994;226(2):377-84.
172. Allahverdiyev AM, Bagirova M, Elcicek S, Koc RC, Ates SC, Baydar SY, et al. Glucose-6-phosphate dehydrogenase deficiency and malaria: a method to detect primaquine-induced hemolysis in vitro. *Dehydrogenases*. 2012:chapt. 4.
173. Szabo P, Purrello M, Rocchi M, Archidiacono N, Alhadeff B, Filippi G, et al. Cytological mapping of the human glucose-6-phosphate dehydrogenase gene distal to the fragile-X site suggests a high rate of meiotic recombination across this site. *Proceedings of the National Academy of Sciences*. 1984;81(24):7855-9.
174. Galgóczy P, Rosenthal A, Platzer M. Human–mouse comparative sequence analysis of the NEMO gene reveals an alternative promoter within the neighboring G6PD gene. *Gene*. 2001;271(1):93-8.
175. Franzè A, Ferrante MI, Fusco F, Santoro A, Sanzari E, Martini G, et al. Molecular anatomy of the human glucose 6-phosphate dehydrogenase core promoter. *FEBS letters*. 1998;437(3):313-8.
176. Riggs M, Whittaker R, Neumann J, Ingram V. n-Butyrate causes histone modification in HeLa and Friend erythroleukaemia cells. *Nature*. 1977;268(5619):462-4.
177. Bertrand P. Inside HDAC with HDAC inhibitors. *European journal of medicinal chemistry*. 2010;45(6):2095-116.
178. Daly K, Shirazi-Beechey SP. Microarray analysis of butyrate regulated genes in colonic epithelial cells. *DNA and cell biology*. 2006;25(1):49-62.
179. Chambers A, Banerjee S, Chaplin T, Dunne J, Debernardi S, Joel S, et al. Histone acetylation-mediated regulation of genes in leukaemic cells. *European journal of cancer*. 2003;39(8):1165-75.
180. Mitsiades CS, Mitsiades NS, McMullan CJ, Poulaki V, Shringarpure R, Hideshima T, et al. Transcriptional signature of histone deacetylase inhibition in multiple myeloma: biological and clinical implications. *Proceedings of the National Academy of Sciences*. 2004;101(2):540-5.
181. Ellis D, Lawman Z, Bonham K. Histone acetylation is not an accurate predictor of gene expression following treatment with histone deacetylase inhibitors. *Biochemical and biophysical research communications*. 2008;367(3):656-62.
182. Ramaiah MJ, Tangutur AD, Manyam RR. Epigenetic modulation and understanding of HDAC inhibitors in cancer therapy. *Life sciences*. 2021;277:119504.
183. Marks PA. Histone deacetylase inhibitors: a chemical genetics approach to understanding cellular functions. *Biochimica et biophysica acta*. 2010;1799(10-12):717-25.
184. Drummond DC, Noble CO, Kirpotin DB, Guo Z, Scott GK, Benz CC. Clinical development of histone deacetylase inhibitors as anticancer agents. *Annu Rev Pharmacol Toxicol*. 2005;45:495-528.
185. Davie JR. Inhibition of histone deacetylase activity by butyrate. *J Nutr*. 2003;133(7 Suppl):2485S-93S.
186. Marks PA, Breslow R. Dimethyl sulfoxide to vorinostat: development of this histone deacetylase inhibitor as an anticancer drug. *Nat Biotechnol*. 2007;25(1):84-90.
187. Finnin MS, Donigian JR, Cohen A, Richon VM, Rifkind RA, Marks PA, et al. Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature*. 1999;401(6749):188-93.
188. Marks PA. Discovery and development of SAHA as an anticancer agent. *Oncogene*. 2007;26(9):1351-6.
189. Curtin M, Glaser K. Histone deacetylase inhibitors: the Abbott experience. *Curr Med Chem*. 2003;10(22):2373-92.
190. Asgar M, Senawong G, Sripa B, Senawong T. Synergistic anticancer effects of cisplatin and histone deacetylase inhibitors (SAHA and TSA) on cholangiocarcinoma cell lines. *International Journal of Oncology*. 2016;48(1):409-20.
191. Glozak MA, Sengupta N, Zhang X, Seto E. Acetylation and deacetylation of non-histone proteins. *gene*. 2005;363:15-23.

192. Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, et al. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*. 2009;325(5942):834-40.
193. Anders L, Guenther MG, Qi J, Fan ZP, Marineau JJ, Rahl PB, et al. Genome-wide localization of small molecules. *Nat Biotechnol*. 2014;32(1):92-6.
194. Luong JH, Vashist SK. Chemistry of biotin–streptavidin and the growing concern of an emerging biotin interference in clinical immunoassays. *ACS omega*. 2019;5(1):10-8.
195. Liu X, Zhang Y, Chen Y, Li M, Zhou F, Li K, et al. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell*. 2017;170(5):1028-43 e19.
196. Kolb HC, Finn M, Sharpless KB. Click chemistry: diverse chemical function from a few good reactions. *Angewandte Chemie International Edition*. 2001;40(11):2004-21.
197. Presolski SI, Hong VP, Finn M. Copper-catalyzed azide–alkyne click chemistry for bioconjugation. *Current protocols in chemical biology*. 2011;3(4):153-62.
198. Tyler DS, Vappiani J, Caneque T, Lam EYN, Ward A, Gilan O, et al. Click chemistry enables preclinical evaluation of targeted epigenetic therapies. *Science*. 2017;356(6345):1397-401.
199. Salisbury CM, Cravatt BF. Optimization of activity-based probes for proteomic profiling of histone deacetylase complexes. *J Am Chem Soc*. 2008;130(7):2184-94.
200. Fujii H, Fujita T. Isolation of Specific Genomic Regions and Identification of Their Associated Molecules by Engineered DNA-Binding Molecule-Mediated Chromatin Immunoprecipitation (enChIP) Using the CRISPR System and TAL Proteins. *International journal of molecular sciences*. 2015;16(9):21802-12.
201. Albrecht C, Hosiner S, Tichy B, Aldrian S, Hajdu S, Nürnberger S. Comparison of lentiviral packaging mixes and producer cell lines for RNAi applications. *Molecular biotechnology*. 2015;57(6):499-505.
202. Fujita T, Fujii H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochem Biophys Res Commun*. 2013;439(1):132-6.
203. Vermeulen M, Dejardin J. Locus-specific chromatin isolation. *Nat Rev Mol Cell Biol*. 2020;21(5):249-50.
204. Dejardin J, Kingston RE. Purification of proteins associated with specific genomic Loci. *Cell*. 2009;136(1):175-86.
205. Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*. 2013;155(7):1479-91.
206. Trevino AE, Zhang F. Genome editing using Cas9 nickases. *Methods in enzymology*. 546: Elsevier; 2014. p. 161-74.
207. Naito Y, Hino K, Bono H, Ui-Tei K. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics*. 2015;31(7):1120-3.
208. Kutner RH, Zhang XY, Reiser J. Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. *Nat Protoc*. 2009;4(4):495-505.
209. Durlak M, Fugazza C, Elangovan S, Marini MG, Marongiu MF, Moi P, et al. A Novel High-Content Immunofluorescence Assay as a Tool to Identify at the Single Cell Level gamma-Globin Inducing Compounds. *PloS one*. 2015;10(10):e0141083.
210. Delrue I, Pan Q, Baczmanska AK, Callens BW, Verdoodt LLM. Determination of the Selection Capacity of Antibiotics for Gene Selection. *Biotechnol J*. 2018;13(8):e1700747.
211. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc*. 2008;3(6):1101-8.
212. Bustin S, Huggett J. qPCR primer design revisited. *Biomol Detect Quantif*. 2017;14:19-28.
213. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
214. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research*. 2013;41(10):e108-e.

215. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*. 2014;15(12):1-21.
216. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*. 2016;44(W1):W90-W7.
217. Wolfer AM, DS Correia G, Sands CJ, Camuzeaux S, Yuen AH, Chekmeneva E, et al. peakPanther, an R package for large-scale targeted extraction and integration of annotated metabolic features in LC–MS profiling datasets. *Bioinformatics*. 2021;37(24):4886-8.
218. Mellacheruvu D, Wright Z, Couzens AL, Lambert J-P, St-Denis NA, Li T, et al. The CRAPome: a contaminant repository for affinity purification–mass spectrometry data. *Nature methods*. 2013;10(8):730-6.
219. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: network analysis and visualization of proteomics data. *Journal of proteome research*. 2018;18(2):623-32.
220. Snyder MP, Gingeras TR, Moore JE, Weng Z, Gerstein MB, Ren B, et al. Perspectives on ENCODE. *Nature*. 2020;583(7818):693-8.
221. Gearing M. SEQUENCE AND LIGATION-INDEPENDENT CLONING (SLIC). *INTRODUCTION TO PLASMIDS* 101.58.
222. Frecha C, Levy C, Cosset FL, Verhoeven E. Advances in the field of lentivector-based transduction of T and B lymphocytes for gene therapy. *Mol Ther*. 2010;18(10):1748-57.
223. Iriguchi S, Yasui Y, Kawai Y, Arima S, Kunitomo M, Sato T, et al. A clinically applicable and scalable method to regenerate T-cells from iPSCs for off-the-shelf T-cell immunotherapy. *Nature Communications*. 2021;12(1):1-15.
224. Schatz PJ. Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in *Escherichia coli*. *Bio/technology*. 1993;11(10):1138-43.
225. Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol*. 2014;32(12):1262-7.
226. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. 2013;31(9):827-32.
227. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen T, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*. 2014;343(6166):84-7.
228. Avery SV. Molecular targets of oxidative stress. *Biochemical Journal*. 2011;434(2):201-10.
229. Dayon L, Hainard A, Licker V, Turck N, Kuhn K, Hochstrasser DF, et al. Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Analytical chemistry*. 2008;80(8):2921-31.
230. Minucci S, Pelicci PG. Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nat Rev Cancer*. 2006;6(1):38-51.
231. Berger SL. The complex language of chromatin regulation during transcription. *Nature*. 2007;447(7143):407-12.
232. Hutt DM, Roth DM, Marchal C, Bouchecareilh M. Using histone deacetylase inhibitors to analyze the relevance of HDACs for translation. *HDAC/HAT Function Assessment and Inhibitor Development: Methods and Protocols*. 2017:77-91.
233. Xu W, Parmigiani R, Marks P. Histone deacetylase inhibitors: molecular mechanisms of action. *Oncogene*. 2007;26(37):5541-52.
234. Kelly WK, Marks PA. Drug insight: histone deacetylase inhibitors—development of the new targeted anticancer agent suberoylanilide hydroxamic acid. *Nature Clinical Practice Oncology*. 2005;2(3):150-7.
235. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*. 2013;14(1):1-14.

236. Li H, Liefke R, Jiang J, Kurland JV, Tian W, Deng P, et al. Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature*. 2017;549(7671):287-91.
237. Duan R, Du W, Guo W. EZH2: a novel target for cancer treatment. *Journal of hematology & oncology*. 2020;13(1):1-12.
238. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012;489(7414):75-82.
239. Chen A, Chen D, Chen Y. Advances of DNase-seq for mapping active gene regulatory elements across the genome in animals. *Gene*. 2018;667:83-94.
240. Leroy G, Dimaggio PA, Chan EY, Zee BM, Blanco MA, Bryant B, et al. A quantitative atlas of histone modification signatures from human cancer cells. *Epigenetics & chromatin*. 2013;6(1):20.
241. Zhang Y, Fonslow BR, Shan B, Baek M-C, Yates III JR. Protein analysis by shotgun/bottom-up proteomics. *Chemical reviews*. 2013;113(4):2343-94.
242. Elias JE, Haas W, Faherty BK, Gygi SP. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nature methods*. 2005;2(9):667-75.
243. Fryer CJ, Archer TK. Chromatin remodelling by the glucocorticoid receptor requires the BRG1 complex. *Nature*. 1998;393(6680):88-91.
244. Wu D, Yan Y, Wei T, Ye Z, Xiao Y, Pan Y, et al. An acetyl-histone vulnerability in PI3K/AKT inhibition-resistant cancers is targetable by both BET and HDAC inhibitors. *Cell reports*. 2021;34(7):108744.
245. Moldovan G-L, Dejsuphong D, Petalcorin MI, Hofmann K, Takeda S, Boulton SJ, et al. Inhibition of homologous recombination by the PCNA-interacting protein PARI. *Molecular cell*. 2012;45(1):75-86.
246. Baba A, Ohtake F, Okuno Y, Yokota K, Okada M, Imai Y, et al. PKA-dependent regulation of the histone lysine demethylase complex PHF2-ARID5B. *Nature cell biology*. 2011;13(6):668-75.
247. Rajendran P, Williams DE, Ho E, Dashwood RH. Metabolism as a key to histone deacetylase inhibition. *Crit Rev Biochem Mol Biol*. 2011;46(3):181-99.
248. Li B, Carey M, Workman JL. The role of chromatin during transcription. *Cell*. 2007;128(4):707-19.
249. Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007;129(4):823-37.
250. Rosenfeld JA, Wang Z, Schones DE, Zhao K, DeSalle R, Zhang MQ. Determination of enriched histone modifications in non-genic portions of the human genome. *BMC genomics*. 2009;10(1):1-11.
251. Garcia BA, Mollah S, Ueberheide BM, Busby SA, Muratore TL, Shabanowitz J, et al. Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nature protocols*. 2007;2(4):933-8.
252. Chang RC, Thomas KN, Mehta NA, Veazey KJ, Parnell SE, Golding MC. Programmed suppression of oxidative phosphorylation and mitochondrial function by gestational alcohol exposure correlate with widespread increases in H3K9me2 that do not suppress transcription. *Epigenetics & chromatin*. 2021;14(1):1-16.
253. Georgiou E, Layton M, Karadimitris A. Inherited GPI deficiency: a disorder of histone hypoacetylation. *Birth Defects Res C Embryo Today*. 2009;87(4):327-34.
254. Almeida AM, Murakami Y, Layton DM, Hillmen P, Sellick GS, Maeda Y, et al. Hypomorphic promoter mutation in PIGM causes inherited glycosylphosphatidylinositol deficiency. *Nat Med*. 2006;12(7):846-51.
255. van Wijk R, van Solinge WW. The energy-less red blood cell is lost: erythrocyte enzyme abnormalities of glycolysis. *Blood*. 2005;106(13):4034-42.
256. Arya R, Layton DM, Bellingham AJ. Hereditary red cell enzymopathies. *Blood Rev*. 1995;9(3):165-75.
257. Almeida A, Layton M, Karadimitris A. Inherited glycosylphosphatidyl inositol deficiency: a treatable CDG. *Biochimica et biophysica acta*. 2009;1792(9):874-80.

258. Dynan WS, Tjian R. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell*. 1983;35(1):79-87.
259. Torigoe T, Izumi H, Wakasugi T, Niina I, Igarashi T, Yoshida T, et al. DNA topoisomerase II poison TAS-103 transactivates GC-box-dependent transcription via acetylation of Sp1. *Journal of Biological Chemistry*. 2005;280(2):1179-85.
260. Kadonaga JT, Courey AJ, Ladika J, Tjian R. Distinct regions of Sp1 modulate DNA binding and transcriptional activation. *Science*. 1988;242(4885):1566-70.
261. Pascal E, Tjian R. Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. *Genes & development*. 1991;5(9):1646-56.
262. Courey AJ, Tjian R. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell*. 1988;55(5):887-98.
263. Courey AJ, Holtzman DA, Jackson SP, Tjian R. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell*. 1989;59(5):827-36.
264. Lu C, Yang D, Klement JD, Colson YL, Oberlies NH, Pearce CJ, et al. H3K9me3 represses G6PD expression to suppress the pentose phosphate pathway and ROS production to promote human mesothelioma growth. *Oncogene*. 2022;1-12.
265. Salisbury CM, Cravatt BF. Activity-based probes for proteomic profiling of histone deacetylase complexes. *Proc Natl Acad Sci U S A*. 2007;104(4):1171-6.
266. Accili D, Arden KC. FoxOs at the crossroads of cellular metabolism, differentiation, and transformation. *Cell*. 2004;117(4):421-6.
267. Matsuzaki H, Daitoku H, Hatta M, Aoyama H, Yoshimochi K, Fukamizu A. Acetylation of Foxo1 alters its DNA-binding ability and sensitivity to phosphorylation. *Proceedings of the National Academy of Sciences*. 2005;102(32):11278-83.
268. Chen J, Lu Y, Tian M, Huang Q. Molecular mechanisms of FOXO1 in adipocyte differentiation. *Journal of molecular endocrinology*. 2019;62(3):R239-R53.
269. Akasaki Y, Alvarez-Garcia O, Saito M, Caramés B, Iwamoto Y, Lotz MK. FoxO transcription factors support oxidative stress resistance in human chondrocytes. *Arthritis & rheumatology*. 2014;66(12):3349-58.
270. Xu W, Trepel J, Neckers L. Ras, ROS and proteotoxic stress: a delicate balance. *Cancer cell*. 2011;20(3):281-2.
271. Smidt MP, Smits SM, Bouwmeester H, Hamers FP, van der Linden AJ, Hellemons AJ, et al. Early developmental failure of substantia nigra dopamine neurons in mice lacking the homeodomain gene Pitx3. 2004.
272. Chen JD, Evans RM. A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature*. 1995;377(6548):454-7.
273. L'honoré A, Commère P-H, Ouimette J-F, Montarras D, Drouin J, Buckingham M. Redox regulation by Pitx2 and Pitx3 is critical for fetal myogenesis. *Developmental cell*. 2014;29(4):392-405.
274. Davie JR. Inhibition of histone deacetylase activity by butyrate. *The Journal of nutrition*. 2003;133(7):2485S-93S.
275. Cao R, Zhang Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Molecular cell*. 2004;15(1):57-67.
276. Kültz D. Molecular and evolutionary basis of the cellular stress response. *Annu Rev Physiol*. 2005;67:225-57.
277. Stacchiotti A. Exploring cellular stress response and chaperones. *MDPI*; 2019. p. 408.
278. Griekspoor A, Zwart W, Neefjes J, Michalides R. Visualizing the action of steroid hormone receptors in living cells. *Nuclear receptor signaling*. 2007;5(1):nrs. 05003.
279. Eick GN, Thornton JW. Evolution of steroid receptors from an estrogen-sensitive ancestral receptor. *Molecular and cellular endocrinology*. 2011;334(1-2):31-8.
280. Stenoien DL, Mancini MG, Patel K, Allegretto EA, Smith CL, Mancini MA. Subnuclear trafficking of estrogen receptor- α and steroid receptor coactivator-1. *Molecular endocrinology*. 2000;14(4):518-34.

281. Scheinman RI, Gualberto A, Jewell CM, Cidlowski JA, Baldwin Jr AS. Characterization of mechanisms involved in transrepression of NF-kappa B by activated glucocorticoid receptors. *Molecular and cellular biology*. 1995;15(2):943-53.
282. Cadepond F, Gasc J-M, Delahaye F, Jibard N, Schweizer-Groyer G, Segard-Maurel I, et al. Hormonal regulation of the nuclear localization signals of the human glucocorticosteroid receptor. *Experimental cell research*. 1992;201(1):99-108.
283. Brkljačić J, Perišić T, Dundjerski J, Matić G. Interaction of rat renal glucocorticoid receptor with Hsp90 and Hsp70 upon stress provoked by mercury. *Journal of Applied Toxicology: An International Journal*. 2007;27(1):43-50.
284. Echeverria PC, Picard D. Molecular chaperones, essential partners of steroid hormone receptors for activity and mobility. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. 2010;1803(6):641-9.
285. Forneris F, Binda C, Vanoni MA, Mattevi A, Battaglioli E. Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. *FEBS letters*. 2005;579(10):2203-7.
286. Forsythe HL, Jarvis JL, Turner JW, Elmore LW, Holt SE. Stable association of hsp90 and p23, but Not hsp70, with active human telomerase. *Journal of Biological chemistry*. 2001;276(19):15571-4.
287. Bali P, Pranpat M, Bradner J, Balasis M, Fiskus W, Guo F, et al. Inhibition of histone deacetylase 6 acetylates and disrupts the chaperone function of heat shock protein 90: a novel basis for antileukemia activity of histone deacetylase inhibitors. *Journal of Biological Chemistry*. 2005;280(29):26729-34.
288. Khurana N, Bhattacharyya S. Hsp90, the concertmaster: tuning transcription. *Frontiers in oncology*. 2015;5:100.
289. Alfarouk KO. Tumor metabolism, cancer cell transporters, and microenvironmental resistance. *Journal of enzyme inhibition and medicinal chemistry*. 2016;31(6):859-66.
290. Hayes JD, Dinkova-Kostova AT, Tew KD. Oxidative stress in cancer. *Cancer cell*. 2020;38(2):167-97.
291. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature*. 2011;469(7330):343-9.
292. Simon JA, Kingston RE. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nature reviews Molecular cell biology*. 2009;10(10):697-708.
293. Cao Q, Wang X, Zhao M, Yang R, Malik R, Qiao Y, et al. The central role of EED in the orchestration of polycomb group complexes. *Nature communications*. 2014;5(1):1-13.
294. McCabe MT, Graves AP, Ganji G, Diaz E, Halsey WS, Jiang Y, et al. Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proceedings of the National Academy of Sciences*. 2012;109(8):2989-94.
295. Caputo VS, Costa JR, Makarona K, Georgiou E, Layton DM, Roberts I, et al. Mechanism of Polycomb recruitment to CpG islands revealed by inherited disease-associated mutation. *Human molecular genetics*. 2013;22(16):3187-94.
296. Voigt P, Tee W-W, Reinberg D. A double take on bivalent promoters. *Genes & development*. 2013;27(12):1318-38.
297. Kumar D, Cinghu S, Oldfield AJ, Yang P, Jothi R. Decoding the function of bivalent chromatin in development and cancer. *Genome research*. 2021;31(12):2170-84.
298. Almeida AM, Murakami Y, Baker A, Maeda Y, Roberts IA, Kinoshita T, et al. Targeted therapy for inherited GPI deficiency. *The New England journal of medicine*. 2007;356(16):1641-7.
299. Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes & development*. 2011;25(10):1010-22.
300. Karmodiya K, Krebs AR, Oulad-Abdelghani M, Kimura H, Tora L. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC genomics*. 2012;13(1):1-18.

301. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006;125(2):315-26.
302. Perillo B, Di Santi A, Cernera G, Galasso G, Pocsfalvi G, Castoria G, et al. Acetylation/methylation at lysine 9 in histone H3 as a mark of nucleosome asymmetry in human somatic breast cells. *Cell death discovery*. 2020;6(1):1-4.
303. Bernhart SH, Kretzmer H, Holdt LM, Jühling F, Ammerpohl O, Bergmann AK, et al. Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer. *Scientific reports*. 2016;6(1):1-18.
304. Margaritis T, Holstege FC. Poised RNA polymerase II gives pause for thought. *Cell*. 2008;133(4):581-4.
305. Gaertner B, Johnston J, Chen K, Wallaschek N, Paulson A, Garruss AS, et al. Poised RNA polymerase II changes over developmental time and prepares genes for future expression. *Cell reports*. 2012;2(6):1670-83.
306. De Raedt T, Walton Z, Yecies JL, Li D, Chen Y, Malone CF, et al. Exploiting cancer cell vulnerabilities to develop a combination therapy for ras-driven tumors. *Cancer cell*. 2011;20(3):400-13.
307. Perillo B, Ombra MN, Bertoni A, Cuozzo C, Sacchetti S, Sasso A, et al. DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. *Science*. 2008;319(5860):202-6.
308. Beck DB, Oda H, Shen SS, Reinberg D. PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes & development*. 2012;26(4):325-37.
309. Paquin KL, Howlett NG. Understanding the histone DNA repair code: H4K20me2 makes its mark. *Molecular Cancer Research*. 2018;16(9):1335-45.
310. Chen C, Wang Y, Wang S, Liu Y, Zhang J, Xu Y, et al. LSD1 sustains estrogen-driven endometrial carcinoma cell proliferation through the PI3K/AKT pathway via di-demethylating H3K9 of cyclin D1. *International journal of oncology*. 2017;50(3):942-52.
311. Li H-G, Tian W-H, Qin C-L, Ye R-R, Liu D-H, Liu H-W. Uhrf1 regulates H3K9me2 modification of mTOR to inhibit the effect of autophagy in myocardial ischemia-reperfusion injury. *Aging (Albany NY)*. 2021;13(7):9704.
312. De Zio D, Bordi M, Cecconi F. Oxidative DNA damage in neurons: implication of ku in neuronal homeostasis and survival. *International Journal of Cell Biology*. 2012;2012.
313. Srinivas US, Tan BW, Vellayappan BA, Jeyasekharan AD. ROS and the DNA damage response in cancer. *Redox biology*. 2019;25:101084.
314. Lesch BJ, Page DC. Poised chromatin in the mammalian germ line. *Development*. 2014;141(19):3619-26.
315. Choate LA, Danko CG. Poised for development. *Nature genetics*. 2016;48(8):822-3.
316. Wu Y-T, Tan H-L, Shui G, Bauvy C, Huang Q, Wenk MR, et al. Dual role of 3-methyladenine in modulation of autophagy via different temporal patterns of inhibition on class I and III phosphoinositide 3-kinase. *Journal of Biological Chemistry*. 2010;285(14):10850-61.
317. Liu Y-L, Yang P-M, Shun C-T, Wu M-S, Weng J-R, Chen C-C. Autophagy potentiates the anti-cancer effects of the histone deacetylase inhibitors in hepatocellular carcinoma. *Autophagy*. 2010;6(8):1057-65.
318. Prince HM, Bishton MJ, Harrison SJ. Clinical studies of histone deacetylase inhibitors. *Clinical cancer research*. 2009;15(12):3958-69.
319. Dokmanovic M, Marks PA. Prospects: histone deacetylase inhibitors. *Journal of cellular biochemistry*. 2005;96(2):293-304.
320. Steliou K, Boosalis MS, Perrine SP, Sangerman J, Faller DV. Butyrate histone deacetylase inhibitors. *BioResearch open access*. 2012;1(4):192-8.

321. Ruwende C, Khoo SC, Snow RW, Yates SN, Kwiatkowski D, Gupta S, et al. Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. *Nature*. 1995;376(6537):246-9.
322. Kung C, Hixon J, Kosinski PA, Cianchetta G, Histen G, Chen Y, et al. AG-348 enhances pyruvate kinase activity in red blood cells from patients with pyruvate kinase deficiency. *Blood, The Journal of the American Society of Hematology*. 2017;130(11):1347-56.
323. Xu JZ, Conrey A, Frey I, Gwaabe E, Menapace LA, Tumburu L, et al. A Phase 1 Dose Escalation Study of the Pyruvate Kinase Activator Mitapivat (AG-348) in Sickle Cell Disease. *Blood*. 2022.
324. Matte A, Federti E, Kung C, Kosinski PA, Narayanaswamy R, Russo R, et al. The pyruvate kinase activator mitapivat reduces hemolysis and improves anemia in a β -thalassemia mouse model. *The Journal of Clinical Investigation*. 2021;131(10).
325. Liu X, Chen Y, Zhang Y, Liu Y, Liu N, Botten GA, et al. Multiplexed capture of spatial configuration and temporal dynamics of locus-specific 3D chromatin by biotinylated dCas9. *Genome biology*. 2020;21(1):1-20.