Dissecting the role of CBP/P300-mediated acetylation in enhancer function

Nicole Jackson

St Cross College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2023

Abstract

Enhancers are major determinants of tissue-specific gene expression. Although it is not entirely clear how enhancers activate target genes, nucleosomes at enhancers tend to have high levels of H3K27ac, which is deposited by the lysine acetyltransferases (KATs) CBP and P300. However, the extent to which histone acetylation is important for enhancer function is not fully elucidated. Recent work from the Milne laboratory (WIMM, University of Oxford) demonstrated that anchoring the transactivation domain of MYB (MYB^{TA}) into a neutral chromatin environment can establish an enhancer-like element. Since MYB directly interacts with CBP/P300 we hypothesised this may be dependent upon the recruitment of CBP/P300 and subsequent deposition of histone acetylation. It was unclear whether this behaviour was unique to the genomic locus targeted and whether CBP/P300-mediated acetylation was necessary for this activity.

To investigate whether this effect was context dependent, I tethered MYB^{TA} to a novel, neutral environment on a distinct chromosome using the TetR system in mouse embryonic stem cells. Surprisingly, despite activating the original locus, the enhancer activity of MYB^{TA} did not extend to distal elements in this new context. This indicates that the enhancer forming ability of MYB is contingent upon the chromatin context within which it is bound. To test the necessity of CBP/P300 for MYB to generate an enhancer, I employed a proteolysis-targeting chimera (PROTAC), dCBP-1, to induce their degradation. This led to a reduction in H3K27ac genome-wide, coupled with transcriptional dysregulation. Despite a reduction in CBP/P300 and H3K27ac, MYB^{TA} was still able to activate distal elements, but less potently. This suggests that CBP/P300 are not necessary for the full transcriptional activity of MYB^{TA} at this locus.

Finally, to directly test the requirement of P300 KAT activity in enhancer function, I isolated the core domain of P300, containing the KAT domain and the bromodomain, and used this in the TetR chromatin targeting assay. This allowed me to examine the function of acetylation in the absence of the other CBP/P300 domains, while eliminating the complexity of transcription from nearby genes. The P300 core domain deposited H3K27ac and, interestingly, this was associated with activation of distal elements, analogous to an endogenous enhancer. Surprisingly, however, MYB^{TA} and $P300^{Core}$ activated different elements within the same genomic region, despite being targeted to the same binding site. This suggests that MYB may drive transcription and enhancer function through additional or alternative mechanisms to P300-mediated acetylation.

Acknowledgements

I would like to express my gratitude to my supervisors Tom Milne, Jim Hughes and Nick Crump for giving me the opportunity to pursue my PhD within their groups and for their continuous guidance, kindness and encouragement. The challenges posed by the COVID-19 pandemic added further complexity to my PhD journey, and I am grateful for their ongoing support and engaging discussions that kept me motivated. The collaborative environment at the WIMM has been invaluable. I would like to thank Philip Hublitz and the rest of the genome engineering team for providing the plasmids used in this thesis and Doug Higgs and the Higgs lab, in particular Caz, for welcoming me as an adopted member of their lab and inspiring discussions during lab meetings. Thank you to Chris Ott, for providing me with the PROTAC for Chapter 4.

To the Milne lab, past and present, thank you for fostering a supportive and friendly environment, offering extensive assistance at the bench and engaging in stimulating discussions during lab meetings and over coffee breaks. I extend my gratitude to Nick for his support as a supervisor, teaching me lab work from basics to complex techniques and creating an environment in which there are no silly questions. Thank you to Vassi and Ana, for their support through the ups and downs of PhD life, Alastair and Joe for their help with cloning and introducing me to the world of bioinformatics, especially Catherine for the ‘Seqnado user therapy’ group and finally, Grace, for her help with FACS plots and always helping me appreciate the little joys of life in the lab.

I am incredibly fortunate to have formed such close bonds in Oxford. I will never forget the joys of extended coffee/lunch breaks in an empty cafeteria after lockdown with Emily and Joe, despite the strange circumstances. Thanks to Susie and Ali and members of my cohort; Becky, Gemma, Ellie, James and Tani for their genuine and profound friendship. To all of my friends outside of the Oxford bubble, who I escaped to London with, especially Maya who navigated her PhD journey alongside mine.

Finally, I would like to thank my family, my mum and sister, grandma and late grandad for always reminding me I am braver than I believe, stronger than I feel and smarter than I think and for their ongoing love and support.

This work was made possible through funding from the Medical Research Council.

Declaration

I declare that this work is my own research, except where explicitly stated. This thesis has not been submitted, either partially or in full, for another qualification at this University, or for a qualification at any other institution.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3C	Chromatin Conformation Capture
ALL	Acute Lymphoblastic Leukaemia
AML	Acute Myeloid Leukaemia
BAC	Bacterial Artificial Chromosome
BPTF	Bromodomain PHD Finger Transcription Factor
BET	Bromodomain and Extraterminal
bp	Base Pair
BRD	Bromodomain
BRD4	Bromodomain containing 4
CBP	CREB-Binding Protein
CDB	ChIP Dilution Buffer
CDK7	Cyclin-Dependent Kinase 7
cDNA	Complimentary DNA
CFP1	CXXC Zinc Finger Protein 1
ChIP	Chromatin Immunoprecipitation
CO_2	Carbon Dioxide
CREB	cAMP-Response Element Binding Protein
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTCF	CCCTC-Binding Factor
CTD	C-terminal Domain
DBD	DNA Binding Domain
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DNase I	Deoxyribonuclease I
DPE	Downstream Promoter Element
DOT1L	Disruptor of Telomeric Silencing 1-like
DSG	Disuccinimidyl Glutarate
DSIF	DRB Sensitivity-Inducible Factor
DTT	1,4-Dithiothreitol
ECL	Enhanced Chemiluminescence
eGFP	Enhanced Green Fluorescent Protein
EB	Embryoid Bodies

eRNA	Enhancer Ribonucleic Acid
FA	Formaldehyde
FACS	Fluorescent Activated Cell Sorting
FBS	Fetal Bovine Serum
FOXA1	Forkhead Box A1
FS2	FLAG-StrepII
gRNA	Guide Ribonucleic Acid
HDR	Homology Dependent Repair
IDR	Intrinsically Disordered Region
INR	Initiator
ISWI	Imitation SWItch
JMY	Junction Mediating Protein
KAT	Lysine Acetyltransferase
kb	Kilobase
KDAC	Histone Deacetylase
KEE	H3K79me2/3 Enhancer Element
KMT	Lysine Methyltransferase
LB	Lysogeny Broth
LIF	Leukaemia Inhibitory Factor
LZ	Leucine Zipper
MCC	Micro Capture-C
mESC	Mouse Embryonic Stem Cells
MLL	Mixed Lineage Leukaemia
MTE	Motif Ten Element
N.P.	No Protein
NAP	Nucleosome Assembly Protein
nChIP-seq	Referenced Normalised ChIP-seq
NEAA	Non-Essential Amino Acids
NELF	Negative Elongation Factor
NR	Negative Regulatory
NURF	Nucleosome Remodelling Factor
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PHD	Plant Homeodomain
PIC	Preinitiation Complex
PcG	Polycomb Group Proteins
PROTAC	Proteolysis-Targeting Chimera

PRC	Polycomb Repressive Complex
iPSC	induced Pluripotent Stem Cell
PTEF-b	Positive Transcription Elongation Factor
PVDF	Polyvinylidene Fluoride
qPCR	Quantitative Polymerase Chain Reaction
RING	Really Interesting New Gene
RNA	Ribonucleic Acid
RNAPII	RNA Polymerase II
RT	Room Temperature
RT-qPCR	Reverse Transcriptase Quantitative Polymerase Chain Reaction
RTS	Rubinstein-Tabi Syndrome
Sas	Something About Silencing
-seq	Sequencing
SEM	Standard Error of the Mean
SDS	Sodium Dodecyl Sulfate
SOC	Super Optimal Broth with Catabolite Repression
SWI/SNF	SWItch/Sucrose Non-Fermentable
TA	Transactivation Domain
TAD	Topologically Associating Domain
TAF	TBP-Associated Factor
TBP	TATA-Binding Protein
TBS	Tris-Buffered Saline
TetO	Tet Operator
TetR	Tet Repressor
TF	Transcription Factor
Tris-HCl	Tris(hydroxymethyl)aminomethane Hydrochloride
TSS	Transcription Start Site
WT	Wild Type

Chapter 1: Introduction

1.1 Transcription

Transcription is a highly dynamic process that needs to be tightly controlled to uphold tissue-specificity and prevent aberrant gene expression. In eukaryotes, transcription is a stepwise process encompassing initiation, elongation and termination, which are guided by the RNA polymerase II (RNAPII) transcription machinery to produce ribonucleic acid (RNA). Tissue-specific transcription factors ensure specificity of initiation through binding to distinct deoxyribonucleic acid (DNA) sequences at transcription start site (TSS)-proximal sites (promoters) and TSS-distal sites (enhancers). Chromatin structure is important for packaging DNA, but it also plays an important role in gene regulation. While chromatin is used by eukaryotes to confine transcription to the appropriate regions of the genome, it can also act as an additional barrier that the RNAPII machinery needs to overcome for transcription to occur. How exactly the cell overcomes this transcriptional impediment is not well understood, but it is thought that histone modifications, including histone acetylation, may be involved. The role of histone acetylation in gene regulation has not been well explored in isolated model systems.

1.1.1 Organisation of DNA and Chromatin

Chromatin structure is important for both packaging of DNA into the nucleus and for regulating key cellular processes including transcription, DNA repair and replication. Further understanding the mechanisms by which chromatin structure is altered is crucial in dissecting the principles of transcription. To ensure DNA is packaged precisely in the nucleus of the cell, DNA is condensed into a tightly compacted structure through the association of histone proteins. 147bp of DNA is wrapped around an octamer of histone proteins, which consists of two copies of the canonical histones

(H2A, H2B, H3 and H4 (Kornberg, 1977; Luger *et al.*, 1997; Cutter and Hayes, 2015)). These structural units, known as nucleosomes, serve as the fundamental unit of chromatin. Each core histone comprises a central globular domain and unstructured histone tails that protrude from the nucleosome core (Luger and Richmond, 1998). Post-translational modifications of amino acid residues on both the core and the histone tails can influence the overall structure and accessibility of chromatin, thereby influencing transcription. Chromatin can be further compacted, through interactions between nucleosomes and this can restrict access to the DNA. This is referred to as heterochromatin, which is associated with low gene expression and appears darker and denser under the microscope (Tonnemacher, Eltsov and Jakob, 2020). Conversely, euchromatin exhibits a more open nucleosome configuration and higher gene expression, as evident by peaks in ATAC-seq and Deoxyribonuclease I (DNase I) hypersensitivity experiments (Lieberman-Aiden, Van Berkum, *et al.*, 2009).

The genome is compartmentalised into chromosomes, which occupy distinct regions of the nucleus. These are termed chromosome territories, where chromosomes interact with themselves more frequently than with other chromosomes (Cremer *et al.*, 1982; Cremer and Cremer, 2001). With the emergence of chromosome conformation capture (3C) technologies (reviewed in (Davies *et al.*, 2017)), it has been possible to investigate the structure of chromatin at high resolution. 3C technologies rely on stabilisation of chromatin contacts by chemical cross-linking (e.g. formaldehyde) and fragmentation of chromatin usually by restriction enzyme digest. Next, the cleaved DNA fragments are ligated back together, with ends close in space (but not necessarily close in sequence) forming ligation products so that the DNA sequence now reflects the 3D chromatin structure. These ligation junctions can be visualised and quantified by PCR or next-generation sequencing (Davies *et al.*, 2017). These techniques have allowed the investigation of the various higher order levels of chromatin structure. “A” and “B” compartments are at the mega-base (Mb) scale and were initially identified using Hi-C,

a 3C technique that generates contact maps showing the frequency of interaction between all parts of the genome in two dimensional heat maps (Lieberman-Aiden, Van Berkum, *et al.*, 2009). “A” compartments contain transcriptionally active, gene-rich chromatin and “B” compartments contain inactive, gene-poor chromatin, broadly correlating with euchromatin and heterochromatin (Lieberman-Aiden, van Berkum, *et al.*, 2009). One model that has been proposed suggests that spatial separation of “A” and “B” compartments may be driven by liquid-liquid phase-separation where low-affinity, non-specific interactions within similar compartments aggregate and exclude opposite compartments (Larson *et al.*, 2017).

Active genomic elements assemble into large, megabase sized chromatin interaction domains, named topologically associated domains (TADs), which were initially identified using Hi-C (Dixon *et al.*, 2012). The most widely accepted mechanism for how TADs are formed is through loop extrusion (Fudenberg *et al.*, 2016). This model posits that loop-extruding factors such as the ring-shaped cohesin molecule are loaded onto chromatin and move along the chromatin, forming a loop structure. This process is halted when cohesin reaches a CCCTC-binding factor (CTCF)-marked boundary element and the loop structure forms a hub containing active genomic elements. CTCF bind to these CTCF sites in convergent orientation and they work in an orientation-dependent manner (Dixon *et al.*, 2012; Nora *et al.*, 2012; Sexton *et al.*, 2012; Rao *et al.*, 2014). Within these TADs, enhancers and promoters are brought into close proximity and interact more frequently, which is thought to play a pivotal role in driving gene expression and contributes to enhancer-promoter specificity. CTCF functions to delimit TADs, preventing elements within one TAD from interacting with those in neighbouring TADs (de Wit *et al.*, 2015). When CTCF sites are inverted, chromatin loops do not form as effectively, which can lead to aberrant gene expression (Stolper *et al.*, 2023).

High resolution 3C techniques have been developed, including Capture-C, which have enabled the visualisation of robust enhancer-promoter interactions within TADs, containing active genes (Sanyal *et al.*, 2012; Hughes *et al.*, 2013; Mifsud *et al.*, 2015; Hay *et al.*, 2016; Hanssen *et al.*, 2017). Most recently, Micro-capture C (MCC) was developed which exhibits base-pair resolution and therefore has been pivotal in dissecting precise enhancer-promoter interactions (Hua *et al.*, 2021; Aljahani *et al.*, 2022).

1.1.2 Histone Modifications

Amino acid residues present on both the histone cores and tails, as well as on non-histone proteins, are subject to post-translational modification. These include but are not restricted to: acetylation of lysine residues, methylation of lysine and arginine residues and phosphorylation of serine and threonine residues (Bannister and Kouzarides, 2011). These are coined 'epigenetic' marks which impact gene regulation without changing the DNA sequence (Waddington C.H., 1942). Histone acetylation generally correlates with active chromatin whilst histone methylation exhibits both active and inactive properties depending on the histone and amino acid residue on which it is located. "Writers" are enzymes that deposit the histone mark, "erasers" are enzymes that remove the histone mark and "readers" are proteins that recognise the histone mark. H3K27ac is deposited by CREB-binding protein (CBP)/P300 lysine acetyltransferases (KATs) and is found enriched at active promoters and enhancers. H3K4me2/3 is a histone mark enriched at promoters, deposited by the lysine methyltransferases (KMTs) within the trithorax group, namely the Mixed Lineage Leukaemia protein 1/2 (MLL1/2, also known as KMT2A/D) (Milne *et al.*, 2002; Nakamura *et al.*, 2002; Dou, Y. *et al.*, 2006) and SETD1A/B via their SET domain. MLL1/2 directly target promoter regions, due to the presence of their CXXC domain which recognises unmethylated CpG islands (Milne *et al.*, 2010; Wang, Z. *et al.*, 2010).

SETD1A/B instead indirectly contact the promoter, by binding to CXXC Zinc Finger Protein 1 (CFP1) which contains a CXXC domain (Lee and Skalnik., 2005). In contrast, MLL3/4 (KMT2B/C) lack a CXXC domain and instead preferentially bind to enhancers via unclear mechanisms, depositing H3K4me1, an active mark enriched at enhancers (Heintzman *et al.*, 2007). Conversely, H3K9me3 and H3K27me3 are associated with repressed chromatin (Shen *et al.*, 2008). The extent to which these histone marks influence chromatin structure and transcription, rather than simply reflecting it, is still under investigation.

1.1.3 Transcription Factors and Co-activators

Transcription factors (TFs) bind DNA at regulatory elements, recruiting co-factors and facilitating the assembly of a multi-protein complex that includes the general transcription machinery. This process is pivotal for controlling gene expression in a temporal and tissue-specific manner, with expression of transcription factors often limited to specific cell types.

Transcription factors bind their cognate DNA sequence which are usually 4-12bp in length through their DNA binding domain (DBD) (Yuh *et al.*, 1994; Vaquerizas *et al.*, 2009). Given the large size of the genome compared to the short recognition sequence, it has long been debated how TFs are able to find their target sequences. Facilitated diffusion is a mechanism which may explain how TFs and specific DNA sequences make more frequent contact than one might expect from passive diffusion (Shvets, Kochugaeva and Kolomeisky, 2018). TFs are constantly sampling chromatin, engaging with all potential target sites by ongoing, weak, non-specific collisions. They display higher affinity for their cognate binding motif, where more binding events occur (von Hippel and Berg, 1989; Hager, McNally and Misteli, 2009; Klose *et al.*, 2013). Single molecule tracking experiments in mouse ES cells using the pluripotency TFs

Sox2 and Oct4 has provided evidence for facilitated diffusion (Chen *et al.*, 2014). Non-specific collisions with DNA lasted 0.75-0.9s whilst a stable binding event with the target DNA binding motif lasted 12.0-14.6s (Chen *et al.*, 2014).

Somatic mutations have the potential to alter DNA binding motifs, resulting in aberrant recruitment of TFs, impacting transcription. For example, a heterozygous somatic mutation found in a subset of T-ALL patients generates a binding motif for the TF MYB, generating a novel super-enhancer upstream of the *TAL1* oncogene. This drives overexpression of *TAL1*, contributing to cancer-specific gene expression (Mansour *et al.*, 2014).

Transcription factors usually have a separate transactivation (TA) domain, separate from the DNA binding domain (Keegan, Gill and Ptashne, 1986). One function of the TA domain is to recruit transcriptional co-activators, such as Bromodomain containing 4 (BRD4) and Mediator (Johnson *et al.*, 2002). Co-activators are proteins that associate with gene activation, but do not directly bind DNA like TFs. Instead, they are thought to be recruited to chromatin through interactions with TFs, chromatin or other co-activators. The TA domain is often intrinsically disordered (Staby *et al.*, 2017) and may be involved in the formation of phase separated condensates, driving non-specific interactions between co-activators (Boija *et al.*, 2018; Sabari *et al.*, 2018). Phase separation is a model which may explain how some specific co-activator complexes are assembled and how they interact with transcription factors and other co-activators. It is currently unclear to what extent co-activators are involved in transcription, but since they do not contain a DNA binding domain themselves it is thought that their association with TFs may explain how they are recruited to chromatin.

1.1.4 Transcription Initiation and Elongation

Initiation

Transcription initiation is a complex process that involves many dynamic protein-DNA and protein-protein interactions. TATA-binding protein (TBP), which is part of TFIID, initially recognises a core promoter element e.g. TATA box, a DNA sequence found upstream of the TSS of many but not all genes (Kim, Nikolov and Burley, 1993). Next, the pre-initiation complex forms, which is a large protein complex made up of TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH and RNAPII (Parvin and Sharp, 1993; Verrijzer *et al.*, 1995; Tsai and Sigler, 2000; Kostrewa *et al.*, 2009; He *et al.*, 2013). The promoter acts as an important docking site for the remaining members of the Preinitiation Complex (PIC) which recruits Mediator (Johnson *et al.*, 2002; Baek, Kang and Roeder, 2006; Esnault *et al.*, 2008; Seizl *et al.*, 2011; Jishage *et al.*, 2012; Plaschka *et al.*, 2015). Initially, short abortive transcripts are produced (Rougvie and Lis, 1988) and promoter clearance needs to occur for productive transcription. TFIIH, part of the RNAPII transcription machinery, has intrinsic helicase activity which helps to separate the DNA strands, forming a 'transcription bubble' where the single-stranded template sits in the RNAPII cleft (Fishburn *et al.*, 2015). The kinase activity of TFIIH, Cyclin-Dependent Kinase 7 (CDK7), phosphorylates serine 5 on the C-terminal domain (CTD) of RNAPII, allowing RNAPII to leave the promoter and elongation to occur (Feaver *et al.*, 1994).

Elongation

Initially, RNAPII transcribes short, abortive transcripts before transitioning to productive elongation, which has been suggested to act as a regulatory step. Positive Transcription Elongation Factor (P-TEFb) is involved in the progression to productive elongation (Peterlin and Price., 2006). P-TEFb has been shown to be recruited to its chromatin targets via interaction with BRD4 (Jang *et al.*, 2005; Yang *et al.*, 2005). P-

TEFb is comprised of CDK9 and Cyclin T1/T2. CDK9 phosphorylates DRB sensitivity-inducible factor (DSIF) and negative elongation factor (NELF), leading to dissociation of DSIF and conversion of NELF to an apparent positive elongation factor (Wada *et al.*, 1998; Yamaguchi *et al.*, 1999; Peterlin and Price, 2006). In addition to this, CDK9 phosphorylates serine 2 of RNAPII CTD. The combination of these phosphorylation events leads to pause release and transcription elongation. P-TEFb, along with ENL and AF4 form the "super elongation complex", which enhance transcription elongation (Takahashi and Yokoyama, 2020).

1.1.5 Impact of chromatin on transcription

Eukaryotic transcription is more complex than prokaryotic transcription, partly due to the presence of nucleosomes, which are essential for eukaryotic transcription but also act as additional obstacles to transcription. The transcription machinery therefore requires additional chromatin modifiers to overcome the nucleosomal barrier to allow productive transcription to occur. For example, Tang *et al* demonstrated that the presence p300 and p53 are essential for transcription of chromatinised DNA *in vitro* (Tang *et al.*, 2013). These additional factors include but are not limited to: "pioneer" transcription factors, co-activators, chromatin remodellers and histone chaperones. These will be discussed further in section 1.2: Gene Regulation in the Cell.

1.2 Gene Regulation in the Cell

A fascinating concept in the field of molecular biology is that a single genome can give rise to multicellular organisms with ~200 different cell types (Gartner and Hiatt, 2006). This is driven, in part, by cis-regulatory elements which, when bound by certain transcription factors, are selectively activated in different cell types to regulate unique subsets of the genome.

Cis-regulatory elements cooperate to orchestrate spatio-temporal gene regulation. There are two distinct regulatory elements in the mammalian genome associated with gene activation; promoters and enhancers. These are characterised by their underlying DNA sequence and additional defining properties. Promoters mark the site of transcription initiation and enhancers amplify transcription. Despite this distinction, it is evident that promoters and enhancers exhibit many overlapping features, making it difficult to distinguish them. For example, low-level transcription is observed at enhancers.

1.2.1 Promoters

Promoters are DNA sequences, typically 100bp-1kb, located upstream of the transcription start site (TSS). The DNA sequence of a promoter is defined by core promoter elements, including the TATA box, initiator (INR), motif ten elements (MTE) and downstream promoter element (DPE) (Benoist and Chambon, 1981; Vo ngoc *et al.*, 2020). However, not every element is present at all promoters in eukaryotes. Active promoters are marked by high levels of H3K4me3 and H3K27ac, in contrast to 'poised' promoters which are not actively transcribed (Heintzman *et al.*, 2007). General transcription factors bind to the promoter, facilitating the formation of the PIC along with loading of RNAPII, resulting in the production of stable, unidirectional transcripts (Sainsbury, Bernecky and Cramer, 2015). Many genes demonstrate multiple TSSs and may also be associated with more than one promoter (Forrest *et al.*, 2014).

1.2.2 Enhancers

Enhancers are canonically short (~100bp-1kb), non-coding DNA elements that regulate tissue-specific transcription independent of its position, distance and orientation (Banerji, Rusconi and Schaffner, 1981; Maniatis, Goodbourn and Fischer, 1987). Super enhancers represent a class of strong enhancer clusters, with around 5-12kb between

each individual enhancer (Blobel *et al.*, 2021). They are marked by at least one order of magnitude higher levels of H3K27ac and H3K4me1 compared to a typical enhancer and elevated binding of transcription factors and co-activators including Mediator and BRD4 (Hnisz *et al.*, 2013; Whyte *et al.*, 2013). Initially, it was thought that all enhancers within the super enhancer cluster contribute equally to their activity. Recent work has attempted to dissect whether super enhancers function in a syngeristic (Shin *et al.*, 2016), additive (Hay *et al.*, 2016) or redundant (Hörnblad *et al.*, 2021) fashion. Furthermore, facilitators have been identified within the alpha globin super enhancer, which have no intrinsic enhancer activity alone but maximise enhancer activity from the other enhancer elements within the cluster (Blayney *et al.*, 2023). However, this may not be universal and instead super enhancers are likely to function in a gene and cell-type dependent manner.

1.2.3 Enhancer Features and Function

Enhancers bind tissue-specific transcription factors and co-activators, allowing for distinct gene expression patterns not only across different cell types but also within the same cell type during development, differentiation and disease. Enhancers are associated with a range of chromatin features, which can be used for bioinformatic identification and to infer their activity. Putative active enhancers can be identified by: 1. Chromatin accessibility and histone variants; 2. Histone modifications, specifically high levels of H3K4me1 and H3K27ac; 3. High levels of transcription factor binding; 4. High levels of co-activators; 5. Enhancer ribonucleic acid (eRNA) transcription; 6. 3D contact with the target promoter (Heintzman *et al.*, 2007; Creyghton *et al.*, 2010).

Chromatin Accessibility and Histone Variants

Enhancers are regions that exhibit enhanced chromatin accessibility and undergo frequent nucleosome exchange (Jin and Felsenfeld, 2007; Jin *et al.*, 2009) to allow

access for DNA binding proteins. Histone chaperones, for example FACT, are thought to promote nucleosome exchange by binding to free histones (Gurova *et al.*, 2018). FACT has been shown to be associated with enhancers in leukaemia (Crump *et al.*, 2023). Chromatin remodellers use ATP hydrolysis to alter the nucleosome structure. This is achieved by sliding, evicting or modifying the space between nucleosomes (Li, Carey and Workman, 2007). Many nucleosome remodellers are recruited to active chromatin, including enhancers, through intrinsic recognition domains. For example, BPTF (bromodomain PHD finger transcription factor), a subunit of the NuRF (nucleosome remodelling factor), contains a bromodomain, a domain that recognises and binds to acetylated residues and a PHD finger that recognises and binds to H3K4me3 (Ruthenburg *et al*., 2011).

The histone H3 variant H3.3 is enriched at enhancers (Martire *et al.*, 2020). Interestingly, depletion of H3.3 from mouse embryonic stem cells (mESCs) results in no change in chromatin accessibility but instead a loss of H3K27ac (Li *et al.*, 2012). Additionally, H3.3 deposition is required to maintain high levels of H3K27ac at active enhancers and promoters in mESCs (Zhang *et al.*, 2020). Therefore, H3.3 may play a role in maintaining the epigenetic landscape at enhancers rather than a role in chromatin accessibility.

Histone modifications

It is widely recognised that H3K4me1 is found at active enhancers whilst H3K4me3 is present at active promoters (Heintzman *et al.*, 2007). These marks are often used to distinguish promoters and enhancers bioinformatically. However, it is important to note that H3K4me1 is enriched not only at enhancers, but also at promoters and in the gene body (Cui *et al.*, 2009), making it difficult to discriminate between promoters and enhancers using this modification.

H3K79me2/3 enhancer elements (KEEs) are a distinct class of enhancers that are dependent on H3K79me2/3 for their full function (Godfrey *et al.*, 2019). Disruptor of telomeric silencing 1-like (DOT1L) is a histone methyltransferase that deposits H3K79me2/3 (Feng *et al.*, 2002). When H3K79me2/3 is lost via DOT1L inhibition, KEE's lose their enhancer activity, including a loss of enhancer-promoter contacts and decreased transcription of target genes.

H3K27ac is found at both active promoters and enhancers (Creyghton *et al.*, 2010). However, there is a myriad of other histone acetylation marks also associated with active regulatory elements (Refer to section 1.3.1). Histone acetylation is generally associated with active chromatin, given its biophysical role in transcription. More specifically, histone acetylation neutralises the positive charge of lysine residues on histones, disrupting the interaction between histones and the negatively charged DNA backbone, altering the chromatin structure so that it is more permissible for transcription (Eberharter and Becker, 2002), for example by increasing accessibility for chromatin remodellers and transcription factors.

Histone acetylation and other marks can also act as a platform for reader proteins, which recognise and bind to these modifications, recruiting additional activities to contribute to enhancer function. Many key transcriptional related proteins contain a bromodomain, which recognises acetylation. Examples include the KAT CBP/P300 and GCN5/PCAF, thereby contributing to the establishment of a positive feedback loop, since they bind to the mark they deposit. Nucleosome remodellers often also contain a bromodomain, including BPTF, a subunit of the NuRF (Ruthenburg *et al*., 2011). The TFIID component TBP-associated factor 1 (TAF1), part of the basal transcription machinery also contains a bromodomain (Narita *et al.*, 2021). Bromodomain and extraterminal (BET) proteins including BRD4 which are involved in RNAPII elongation bind to acetylation via their bromodomain (Tafessu and Banaszynski, 2020). YEATS

domains are also readers of lysine acetylation (Zeng *et al.*, 2010; Zhao *et al.*, 2017) which are found in the transcription elongation factors ENL and AF9, linking acetylation to transcription elongation (Boyson *et al.*, 2021).

Transcription Factor Binding

A subset of transcription factors can bind directly to nucleosomal DNA and either independently or through recruiting additional chromatin remodellers, can create a permissive chromatin environment. These have been coined “pioneer” factors. Examples include GATA4, forkhead box A1 (FOXA1) (Cirillo *et al.*, 2002) and MYB (Fuglerud *et al.*, 2017). They recruit chromatin remodellers and co-activators, creating a more accessible chromatin environment for further activation.

Co-Activators

Co-activators are proteins that are correlated with gene activation but do not directly bind DNA like TFs. Co-activators, including histone modifying enzymes such as CBP/P300, are often enriched at enhancers. P300 contains multiple TF interaction domains which in combination help P300 to associate with chromatin (Ferrie *et al.*, 2023). P300 deposits histone acetylation to maintain an open chromatin structure (Eberharter and Becker, 2002). This, in turn, facilitates the recruitment of co-activator proteins including BRD4. BRD4 further recruits positive elongation factor b (P-TEFb), leading to CDK9-mediated phosphorylation of RNAPII, promoting productive elongation (Jang *et al.*, 2005; Yang *et al.*, 2005). BRD4 also associates with Mediator, a bridging complex that may play a role in enhancer-promoter interactions (Cho *et al.*, 2018). These coordinated processes not only function to stimulate transcription, but also contribute to the stabilisation of transcription complexes on DNA

eRNA transcription

An emerging area of research focuses on the short, bidirectional transcripts originating from the enhancer, referred to as enhancer RNAs (eRNAs) (Sartorelli and Lauberth, 2020), which play a role in enhancer function. Depletion of eRNAs correlates with a loss of accessibility at enhancers and a reduction in transcription of the MYOD1 locus (Mousavi *et al.*, 2013). eRNAs can directly contact CBP, stimulating its KAT activity and stabilising BRD4, to further prime enhancers for transcription (Bose and Berger, 2017; Bose *et al.*, 2017; Rahnamoun *et al.*, 2018). Although there is no direct evidence that loss of eRNAs contributes to loss of enhancer-promoter contacts, eRNAs may help to stabilise enhancer-promoter loops through interacting with Mediator (Lai *et al.*, 2013), CTCF (Hansen *et al.*, 2019) and cohesin (Li *et al.*, 2013).

Enhancer-Promoter Interactions

It is thought that promoters may be required to physically contact the enhancer for transcription initiation to occur (Furlong and Levine, 2018). However, there is some evidence that in some cases transcription from the promoter occurs independently of physical proximity to its enhancer. For example, using live-cell imaging in mESCs, it was shown that when *Sox2* is transcribed there is no physical contact with the *Sox2* control region (SCR) (Alexander *et al.*, 2019). Considerable research has focussed on elucidating how enhancer and promoter contacts are established and assessing the significance of these interactions for transcriptional regulation of target genes. Two mechanisms by which enhancer-promoter interactions may occur are loop extrusion and phase separation (discussed in Section 1.1.1). Recent work from the Milne group showed that dissolving phase condensates using a BRD4 inhibitor caused a reduction of transcription but no loss of enhancer-promoter contacts (Crump *et al.*, 2021). This suggests that phase separated condensates are not necessary for the maintenance of enhancer-promoter contacts and that transcription and enhancer-promoter contacts are separable events.

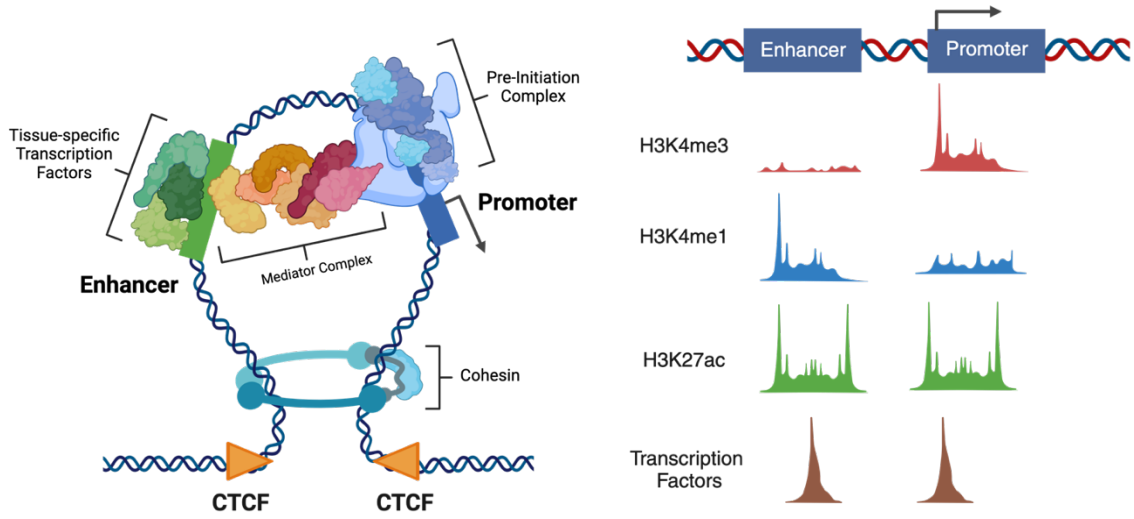


Figure 1.1 Principles of gene regulation

Left side: The enhancer and promoter are brought into close proximity by DNA looping with the aid of additional factors including the bridging complex, Mediator. Transcription factors and co-factors bound at enhancers and promoters help assemble the RNAPII machinery at the promoter.

Right side: Enhancers and promoters display similar and distinct features. Promoters are proximal to the TSS of the gene and marked with H3K4me3, H3K27ac and binding of TFs, which can be visualised by chromatin accessibility (ATAC-seq). Enhancers are distal to the TSS and marked with H3K4me1, H3K27ac and binding of TFs.

Although the precise mechanisms of enhancer function remain elusive, it is evident that the chromatin profile at enhancers appears to be a key aspect of their activity. In particular, the presence of CBP and P300 along with increased levels of histone acetylation are commonly observed at enhancers. The extent to which CBP/P300-mediated histone acetylation contributes to enhancer function is not yet fully elucidated.

Due to the complexity of enhancer function and the multitude of enhancer properties mentioned in this chapter, I chose to dissect one specific aspect of enhancer function - histone acetylation deposited by CBP and P300 lysine acetyltransferases. Given that

enhancer properties are intrinsically interconnected, modulating acetylation is likely to impact all of these features.

1.3. Lysine Acetyltransferases (KATs)

Lysine acetyltransferases (KATs) are epigenetic modifier enzymes that have the ability to acetylate lysine residues on both histone and non-histone proteins. Histone acetylation is highly correlated with active transcription however it is unclear whether this is solely correlative or if histone acetylation plays a functional role in transcription. It is thought that histone acetylation can alter chromatin structure to an open configuration by neutralising the positive charge of histone proteins and act as a scaffold for aiding assembly of the transcription machinery to chromatin, thereby positively influencing transcription (Eberharter and Becker, 2002).

1.3.1 Classes

Several eukaryotic proteins have intrinsic KAT activity. The first KATs to be identified were Gcn5 and Hat1 in yeast (yHpa2 and yHpa3), which form part of the GNAT family (Marmorstein and Zhou, 2014). Yeast also possess several members of the MYST family (yEsa1, ySas2, ySas3) (Marmorstein and Zhou, 2014). Finally, yRtt109 is most structurally similar to metazoan CBP/P300, sharing a conserved central core region for acetyl-CoA binding but instead yRtt109 acetylates H3K56 (Wang *et al.*, 2008) . KATs are evolutionary conserved from yeast to humans, but humans have several more KATs within each class, likely arising via gene duplication, summarised in table 1.1.

Table 1.1 Major lysine acetyltransferase family members in mammals and putative histone substrates

Table summarising the metazoan lysine acetyltransferase enzymes and their histone substrates as demonstrated in a mix of *in vitro* and *in vivo* experiments. Table adapted from (Shvedunova and Akhtar, 2022)

Family	Member	Histone Substrate
GNAT	HAT1	H4K5, H4K12 (Marmorstein and Zhou, 2014)
	GCN5 (KAT2A)	H3K9 (Grant *et al.*, 1999; Jin *et al.*, 2011)
	PCAF (KAT2B)	H3K9 (Grant *et al.*, 1999; Jin *et al.*, 2011)
P300/CBP	CBP (KAT3A)	H2AK4, H2AK5, H2AK7, H2AK9, H2AK11, H2AK13, H2BK5, H2BK11, H2BK12, H2BK15, H2BK16, H2BK20, H2BK21, H2BK23, H2BK24, H3K18, H3K27, H3K36, H4K5 (Weinert *et al.*, 2018)
	P300 (KAT3B)	H2AK4, H2AK5, H2AK7, H2AK9, H2AK11, H2AK13, H2BK5, H2BK11, H2BK12, H2BK15, H2BK16, H2BK20, H2BK21, H2BK23, H2BK24, H3K18, H3K27, H3K36, H4K5 (Weinert *et al.*, 2018)
MYST	TIP60 (KAT5)	H2AK5, H2AK15, H3K14, H4K5, H4K8, H4K12, H4K16 (Kimura and Horikoshi, 1998; Jacquet *et al.*, 2016)
	MOZ (KAT6A)	H3K9, H3K14, H3K23 (Doyon *et al.*, 2006; Voss *et al.*, 2012; Lv *et al.*, 2017)
	MORF (KAT6B)	H3K14, H3K23 (Doyon *et al.*, 2006; Simó-Riudalbas *et al.*, 2015)
	HBO1 (KAT7)	H3K9, H3K14, H4K5, H4K8, H4K12 (Doyon *et al.*, 2006; Mishima *et al.*, 2011)
	MOF (KAT8)	H4K5, H4K8, H4K16 (Radzisheuskaya *et al.*, 2021)

1.3.2 Structure

KATs harbour chromatin-binding domains that allows them to recognise and bind to modified histone tails. CBP/P300 and members of the GNAT family harbour a bromodomain, which recognises acetylated lysine residues (Josling *et al.*, 2012). Conversely, the MYST family have a chromodomain or plant homeodomain (PHD) fingers which bind methylated lysine residues (Yang and Ullah, 2007). This distinction suggests diverse recruitment mechanisms to chromatin and may influence where in the genome they exhibit their activity. Most KATs are found as part of large protein complexes, whereas CBP/P300 rely on domains that allow them to interact with a large number of other proteins. For example, CBP/P300 have a KIX domain, which mediates interactions with transcription factors, further explored in section 4.1.1. The KATs are

highly related within their respective families and there is the greatest diversity within the MYST family (Figure 1.2).

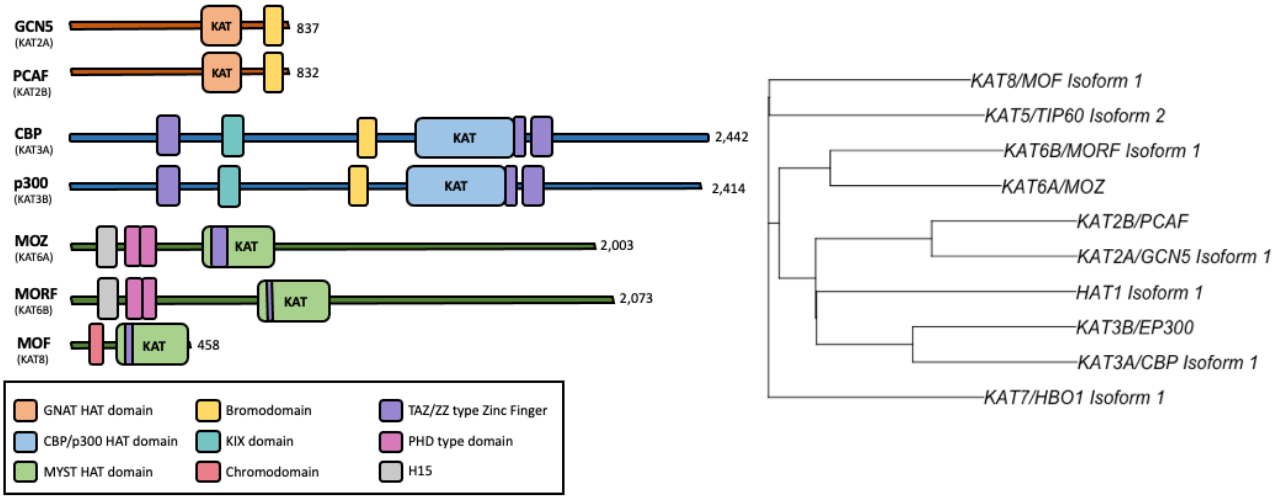


Figure 1.2 Lysine acetyltransferases structure and evolution

Left Side: Domains of human KAT proteins from three different families (GNAT, CBP/P300, MYST). Domains and sequence length information obtained from UniProt.

Right Side: Phylogenetic Tree of KAT proteins. The input is the full-length human protein. Analysis using the ClustalW algorithm in R – tidyverse, seqinr, msa and ape.

1.3.3 General Function

The role of KATs in gene regulation is complicated by the fact that different KATs within and between groups have been found to acetylate an ever-growing list of histone substrates, as well as non-histone proteins (Weinert *et al.*, 2018).

Histone and non-histone acetylation influence a variety of cellular and physiological processes, including transcription, autophagy, mitosis, differentiation and neural function (Shvedunova and Akhtar, 2022). Histone acetylation is a highly dynamic

process, opposed by KDACs that deacetylate histone and non-histone substrates. The balance between KATs and KDACs is important for coordinating a transcriptional response and for faithful cellular function (Halsall *et al.*, 2015). KDACs are more promiscuous, as one KDAC can deacetylate a myriad of acetylation marks on multiple residues, whereas KATs have a preference for particular lysine residues (Table 1.1). This specificity may change based on other proteins they complex with. For example, SAGA is a nucleosome remodelling complex that has intrinsic KAT activity through association with GCN5 which preferentially acetylates H3K9. However, GCN5 is also part of the NuA3 complex where it instead prefers to acetylate H3K14ac (John *et al.*, 2000).

Some lysine residues exhibit an acetylation-methylation equilibrium. For example, H3K27 can be modified by CBP/P300 to H3K27ac or by EZH2, a methyltransferase component of the polycomb repressive complex (PRC)2 repressor complex, to H3K27me1/2/3 (Ferrari *et al.*, 2014) Conversely, KDM6A, a histone demethylase, can return H3K27me3 to its unmodified state. Loss of function mutation of PRC2 results in a loss of H3K27me3 and an increase in H3K27ac by CBP/P300, plus an increase in gene expression at these targets (Hogg *et al.*, 2021). CBP/P300 and KDM6A may work together to maintain an acetylated state at H3K27 sustain an active transcriptional site.

1.3.4 Role of CBP/P300 in Gene Regulation

CBP/P300 are acetyl lysine ‘writers’, depositing the most well studied histone acetylation mark, H3K27ac. However, they can also acetylate other residues on H3 including H3K14, H3K18 and H3K23 (Weinert *et al.*, 2018) arising from the interaction between its ZZ-type zinc finger domain and the N-terminal of H3 (Zhang *et al.*, 2018). In addition, they can acetylate lysine 5 on H4 (Weinert *et al.*, 2018) (McManus and Hendzel, 2003) and many residues on H2A and H2B (H2AK4, H2AK5, H2AK7, H2AK9,

H2AK11, H2AK13, H2BK5, H2BK11, H2BK12, H2BK15, H2BK16, H2BK20, H2BK21, H2BK23, H2BK24) (Weinert *et al.*, 2018). More recent *in vivo* experiments have revealed that H2B acetylation is an increasingly important enhancer signature, which exclusively marks active enhancers rather than active promoters (Narita *et al.*, 2023). This mark, which is deposited by CBP/P300, may serve an important role in identifying enhancers, although its exact role in enhancer activity thus far is understudied.

In addition to acetylating histones, CBP/P300 also acetylate many non-histone proteins, regulating their function through changes in conformation, stability, hydrophobicity, cellular localisation or blocking the addition of different post-translational modification (Glozak *et al.*, 2005). Examples include MYB (Tomita *et al.*, 2000; Sano and Ishii, 2001), p53 (Gu and Roeder, 1997; Sakaguchi *et al.*, 1998; Liu *et al.*, 1999) and members of the basal transcription machinery (TFIIE and TFIIF) (Imhof *et al.*, 1997). This can alter the function of the protein, for example, the acetylation of GATA1 (Boyes *et al.*, 1998; Hung *et al.*, 1999) by CBP/P300 stabilises its association with chromatin, promoting erythroid-specific gene expression (Lamonica, Vakoc and Blobel, 2006; Lamonica *et al.*, 2011). CBP/P300 even acetylates itself (Thompson *et al.*, 2004), exhibiting an autoregulatory loop. The bromodomain of CBP/P300 acts as an acetyl lysine ‘reader’, recognising and binding to acetylated lysines deposited by itself or other KATs, creating a positive feedback loop to maintain an active chromatin state.

Further to this, CBP/P300 can utilise its additional domains in a structural role in transcription, for example functioning as a scaffold protein, by interacting with a multitude of transcription-related proteins (Figure 1.3). These include MYB (Zor *et al.*, 2004), MLL (Ernst *et al.*, 2001) cAMP-response element binding protein (CREB) (Goto *et al.*, 2002) , other KATs including PCAF (Ogryzko *et al.*, 1996) and over 400 other

proteins (Dancy and Cole, 2015). They also co-localise with Nanog, Oct4 and Sox2, important pluripotency factors in ES cells (Chen *et al.*, 2008) suggesting that CBP/P300 may be recruited by these master transcription factors to facilitate ESC-specific gene expression. Many TFs do not directly interact with the RNAPII machinery, therefore CBP and P300 can bridge them together (Chan and La Thangue, 2001). CBP and P300 interact with members of the basal transcription machinery: TFIIB (Felzien *et al.*, 1999), TBP (Yanazume *et al.*, 2003) and RNAPII itself through RNA helicase A (Nakajima *et al.*, 1997), indicating a role in stabilising the transcription machinery to their DNA targets.

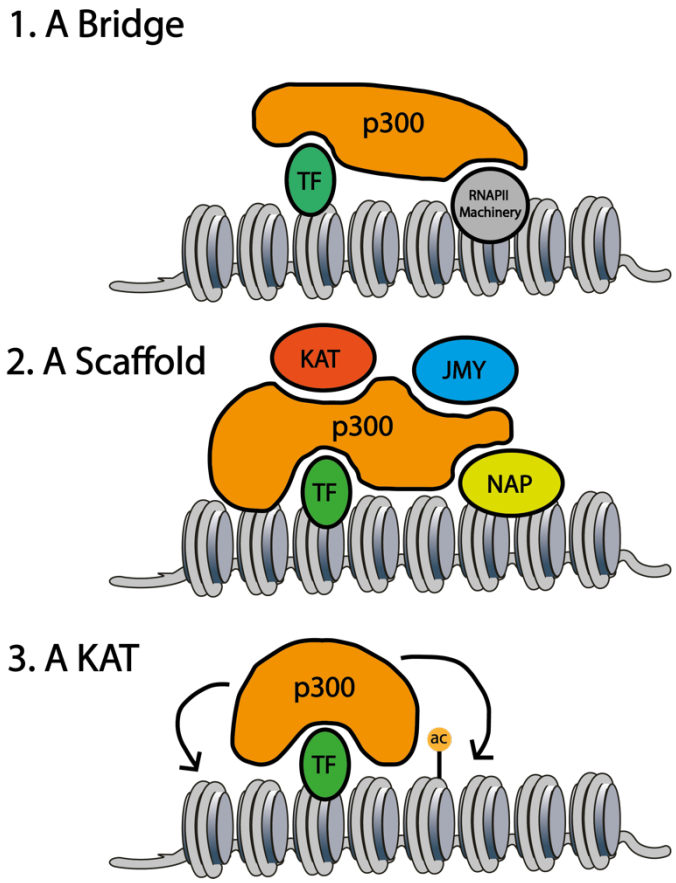


Figure 1.3 Models for the mechanisms of action of CBP and P300 on transcription

1. Acting as a bridge, CBP/P300 connect DNA binding transcription factors to RNAPII transcription machinery.

2. As a scaffold; CBP/P300 aid assembly of multisubunit complexes which stabilise the transcription machinery e.g. KATs (lysine acetyltransferases), junction mediating and regulatory protein (JMY) and nucleosome assembly proteins (NAP).
3. The intrinsic KAT activity of CBP/P300 acetylates histone and non-histone substrates.
Adapted from (Chan and La Thangue, 2001).

1.3.5 Role of CBP/P300 in Development and Disease

CBP and P300 are indispensable for normal development. Double knockouts of either gene, or double heterozygosity ($P300^{+/-}/CBP^{+/-}$) are embryonic lethal (Yao *et al.*, 1998; Tanaka *et al.*, 2000). Mutation of CBP and P300 are also implicated in human disease. Mutations in the CBP gene are notably associated with Rubinstein-Taybi Syndrome (RTS) in humans, a disease characterized by learning difficulties, craniofacial abnormalities, broad toes and thumbs, and a predisposition to haematological malignancies (Petrif *et al.*, 1995).

CBP/P300 associate with MYB, an important transcription factor and haematopoietic regulator. c-Myb (cellular Myb) is involved in gene regulation and required for normal development particularly of hematopoietic cells. v-Myb is the viral form originally identified in avian myeloblastosis and E26 leukaemia virus (Chen and Lipsick, 1993). Mice with homozygous mutations in the KIX domain of P300, which disrupts its interaction with MYB and other transcription factors, develop anaemia, B-cell deficiency, thymic hypoplasia, megakaryocytosis and thrombocytosis (Kasper *et al.*, 2002). Furthermore, the MYB-P300 interaction plays a pivotal role in leukaemic transformation by the fusion oncogenes AML1-ETO and MLL-AF9 *in vitro* likely by helping these cells to retain their self-renewal capacity leading to a differentiation block (Pattabiraman *et al.*, 2014). The *Booreana* mouse has a mutation in *c-Myb*, E308G, that interferes with its interactions with Cbp/p300. *Booreana* cells transduced with AML1-ETO9a or MLL-AF9 retroviruses fail to generate leukaemia upon transplantation

into irradiated recipients (Pattabiraman *et al.*, 2014). This observation led to the development of the inhibitory peptide which interferes with the assembly of the MYB:CBP/P300 complex. Treatment of AML cells with this MYBMIM peptide causes the redistribution of CBP/P300 away from oncogenic enhancers and the downregulation of MYB-dependent gene expression including *MYC* and *BCL2* (Ramaswamy *et al*., 2018) and colocalization of CBP/P300 with alternative TFs including RUNX1 (Takao *et al*., 2021). This overall leads to an induction of myeloid differentiation and apoptosis (Ramaswamy *et al*., 2018, Takao *et al*., 2021).

Additionally, CBP/P300 contribute to AML disease progression by acetylating MYB and AML1-ETO which leads to upregulation of genes such as Id1, p21 and Egr1 (Sun *et al.*, 2015), blocking differentiation and promoting self-renewal of cancer cells.

Point mutations in CBP/P300 are observed in both solid and haematological cancers, summarised in Figure 1.4. Mutations in the inhibitory really interesting new gene (RING) domain of CBP/P300 are observed in melanoma, endometrial and colorectal cancers (Forbes *et al.*, 2015; Attar and Kurdistani, 2017) and may be associated with hyperacetylation. Both CBP and P300 are also involved large-scale chromosomal translocations including: MLL-P300 (Ida *et al.*, 1997) and MLL-CBP (Borrow *et al.*, 1996) fusion proteins leading to acute myeloid leukaemia (AML). These fusion proteins retain the bromodomain and KAT domain, suggesting the acetylation activity is likely to be important for leukemic transformation in these cancers. There are examples of oncogenic translocations that fuse CBP to MOZ, a KAT in the MYST family, creating a ‘super-KAT’ with two KAT domains (Borrow *et al.*, 1996; Yang and Ullah, 2007).

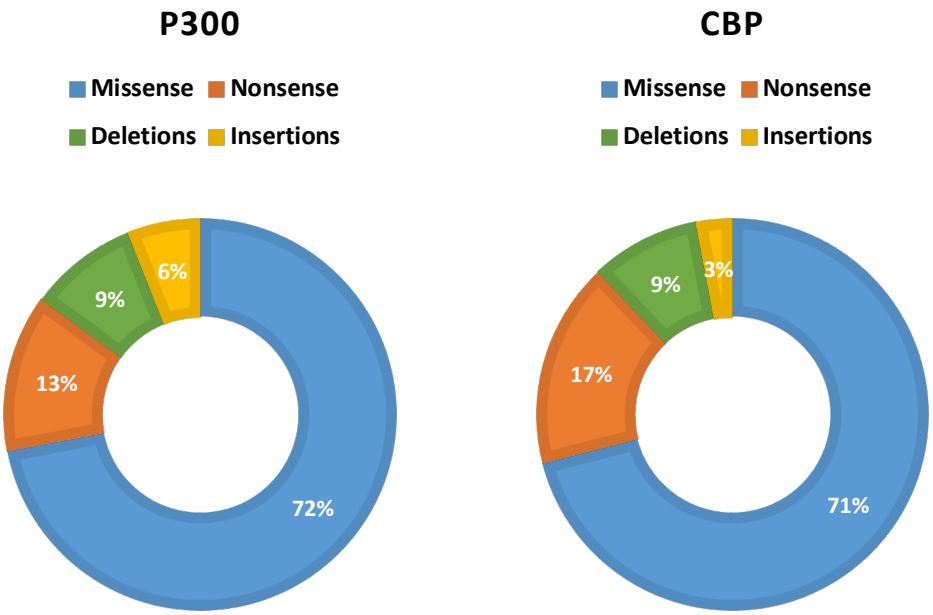


Figure 1.4 Mutation frequency of *P300* and *CBP* in human cancers

The most common mutation of *P300* and *CBP* in cancer is a missense mutation, followed by nonsense mutations, deletions and then insertions (Attar and Kurdistani, 2017).

CBP and P300 inhibitors

P300/CBP have more recently become desirable targets for therapeutic intervention in human cancers, with the aim to disrupt malignant acetylation and reverse oncogenic gene expression. P300/CBP dual inhibitors have been developed that target either the KAT domain or the bromodomain, with high selectivity and potency. Circumin was identified to specifically induce degradation of P300 and CBP, leading to a reduction of histone acetylation and acetylation of p53, a non-histone substrate (Balasubramanyam *et al.*, 2004). C646 was produced using virtual ligand screening, as an inhibitor that targets the KAT domains of P300 and CBP (Bowers *et al.*, 2010). However, this early drug exhibited issues with potency and selectivity.

More recently, the emergence of A-485 marked a significant advanced in the field. A-485 was designed to specifically target the catalytic KAT domain of both CBP and P300 with relatively high affinity, allowing for improved selectivity and potency in cancer treatment (Lasko *et al.*, 2017). However, this drug is not currently in clinical trials, but is used as a biological tool. Treatment with A-485 results in a genome-wide loss of H3K27ac with a concurrent reduction in cell-type specific gene expression including pluripotent-specific gene expression in induced pluripotent stem cells (iPSCs) (Pelham-Webb *et al.*, 2021) and oncogenic transcription in multiple myeloma (Hogg *et al.*, 2021).

Another strategy for inhibiting P300 and CBP activity is through targeting its bromodomain. GNE-049 (Raisner *et al.*, 2018), GNE-272 (Crawford *et al.*, 2016), GNE-781 (Romero *et al.*, 2017) and CBP-30 (Brooks *et al.*, 2021) are examples of early bromodomain inhibitors which specifically disrupt the bromodomain, leading to a loss of H3K27ac (Raisner *et al.*, 2018). CCS1477 (Inobrodib) is a more recent bromodomain inhibitor which is in Phase I/II clinical trials for AML, non-Hodgkin lymphoma and multiple myeloma and prostate cancer, demonstrating the therapeutic potential of targeting KAT proteins for cancer (Nicosia *et al.*, 2023).

Nonetheless, a significant issue with these drugs is the potential for patient relapse. Resistance and relapse is often driven by changes in enhancer usage and differential gene expression (Hogg *et al.*, 2021). For example, inhibiting BRD4 leads to upregulation of genes driven by enhancers independent of BRD4 (Rathert *et al.*, 2015). Understanding the mechanisms by which P300/CBP influence gene expression will be critical in identifying potential combination therapies to increase the efficacy of P300/CBP inhibitor monotherapy. In particular, the extent to which CBP and P300 are functionally involved in enhancer activity, and how this might impact cancer progression, is not yet fully established.

1.4 Summary and Aims

Previous work in our lab has shown that MYB^{TA} can establish enhancer activity and is associated with CBP/P300-mediated acetylation. What is not clear is whether the link between acetylation and transcription is solely correlative or whether acetylation is functionally involved in establishing enhancer-driven transcription. Therefore, utilising the *in vivo* TetR recruitment system, the following aims of this thesis are:

1. Is MYB^{TA} sufficient for establishing an enhancer in different genomic contexts?
2. Are CBP/P300 necessary for enhancer activity, including MYB^{TA}-driven enhancer function?
3. Is CBP/P300-mediated acetylation sufficient for establishing an enhancer?

Chapter 2: Materials and Methods

2.1 Cell Culture

E14 (TOT2N variant) mouse embryonic stem cells (mESCs) were a gift from Rob Klose (University of Oxford). mESCs were maintained at 37°C, 5% CO_2. Cells were grown in Dulbecco's Modified Eagle Medium (DMEM) (GibcoTM, A12860-01), supplemented with 10% Fetal Bovine Serum (FBS) (GibcoTM), 1X GlutamaxTM (GibcoTM, 35050061), 1X Non-Essential Amino Acids (NEAA) (GibcoTM, 11140050), 0.1µg/ml LIF (Leukaemia Inhibitory Factor) (Cambridge Bioscience, GFM200), 0.01M β mercaptoethanol (GibcoTM, 31350010). Cells passaged when 80-90% confluent every 2-3 days. Briefly, cells were washed with Phosphate Buffered Saline (PBS pH7.4, 10010-015, Gibco, Life Technologies). 0.05% trypsin (diluted in PBS) was added to the dish. Cells were incubated at 37°C for 3-5 minutes until the cells had detached. Trypsin was inactivated using fresh mESC medium and cells were plated into a new 10cm dish at a 1:6 dilution. Dishes were gelatinised prior to use using 0.1% gelatin (gelatin solution, Sigma-Aldrich, G1393), diluted in PBS and autoclaved in house. Cells were washed and trypsinised as above before any downstream applications.

Drosophila S2 cells were grown at 26°C and maintained at $0.5x10^6$-$2x10^6$ cells/ml as a semi-adherent monolayer, with passaging occurring every 3-4 days. S2 cells were derived from late stage (20-24 hours old) *Drosophila melanogaster* embryo (Schneider, I., 1972).

All cell lines used in this thesis were mycoplasma free and were tested using Lonza MycoAlert Mycoplasma detection kit, as per manufacturer's instructions.

2.1.1 Storage and Recovery of Cell Lines

Cells at 80-90% confluency were trypsinised, then centrifuged at 300g for 5 minutes and resuspended in fresh mESC medium plus 10% Dimethyl Sulfoxide (DMSO, Hybri-MAX®, #D2650, Sigma Aldrich). 1/3 10cm dish of cells (~7 million cells) were aliquoted to 500µl into cryovials and transferred to -80^{o}C in a freezing unit before being transferred to liquid nitrogen for long term storage.

To thaw cells, warm fresh mESC medium was added to the cells, then transferred to a 15ml tube, diluted in 10ml medium and centrifuged at 300g for 5 minutes and DMSO medium removed. Cells were resuspended in 10ml of fresh mESC medium and plated onto gelatinised 10cm plates to be maintained at 37^{o}C. The medium was changed the next day.

2.2 Molecular and Cloning Methods

2.2.1 Polymerase Chain Reaction

For polymerase chain reaction (PCR), KAPA (KK2601, Roche) or DreamTaq MasterMix (K1071, ThermoFisher) was used. 10-100ng of plasmid was used as template with primers at 10µM concentration. For gDNA, 2-4µl of gDNA extract was used as a template with primers at 10µM concentration. The PCR reaction programmes for each MasterMix are displayed in Table 2.1 and Table 2.2.

Table 2.1 KAPA PCR reaction programme

Step	Time (m.s)	Temperature (°C)
1	3.00	95
2	0.30	98
3	0.30	58-68
4	3.00	72 (Repeat steps 2-4, 25x-35x)
5	5.00	72
6	Hold	4

Table 2.2 DreamTaq PCR reaction programme

Step	Time (m.s)	Temperature (°C)
1	3.00	95
2	0.20	98
3	0.30	58-68
4	3.00	72 (Repeat steps 2-4, 25x-35x)
5	5.00	72
6	Hold	4

PCR product size and purity was confirmed by agarose gel electrophoresis. Fragments were purified with Qiagen PCR purification kit (28106), or Gel Extraction Kit (28704) and concentration/purity determined using Nanodrop (ThermoFisher Scientific).

2.2.2 Restriction Enzyme Digestion

For restriction enzyme digestion, 1µg of plasmid was incubated with the appropriate buffer and enzyme as per manufacturer's instructions (New England Biosciences). The digested product was isolated on an agarose gel and purified using Qiagen Gel Extraction Kit (28704).

2.2.3 Bacterial Transformation

Stellar Competent cells, an *E. coli* strain (636763, Takara), were used for transformation. 5µl DNA from the ligation reaction was added to 50µl of stellar competent cells and incubated on ice for 30 minutes. Bacterial cells were then heat shocked at 42^{o}C for 45 seconds and then cooled on ice for 2 minutes. 450µl of Super Optimal Broth with Catabolite repression (SOC) medium was added and the cells were incubated at 37^{o}C with horizontal shaking for 1 hour. After incubation, a range of volumes (20-200µl) were plated on Lysogeny broth (LB)-agar plates with the appropriate antibiotic to allow for formation of individual colonies. Plates were incubated overnight at 37^{o}C until visible colonies had formed. Individual colonies were picked into 3ml LB supplemented with 50µg/ml of appropriate antibiotic and incubated overnight at 37^{o}C, shaking. Plasmid DNA was purified using Qiagen MiniPrep Kit (27104), according to manufacturer's instructions. Plasmid DNA was sequenced by Sanger sequencing using appropriate primers. Plasmid DNA with the correct sequence were purified from 200ml bacterial cultures using the Qiagen MidiPrep Kit, according to manufacturer's instructions.

2.3 Western Blot

2.3.1 Salt Soluble Cell Extraction

1x10^{6} cells were washed in PBS and then lysed in 50µl BC300 (20 mM Tris- HCl, pH 7.5; 20% glycerol; 300 mM KCl; 5 mM EDTA) + 0.5% NP-40 (IGEPAL® CA-630, #I8896-100mL, Sigma) + 1:1000 protease inhibitor cocktail (Roche) and incubated on ice for 30 minutes. Samples were centrifuged at 20,000g at 4^{o}C and the supernatant isolated and adjusted to 1 x NuPAGE LDS sample buffer (NP0008) plus 1% β-mercaptoethanol. The pellet was retained for histone acid extraction if required. Samples were incubated at 95^{o}C to denature proteins, then stored at -20^{o}C prior to western blotting.

2.3.2 Histone Extraction

50µl 0.4M HCl was added to the pellet following BC300 treatment and incubated on ice for 30 minutes. The samples were centrifuged at 20,000g for 5 mins at 4°C and the supernatant isolated. 1ml of ice-cold acetone was added to the supernatant and incubated overnight at -20°C to precipitate proteins. The sample was centrifuged at 20,000g for 5 minutes at 4°C and acetone removed. The pellet was washed twice with ice cold acetone and again centrifuged at 20,000g for 5 minutes at 4°C and acetone removed. The pellet was airdried at room temperature (RT) and resuspended in 66µl 1 x NuPAGE LDS sample buffer (NP0008) plus 1% β- mercaptoethanol. Samples were incubated at 95°C to denature proteins and stored at -20°C prior to western blotting.

2.3.3 SDS PAGE

Pre-cast 4-12% acrylamide Bis-Tris gels (Novex, Life Technologies) was used to analyse proteins up to 200kDa, 12% Bis-Tris gels were used for histones; both were run in MOPS sodium dodecyl sulfate (SDS) Running Buffer (NuPAGE MOPS SDS Running Buffer (20x) Np0001, Novex, Life Technologies). 3-8% acrylamide Tris-Acetate gels were used for proteins above 200kDa with Tris-Acetate Running Buffer. Samples were denatured at 95°C prior to loading. 5µl of sample (~75,000 cells) was loaded for histones and 10µl (~150,000 cells) for other proteins.

Gels were run at 180V for 40 minutes for histones, 1 hour for proteins up to 200kDa and 2 hours for proteins above 200kDa. Molecular weights were estimated using the Precision Plus Protein Kaleidoscope protein molecular weight ladder (Biorad 110375).

2.3.4 Western Blotting

Proteins were transferred to a polyvinylidene fluoride (PVDF) membrane (Immobilon®-P Transfer Membrane, #IPVH00010, EMD Millipore) in transfer buffer (192mM glycine;

25mM Trizma® base; 10% methanol), with the addition of SDS for proteins above 200kDa. Transfer was conducted at 100V for 1 hour at 4°C, or 25V overnight at 4°C for proteins above 200kDa.

Membranes were stained with Ponceau-S stain (Sigma-Aldrich P-7170-1L) to check for presence of protein and then washed in methanol. The membrane was air-dried and then incubated in 5% milk and primary antibody at 4°C overnight for proteins and 1 hour at room temperature for histones (see Table 2.3 for antibodies). The membrane was washed three times in Tris-buffered saline (TBS)/0.05% Tween 20 (0.05% TBS-Tween) for 5 minutes each. Membranes were incubated in 5% milk with the appropriate HRP-conjugated secondary antibody for 1 hour at room temperature. The membranes were washed again three times in 0.05% TBS-Tween for 5 minutes. The membrane was then incubated with enhanced chemiluminescence (ECL) (ECL Prime Western Blotting Detection Reagent, #RPN2232, Amersham, GE Healthcare). Solutions A and B were mixed at a 1:1 ratio and applied to the membrane for 10 seconds. Excess solution was dabbed away, and the membrane visualised in the Chemi-Doc MP Imaging System (Bio-Rad, 12003154).

Table 2.3 Antibodies for Western Blot and ChIP

Antibody	Company/Catalog Number	Host	Working Concentration
CBP	Bethyl/A300-362A	Rabbit	1:2000 (WB
CBP	Santa-Cruz BioTech sc-369	Rabbit	1:500 (ChIP)
GAPDH	Bethyl/A300-641A	Rabbit	1:10000 (WB)
HA	Cell Signalling 3724T	Rabbit	1:20000 (WB), 1:500 (ChIP)
H3	Abcam/ab1791	Rabbit	1:50000 (WB)
H3K27ac	Diagenode/C15410196	Rabbit	1:20000 (WB), 1:1000 (ChIP)
P300	Bethyl/A300-358A	Rabbit	1:5000 (WB), 1:500 (ChIP)
TetR (FS2)	Klose Lab	Rabbit	1:1000000 (WB), 1:500 (ChIP)
Vinculin	Abcam/129002	Rabbit	1:5000-1:10000 (WB)
Anti-Rabbit HRP 2°	Sigma-Aldrich/A6667	Goat	1:5000-1:20000 WB)

2.4 Generating Cell Lines

2.4.1 Transient Transfection of TetR plasmid

Cells were grown to 60-70% confluency in 10cm plates. 37µl Lipofectamine 2000 (Invitrogen, Life Technologies, 11668019) was diluted in 250µl OptiMEM medium (#51985034, Gibco, Life Technologies) in an Eppendorf tube. 14µg of plasmid DNA was diluted in 250µl OptiMEM in a separate tube. The contents of both tubes were combined and incubated for 5 minutes at room temperature. Cells were replenished with fresh mESC medium and then the mixture was added dropwise with mixing, then cells were incubated at 37°C. 24 hours post-transfection, cells were Fluorescent Activated Cell Sorted (FACS) based on the presence of enhanced GFP (eGFP).

2.4.2 FACS sorting and gating

Cells were single cell sorted on the presence of eGFP+ (enhanced green fluorescent protein) using BD FACSAria™ Fusion Flow Cytometer into 15ml falcon tubes. Cells

were gated and selected based on the following phenotypes: FSC-A:SSC-A, FSC-A:FSC-H and 488-670/30-A:488-530/30-A (GFP+). Cells were then harvested for downstream applications in the case of transient transfection experiments or plated at low density to create stably expressing cells lines, as in section 2.4.4.

2.4.3 Creating of Stable mESC Lines by Random Integration to express TetR fusion proteins

mESCs were grown to 60-70% confluency in 10cm plates. In order to obtain consist expression of TetR fusion proteins between replicates, where possible, stable cells lines were generated expressing the TetR fusion protein. Transfection was using Lipofectamine 2000 refer to section 2.4.1. The TetR plasmids used for generating stable cell lines are in the Appendix. 24 hours post-transfection the cells were replenished with fresh mESC medium supplemented with 1µg/ml puromycin (Sigma) to select for cells stably expressing the plasmid. After another 24 hours, cells were plated at low density (~2,000-10,000 cells per 10cm plate) in mESC medium supplemented with puromycin to allow for isolation of colonies derived from a single clone. After 7-10 days, once colonies had formed, distinct colonies were picked into wells of a gelatinised 96-well plate (Corning, Life Sciences, #3595) with puromycin-supplemented mESC medium. Cells were expanded and sequentially transferred to wells of a 24-well, 12-well and 6-well plate (Corning, Life Sciences, #3526, #3513, #3516 respectively) for screening, maintenance and downstream applications. Screening was conducted by western blotting to detect the stable expression of the construct that was transfected (Refer to Section 2.3).

2.4.4 Creating of Stable Cell Lines by CRISPR-Cas9-mediated Homology Dependent Repair to edit specific loci

In collaboration with Philip Hublitz and Chris Babbs (Genome Engineering, Weatherall Institute of Molecular Medicine (WIMM)), three guide RNAs (gRNA) targeting each

locus to be edited were designed. Genome Engineering services (WIMM) cloned these guides into pX458 (addgene, #48138,(Ran *et al.*, 2013)) and conduct a Surveyor assay to assess gRNA quality. Based on this, one gRNA was chosen for transfection (Table 2.4). pX458 contains genes encoding Cas9 from *Strep. Pyogenes* and eGFP under the control of constitutively active promoters (Ran *et al.*, 2013). The sequence to be inserted was cloned into a pMK-RQ plasmid backbone (GeneArt, ThermoFisher) with flanking homology arms (~500bp).

Table 2.4 guide RNA target sequences

Target	Guide Sequence (5'-3')
Neutral Region on Chromosome X	GGAGAGTAGTGGCCCAACTCT
Cbp (C-terminus)	GTCCCTGGTTGGTGATACCACGG

eGFP-positive cells were isolated by FACS, indicating the presence of the guide plasmid. eGFP+ cells were plated at low density (2,000-10,000 cells per 10cm plate) to allow for isolation of colonies derived from a single clone. After 7-10 days, once colonies had formed, distinct colonies were picked into wells of a gelatinised 96-well plate (Corning, Life Sciences, #3595) with fresh mESC medium. Once the majority of wells were confluent, cells were split between two plates, with one to maintain the clones and the other to extract DNA and screen by PCR for the integration of the desired stretch of DNA (Refer to 2.4.5).

2.4.5 Genotyping

Cells were transferred to a 96-well PCR plate using a multichannel pipette and lysis buffer was added to each well (Tris-HCl, Triton X-100, Proteinase K). The plate was vortexed using a plate vortexer for 30 seconds, then centrifuged at 1000rpm for 5 minutes. The lysate was transferred to a fresh 96-well plate and stored at 4^{o}C.

The DNA was used as a template for PCR using primers in Table 2.5. The PCR product was run on a 1% agarose gel in 1X TAE buffer for 40 minutes-1 hour. Pooled clones were screened with internal primers (primer ID No. 2 and 3), where one primer binds inside the inserted sequence and one in the endogenous DNA. Potential positive clones were further screened with spanning primers, which bind outside of the homology arms (primer ID No. 1), to ensure the presence of homozygous clones (in the case of *Cbp* degron) and to ensure no contamination of wild type, which would result in the presence of a smaller PCR product.

Table 2.5 PCR primers used for genotyping of edited mESCs

ID	Target	Type	Primer	Sequence
ChrX:1	Chr X, Neutral Region, TetO array	Spanning	Forward	GTGCTGCAATGAATGTCAGAAGG
			Reverse	GGAAGGGAGAGAGACATGGAGAG
ChrX:2	Chr X, Neutral Region, TetO array	Internal (LH)	Forward	GTGCTGCAATGAATGTCAGAAGG
			Reverse	ACCTGCCCACATCTCTATCAC
ChrX:3	Chr X, Neutral Region, TetO array	Internal (RH)	Forward	AGACCAAATCTCCTGGGTAAGAG
			Reverse	GGAAGGGAGAGAGACATGGAGAG
Cbp:1	Cbp (C terminus)	Spanning	Forward	TTGGAGATGCCCACAGAGTTCAC
			Reverse	CCCCAACATGACAAACATGAATCC
Cbp:2	Cbp (C terminus)	Internal (LH)	Forward	TTGGAGATGCCCACAGAGTTCAC
			Reverse	GCCACTCTCGTCTTCGATGTGG
Cbp:3	Cbp (C terminus)	Internal (RH)	Forward	AACTTAAAGGGCTTGTTTCTGTCGC
			Reverse	CCCCAACATGACAAACATGAATCC

To ensure the edit occurred in the correct place with no additional alterations to DNA sequence, DNA from positive clones was purified from mESCs using the Qiagen DNA Blood and Tissue kit, according to manufacturer's protocol and used directly as a template for PCR amplification using spanning primers in table 2.5. The PCR product was run on a 1% agarose gel in 1X TAE buffer for 40 minutes-1 hour and gel extracted using QIAquick Gel Extraction Kit (Qiagen) and sequenced by Sanger sequencing.

2.5 Gene perturbation experiments

2.5.1 dTAG and PROTAC treatment

Cells were seeded at ~1x10^6 cells/ml in 6-well plate or 1x10^7 cells/ml in 10cm plates. Cells were treated with dTAG-13 (Tocris) at 1μM or PROTAC (dCBP-1; the kind gift from Dr Chris Ott) at 500nM final concentration or 0μM (DMSO only control). After various hour time-points, cells were harvested for downstream applications.

2.6 ChIP Methods

2.6.1 Chromatin Immunoprecipitation (ChIP)

Fixation

Cells were harvested by centrifugation (300g, 5min) followed by PBS wash. 1x10^7 cells were used for ChIP-qPCR and 3x10^7 for ChIP-seq (fixed as 1x10^7 aliquots). For double fixation (transcription factors), cells were resuspended in 2mM Disuccinimidyl Glutarate (DSG, 80424-50MG-F, Sigma) in PBS and rotated at room temperature for 30 minutes. Cells were spun at 10,000g for 1 minute and DSG removed. The cell pellet was then resuspended in 1ml 1% formaldehyde (FA) in PBS and rotated for a further 30 minutes at room temperature. For single fixation (histone modifications), cells were only resuspended in 1ml 1% FA and rotated at room temperature for 10 minutes. Cells were spun again at 10,000g for 1 minute and FA removed. Finally, the pellet was washed in 1ml PBS and pellets snap-frozen on dry ice.

Lysis, Sonication and Pre-Clear

For referenced normalised ChIP-seq (nChIP-seq), fixed Drosophila S2 cells were added to fixed ESCs prior to sonication at a ratio of 1:5 S2:ESC. After sequencing, reads were adjusted based on the ratio of dm6:mm10 reads in the input and IP samples for each condition.

Fixed pellets were thawed on ice and resuspended in 120µl lysis buffer (1% SDS; 10mM EDTA; 50mM Tris-HCl pH8) plus 1:200 protease inhibitor cocktail and syringe-passaged through a 27-gage (27G) needle. After 10 minutes on ice, the lysate was transferred to Covaris micro-tubes (microTUBE AFA Fiber Pre-Slit Snap_CAP 6x16mm, 520045, Covaris Inc) and sonicated in a Covaris sonicator (S220, Covaris Inc.) with the conditions in Table 2.6.

Table 2.6 Sonication conditions for ChIP

Fixation	Peak Power	Duty Factor	Cycles per Burst	Total Time (seconds)
Double	75	25%	500	600
Single	75	25%	500	300

After sonication, the sonicated chromatin was transferred to Eppendorf tubes and spun for 10 minutes at 20,000g at 4°C to remove insoluble chromatin and nuclear debris. The supernatant containing soluble chromatin was then diluted 10X with ChIP dilution buffer (CDB) (0.01% SDS, 1.1% TritonX-100 (X100-100ML, Sigma), 1.2mM EDTA, 16.7mM Tris-HCl pH8, 167mM NaCl, PIC). 5µl of a 1:1 mix of Protein A/G Dynabeads (10001/4D, Novex, Life Technologies) resuspended in CDB was added for pre-clearing and tubes were rotated at 4°C for 20 minutes. Then chromatin was spun at 1000g for 2 minutes and the pellet discarded on a magnet to remove chromatin that binds non-specifically to the dynabeads. 50µl of the chromatin was used as input material. 1ml of chromatin was used for immunoprecipitation and 2µg of antibody was added and rotated overnight at 4°C (See Table 2.3 for antibody dilutions).

Washes and Elution

The next day, 15µl of Protein A/G Dynabeads are added to each 1mL IP aliquot and rotated at 4°C for 3-5 hours to bind to the antibody. Beads were washed twice with

RIPA buffer (50mM Hepes-KOH pH7.6, 500mM LiCl, 1mM EDTA, 1% NP-40, 0.7% Na-Deoxycholate, ddH2O) using a magnetic stand. Beads were washed in 1ml TE (10mM Tris-HCl pH8, 1mM EDTA, 50mM NaCl) and spun at 1000g for 2 minutes. The supernatant was removed, and beads eluted in 105µl elution buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl pH8) at 65°C for 30 minutes. The 100µl eluant was then isolated and chromatin was then treated with 1µl 500µg/ml RNase (EN0531, ThermoScientific) (at 37°C for 30 min) and 1µl 20mg/ml Proteinase K (EO0491, ThermoScientific) and incubated overnight at 65°C to reverse the DNA-Protein crosslinks. Input material was diluted to 100µl with elution buffer and treated along the IP samples.

DNA was purified using the Zymoclean PCR purification kit (Zymoresearch) For ChIP-qPCR DNA was eluted in 200µl water. For ChIP-seq, DNA was eluted in 53µl water.

2.6.2 ChIP-qPCR and Analysis

SYBR green (Fast SYBR® Green Master Mix (#4385612, Applied Biosystems)) was used for qPCR analysis. The reaction mix below (Table 2.7) was set up in individual wells of a 96-well PCR plate, with each reaction conducted in duplicate (MicroAmp Fast Optical 96-well Reaction Plate, 4346906, Applied Biosystems).

Table 2.7 ChIP-qPCR Reaction Mix

Component	Volume (µl)
10µM Forward Primer	0.5
10µM Reverse Primer	0.5
Fast SYBR Green Master Mix (4385612, Applied Biosystems)	10
Nuclease-Free Water	4.5
DNA	5

PCR plates were sealed with MicroAmp Adhesive Film (4311971, Applied Biosystems) and briefly centrifuged. QuantStudio 3 Real-time PCR machine (Applied Biosystems) was used for the PCR reaction with the following RT- qPCR program. DNA was quantified by qPCR, with ChIP samples normalised using input chromatin.

ChIP signal was calculated as a percentage of input using the following calculation:

$$\Delta Ct = Ct(Input) - Ct(Sample)$$

$$\%Input = 2^{\Delta Ct} \times 5$$

2.6.3 ChIP-seq Library Preparation and Sequencing

50ng of DNA was used as input for library preparation. Next generation sequencing libraries were generated using the Ultra II DNA library preparation kit (E7645, NEB) and NEBNext Multiplex oligos for Illumina (E7335L/E7500L). Libraries were then sequenced in-house on a NextSeq 500 (Illumina), 75bp paired-end sequencing or by Novogene on a NovaSeq 6000, 150bp paired-end sequencing.

2.6.4 Quality Control

Library quality was assessed using the DNA 1000 high sensitivity Screen Tape (Agilent 2200 TapeStation system) and quantified by RT-qPCR (KAPA library quantification kit KK4828).

2.6.5 ChIP-seq Analysis

I used the SeqNado (https://pypi.org/project/seqnado) pipeline on the WIMM CCB server developed in the Milne lab for next generation sequencing analysis. SeqNado maps and post-processes mapped reads including deduplication and blacklist filtering. SeqNado generates bigwigs for visualisation and performs peak calling with Macs2 and

Lance-O-Tron. I used deeptools 'computeMatrix' and 'plotHeatmap' functions to plot metaplots and heatmaps for ChIP-seq signal. For nChIP-seq, reads were adjusted based on the ratio of dm6:mm10 reads in the input and IP samples for each condition.

2.7 Gene expression analysis by RT-qPCR

2.7.1 Total RNA extraction

$1x10^6$ cells were harvested by centrifugation (300g x 5min) followed by PBS wash. Total RNA was extracted using the RNeasy Mini Kit (Qiagen 74104/74106) followed by DNase treatment using Turbo DNA-free Kit (Invitrogen, AM1907). RNA concentration was measured using the Nanodrop (ThermoFisher Scientific).

2.7.2 cDNA synthesis

1µg of RNA in 8µl was incubated with 1µl random hexamers (3µg/µl, 58875), 1µl dNTPs (10mM 18427-013) and Nuclease-free H_2O at 65^oC for 5 minutes. The complementary DNA (cDNA) was synthesised from the RNA with the cDNA synthesis reaction mix in Table 2.8 and the cDNA synthesis reaction in Table 2.9.

Table 2.8 cDNA synthesis reaction mix

Component	Volume (µL)
5X First Strand Buffer	4
Nuclease-free H_2O	2
1,4-Dithiothreitol (DTT)	2
RNaseOUT 00000840)	1
SuperScript III (200U/µl, 56575	1

Table 2.9 cDNA synthesis reaction

Temperature (ºC)	Time (m.s)
25	5.00
50	60.00
70	15.00

2.7.3 RT-qPCR

cDNA samples were diluted 10x for Reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) analysis. The reaction mix below was set up in individual wells of a 96-well PCR plate (MicroAmp Fast Optical 96-well Reaction Plate, 4346906, Applied Biosystems) SYBR green was used for RT-qPCR analysis (Table 2.10).

Table 2.10 RT-qPCR reaction mix

Component	Volume (µL)
10µM Forward Primer	0.5
10µM Reverse Primer	0.5
Fast SYBR Green Master Mix (4385612, Applied Biosystems)	10
Nuclease-Free Water	4.5
DNA	5

PCR plates were sealed with MicroAmp Adhesive Film (4311971, Applied Biosystems) and briefly centrifuged. QuantStudio 3 Real-time PCR machine (Applied Biosystems) was used for the PCR with the RT-qPCR programme in Table 2.11.

Table 2.11 RT-qPCR Programme

Temperature (ºC)	Time (m.s)	Cycles
95	3.00	1
98	0.20	30
65	0.15	
72	1.00	
72	2.00	1

SYBR Primers were designed using IDT primer quest tool, primers used for RT-qPCR are displayed in the appendix.

2.8 RNA-seq methods

2.8.1 PolyA+ and PolyA- RNA-seq library preparation

$1x10^6$ cells were harvested by centrifugation (300g x 5min) followed by PBS wash. Total RNA was extracted using the RNeasy Mini Kit (Qiagen 74104/74106) followed by DNase treatment using Turbo DNA-free Kit (Invitrogen, AM1907). RNA concentration was measured using the Nanodrop (ThermoFisher Scientific) and RNA quality assessed using TapeStation. RNA depletion was conducted using the RiboMinus™ Eukaryote System v2 (A15020) and RiboMinus™ Eukaryote Kit v2 (A15026). PolyA+ and PolyA- RNA were separated using the NEBNext Poly(A) mRNA magnetic isolation module (E7490, NEB). Briefly, 1µg of ribodepleted RNA was incubated with oligo dT beads in 1X RNA binding buffer at 65°C for 5 minutes and then cooled to 4°C. Material not attached to the beads containing the PolyA- fraction was kept and precipitated. The beads were washed twice in wash buffer (E7490, NEB) and resuspend in 50µl Tris-HCl and incubated at 80°C for 2 minutes to elute the RNA. Beads were added again with 1X RNA binding buffer and RNA was rebound to the beads in an incubation at room temperature for 5 minutes.

Next generation sequencing libraries of the PolyA+ and PolyA- fraction were generated using the NEBNext Ultra II Directional RNA library preparation kit (E7760, NEB). Library quality was assessed using the DNA 1000 high sensitivity Screen Tape (Agilent 2200 TapeStation system) and quantified by qPCR (KAPA library quantification kit KK4828). Libraries were then sequenced by in-house on a NextSeq 500 (Illumina), 150bp paired-end sequencing or Novogene on a NovaSeq 6000, 150bp paired-end sequencing.

2.8.2 TT-seq

For referenced normalised TT-seq, thiouridine-labelled spike-in RNA was generated by *in vitro* transcription of exogenous plasmid sequences in the presence of 4S-UTP (Jena Bioscience) using the MEGAscript kit (ThermoFisher Scientific). Cells were seeded into 3x 10cm plates. When 70-80% confluent, 50mM 4-thiouridine (4SU) was added to the cells to a 500μM final concentration and plates were incubated at 37ºC for 5 minutes.

After 5 minutes, media was aspirated from the plates and 2ml of trizol (Thermofisher 15596026) was added. The cells were scraped from the plates and aspirated into a 15ml tube and snap frozen on dry ice, before storage in -80ºC.

For referenced normalised TT-seq, 1.2 ng of labelled RNA was added to the Trizol once thawed. To extract RNA, 1ml chloroform was added and the tubes shaken vigorously for 15 seconds followed by a 3-minute incubation at room temperature. Samples were centrifuged at 4500rpm for 15 minutes at 4ºC. The aqueous layer was transferred to a fresh 15ml tube and 5ml of isopropanol was added, mixed well and incubated for 10 minutes at room temperature. Samples were centrifuged at 4500rpm for 10 minutes at 4ºC to precipitate the RNA. The supernatant was removed, and the pellet washed with 5ml of 75% ethanol. Samples were centrifuged at 4500rpm for 10 minutes at 4ºC and the ethanol removed. The pellet was resuspended in 1ml of 100% ethanol and transferred to an Eppendorf tube. The cells were centrifuged at 13000g for 5 minutes at 4ºC. The supernatant was removed, and the pellet was air-dried then resuspended in 87μl of RNase-free water, followed by incubation at 65ºC for 10 minutes to dissolve the pellet. Samples were treated with DNase using the TURBO DNA-free kit (AM1907, Thermofisher) following the manufacturer's instructions. The RNA concentration was measured by Nanodrop and RNA stored at -80ºC.

Sonication

200µg of RNA was diluted with Nuclease-free water to 240µl and transferred to Covaris tubes and sonicated using the conditions in Table 2.12:

Table 2.12 Sonication conditions for TT-seq

Peak Power	Duty Factor	Cycles per Burst	Total Time (m.s)
70	20%	1000	0.10

Biotinylation of labelled RNA
100µg of RNA was biotinylated in the reaction in Table 2.13, twice for each sample:

Table 2.13 Biotinylation Reaction for TT-seq

Component	Volume (µL)
RNA (100µg)	120
HPDP-Biotin 1mg/ml Pierce, 50mg EZ-link Biotin-HPDP 21341	200
10X Biotinylation Buffer (100mM Tris pH 7.48 10mM EDTA)	100
Nuclease-Free Water	580

This mixture was rotated at room temperature for 90 minutes. Each reaction was then split to into two Phase-Lock Gel Heavy tubes (Quantabio 2302830). 500µl chloroform/isoamylalcohol (24:1) was added and mixed vigorously, followed by centrifugation at 20,000g for 5 minutes. This step was repeated with a further 400µl chloroform/isoamylalcohol. The aqueous fraction was transferred to Eppendorf tubes. 50µl 5M NaCl and 500µl isopropanol was added to each tube and samples were centrifuged at 20,000g for 20 minutes at 4^{o}C. The supernatant was removed, and the pellet was washed in 1ml 75% ethanol. The pellet was air-dried and then resuspended in 25µl 1X TE, and samples pooled to a final volume of 100µl.

Isolation of labelled RNA

Biotinylated RNA was denatured at 65ºC for 10 minutes then cooled on ice for 5 minutes. 100µl streptavidin beads were added (µMACS Streptavidin Kit, Miltenyi 130-074-101) and rotated for 15 minutes at RT. The µMACS columns (µMACS Streptavidin Kit, Miltenyi 130-074-101) were equilibrated on a magnetic stand using 900µl of Washing Buffer (100mM Tris pH 7.5; 10mM EDTA; 1M NaCl; 0.1% Tween20). The RNA:bead mix was added to the column and flow-through collected. The flow-through was re-applied to the column and then discarded. This was washed three times with 900µl of 65ºC Washing Buffer and three times with RT Washing Buffer. The RNA was eluted directly into 700µl Buffer RLT (RNeasy MinElute Cleanup Kit, Qiagen 74204) by applying 2X 100µl 100mM DTT to the column. RNA was purified using the MinElute PCR purification Kit (Qiagen 28004) as per manufacturer's instructions and eluted in 13µl RNase-free water. The RNA concentration was measured using Qubit RNA high sensitivity kit and RNA stored at –80ºC.

Next generation sequencing libraries of the labelled RNA were generated using the NEBNext Ultra II Directional RNA library preparation kit (E7760, NEB) and NEBNext Multiplex oligos for Illumina (E7335L/E7500L) (NEB E7530) following manufacturer's instructions. Libraries were then sequenced in-house on a NextSeq 500 (Illumina), 150bp paired-end sequencing or by Novogene on a NovaSeq 6000, 150bp paired-end sequencing.

2.8.3 Quality Control

Library quality was assessed using the DNA 1000 high sensitivity Screen Tape (Agilent 2200 TapeStation system) and quantified by qPCR (KAPA library quantification kit KK4828).

2.8.4 RNA-seq analysis

I used the SeqNado (https://pypi.org/project/seqnado) pipeline on the WIMM CCB server developed in the Milne lab for next generation sequencing analysis. SeqNado maps and post-processes mapped reads including deduplication and blacklist filtering. SeqNado generates bigwigs for visualisation and performs peak calling with Macs2 and Lance-O-Tron. RNA-seq data is quantified using either STAR counts or featureCounts and differential expression analysis using DESeq2 is performed for pair-wise comparisons within SeqNado. For TT-seq performed on dCBP-1 treated cells, spike-in exogenous RNA was used to normalise for total RNA levels. These are generated by transcription of plasmid sequences in the presence of 4S-UTP. Spike-in RNA levels were quantified by mapping to a custom genome using featureCounts and used to normalise the output of DESeq2.

2.9 ATAC-seq Methods

The ATAC-seq protocol is adapted from (Buenrostro *et al.*, 2015). 50,000 cells were harvested by centrifugation (300g, 5min) followed by PBS wash. Cells were resuspended in 50µl cold lysis buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1% IGEPAL CA-630) and spun at 500g for 10 minutes at 4^{o}C. The supernatant was discarded, and the pellet was resuspended in freshly made transposase buffer (Table 2.14) and incubated at 37^{o}C for 30 minutes.

Table 2.14 Transposase Buffer for ATAC-seq

Component	Volume (µl)
RNA (100µg)	120
2X TD Buffer Illumina Tagment DNA Buffer, 20034197	25
Tn5 Transposase (Illumina Tagment DNA Enzyme, 20034197)	2.5
Nuclease-Free Water	2.5

Following transposition, DNA was purified using the MinElute PCR purification Kit (Qiagen 28004) as per manufacturer's instructions and eluted in 10µl 10mM Tris pH8.

2.9.1. ATAC-seq library preparation and sequencing

The transposed DNA was amplified in a PCR reaction in PCR tubes in Table 2.15 and transferred to a thermocycler and run with the conditions in Table 2.16:

Table 2.15 ATAC-seq library preparation reaction mixture

Component	Volume (µL)
Transposed DNA	20
25µM Nextera Index Primer	2.5
25µM Nextera Universal Primer	2.5
2x NEBNext Ultra II Q5 MasterMix (NEB M0544S)	25

Table 2.16: ATAC-seq library preparation PCR amplification conditions

Temperature (°C)	Time (m.s)	Cycles
72	5.00	1
98	0.30	1
98	0.10	12
63	0.30	
72	1.00	

DNA was purified using the MinElute PCR purification Kit (Qiagen 28004), eluting in 20µl of elution buffer, into Lo-bind tubes (Eppendorf Z666548-250E) and stored at -20°C. Libraries were sequenced on NextSeq 500 (Illumina), 75bp paired-end sequencing or by Novogene on a NovaSeq 6000 ,150bp paired-end sequencing.

2.9.2 Quality Control

Library quality was assessed using the DNA 1000 high sensitivity Screen Tape (Agilent 2200 TapeStation system and quantified by qPCR (KAPA quantification kit KK4828).

2.9.3 ATAC-seq analysis

I used the SeqNado (https://pypi.org/project/seqnado) pipeline on the WIMM CCB server developed in the Milne lab for next generation sequencing analysis. SeqNado

maps and post-processes mapped reads including deduplication, blacklist filtering and shifting for ATAC-seq. SeqNado generates bigwigs for visualisation and performs peak calling with Macs2 and Lance-O-Tron.

Chapter 3: The ability of MYB to establish an enhancer-like element is context-dependent

3.1 Introduction

Dissecting the effects of a single transcription factor or co-activator is very challenging, given that transcription is a highly complex and interlinked process, relying on several multivalent interactions in a spatial and temporal manner (Näär, Lemon and Tjian, 2001; Bulynko and O'Malley, 2011). In this chapter, I set out to investigate the role of MYB, a transcription factor that is thought to function through recruitment of the lysine acetyltransferases CBP and P300 (Zor *et al.*, 2004; Pattabiraman *et al.*, 2009, 2014). I build upon previous findings from our group, where we demonstrated that MYB alone possesses the ability to establish an enhancer-like element in the absence of other DNA-bound regulatory elements. However, it was not clear from this result whether the specific genetic locus used may have contributed to the ability of MYB to activate it. By anchoring MYB to a distinct genomic locus, I show that the enhancer forming capabilities of MYB are context dependent. This sheds further light on the role of MYB, and indirectly the contributions of CBP and P300, in enhancer function when isolated from the influence of other transcription factors and active domains.

3.1.1. MYB structure and function

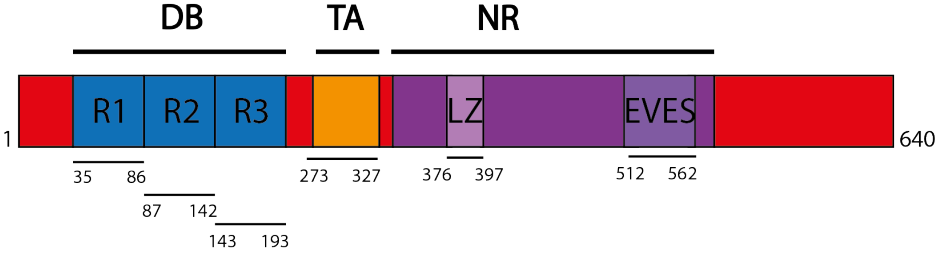


Figure 3.1 Schematic of MYB protein containing conserved domains

N-terminal DNA binding domain (DB) in blue contains tandem repeats R1, R2 and R3, transactivation (TA) domain in orange, negative regulatory (NR) domain including the leucine zipper (LZ) and EVES motif in purple. Amino acid residues are noted for the human MYB. Figure adapted from Dr I-Jun Lau's thesis (Lau, I. 2019).

MYB is a transcription factor that is known to interact with P300 and CBP, lysine acetyltransferases associated with gene activation. MYB consists of three highly conserved domains, the N-terminal DNA binding (DB) domain, the TA and the C-terminal negative regulatory (NR) domain (Figure 3.1). The TA domain is notably responsible for engaging with co-activator proteins including P300 and CBP through their KIX domain (Dai *et al.*, 1996; Sandberg *et al.*, 2005; Pattabiraman *et al.*, 2009). Point mutations within the TA domain of MYB have been observed to disrupt its interaction with P300 and CBP, leading to a loss of MYB's transforming ability.

Furthermore, it has been established that the TA domain of MYB directly interacts with TAF12, linking it to the transcription machinery (Xu *et al.*, 2018). These interactions suggest that MYB may play a role in transcription activation by stabilising these complexes on chromatin at both promoters and enhancers.

3.1.2 The TetR System for investigating components of transcription in isolation

Several challenges arise when aiming to interrogate the role of a single transcription factor. To completely understand its function, an appropriate cellular context needs to be provided. However, this introduces further complexity as endogenous regulatory elements tend to include binding sites for multiple TFs, making it difficult dissect the specific role of a single TF. Conversely, attempting to isolate these TFs often involves

the use of *in vitro* experiments, where certain molecular components are introduced systematically, in different combinations. Unfortunately, such approaches require pre-selection of factors for inclusion and therefore fail to reveal unpredicted protein:protein interactions, and the *in vivo* relevance of interactions detected *in vitro* remains unclear. To address these caveats, there is a need for a technique that enables the functional isolation of the TF in question, whilst preserving a physiologically relevant environment. To achieve this, we employed the TetR system. This novel approach allowed us to examine the function of MYB, isolating the MYB TA domain from the complexity of active transcription, other transcription factors and endogenous regulatory elements, whilst maintaining a physiologically relevant, chromatin context.

The Tet Repressor (TetR) system, originally developed by the Klose lab, has proven instrumental in previous studies examining protein:protein interactions within a chromatin context, such as those involving PcG (polycomb group proteins) and MLL-AF4 (Blackledge *et al.*, 2014; Kerry *et al.*, 2017). This system exploits the high affinity interaction between the bacterial Tet repressor protein (TetR) and the Tet operator *(TetO)* DNA sequence to target protein domains to a specific genomic locus and investigate the consequences on the local chromatin environment.

In this system, mouse embryonic stem cells (mESC) are used that have been genetically engineered to incorporate a series of 14 bacterial *TetO* DNA binding sequences, each separated by approximately 20bp of intervening sequence, collectively forming what is known as the *TetO* array. This array is integrated into a ~180kb neutral region taken from human chromosome 7, inserted into mouse chromosome 8 of ESCs (Figure 3.2). Importantly, this neutral region lacks identifiable genes, CpG islands, the presence of common histone modifications (H3K4me3, H3K27me3, H3K9ac, H2AK119ub, H3K36me3, H4K20me2 and H3K9me3) as well as RNAPII occupancy (Blackledge *et al.*, 2014).

Given the strong binding affinity of the bacterial TetR to the *TetO* array and the use of mouse ES cells, the interaction is highly specific. A fusion protein is engineered, consisting of the TetR DNA binding domain fused with the protein or domain of interest. This fusion protein is stably expressed, effectively anchoring the protein or domain of interest to the *TetO* array. This system has proven valuable for investigating the transcriptional potential associated with the protein of interest by combining multiple next generation sequencing methods including ChIP-seq, ATAC-seq and RNA-seq and observing changes around the *TetO* locus driven by the protein binding there.

A key advantage of this system is that upon the addition of doxycycline, the TetR fusion is removed from the *TetO* array. This allowed us to assess the dynamics of the interactions we reveal, by evaluating how long they persist at the *TetO* array after the removal of the TetR fusion. Additionally, it serves as another important negative control, validating that the presence of histone modifications and TFs are attributable to the presence of the TetR fusion.

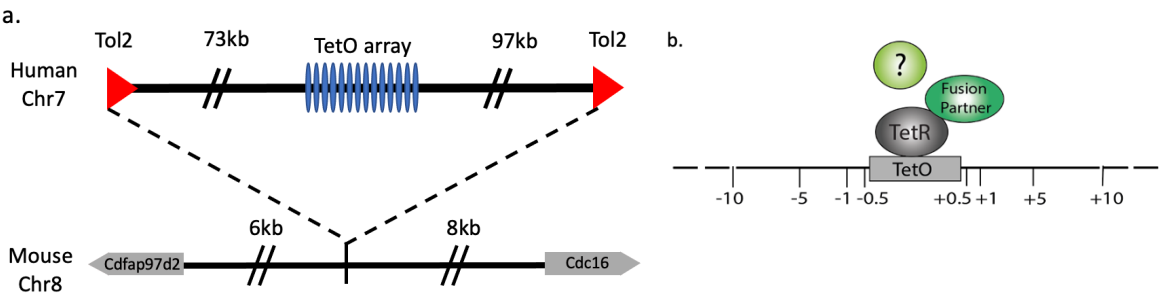


Figure 3.2 The TetR System

a | The *TetO* array, 14 repeats of the *TetO* sequence, each separated by ~20bp, inserted into a 180Kb gene desert in human chromosome 7. This is further integrated into mouse chromosome 8 via bacterial artificial chromosome (BAC) insertion to create the E14 mESC TOT2N variant cell line.

b | The TetR DNA binding domain binds to the *TetO* array, bringing in its fusion partner. Numbers represent the positions of qPCR primers in kb, relative to the *TetO* array. Figure adapted from (Blackledge et al., 2014).

3.1.3 The MYB transactivation domain is sufficient to generate an enhancer-like element.

Previous work from our group (Dr I-Jun Lau, unpublished data) employed the TetR system and engineered a TetR fusion with the TA domain of MYB, resulting in TetR-MYB^{TA}, which binds strongly to the *TetO* array. We conducted experiments with the TetR-MYB^{TA} in comparison with two controls: mESCs stably expressing TetR alone (TetR) or no TetR protein (N.P.). This allowed us to draw conclusions about whether any transcriptional effects we observe were attributed to MYB^{TA} and are not a general consequence of TetR binding at the *TetO* array.

A striking result observed was the ability of MYB^{TA} to induce several key attributes of an active enhancer at the *TetO* locus. As highlighted in section 1.2, properties of an active enhancer include TF binding, presence of active histone modifications (H3K27ac, H3K4me1), chromatin accessibility, eRNA transcription and interactions with target promoters.

ATAC-seq analysis provided evidence that the binding of TetR and TetR-MYB^{TA} at *TetO* results in an increase in chromatin accessibility at the *TetO* locus*,* consistent with the idea that protein binding to DNA promotes chromatin accessibility. Notably, in the cells with no TetR expressed (N.P.), the chromatin at the *TetO* array remained in a closed state. Furthermore, distinct, ATAC-seq peaks, distal to the *TetO* array, were identified across all conditions, with a notable increase in accessibility observed in cells expressing TetR-MYB^{TA}. These peaks are represented as *1-4 in Figure 3.3.

Interestingly, these ATAC-seq peaks are also marked with H3K27ac. Strikingly, polyA-transcription, which captures ncRNA including eRNAs, was detected at the *TetO* array and is potentially initiating from the ATAC-seq peaks *1-4, indicating that these peaks may represent transcription start sites. The distal elements that were activated by MYB^{TA} are already partially accessible in the N.P./TetR cell lines, suggesting that there is some cryptic TF binding already occurring there that get further activated when MYB^{TA} binds to the *TetO*. This suggests that enhancer activity by MYB^{TA} may be context-dependent, relying on the presence of specific TF binding motifs defining cryptic elements.

These findings collectively suggest that MYB^{TA} possesses the ability to generate an enhancer-like element at the *TetO*, which is capable of activating previously latent cryptic regulatory sites. The observation that these sites are distal (up to ~50Kb away) to the TetR-MYB^{TA} binding site, indicates that MYB^{TA} may act via 3D genomic interactions with the distal sites. However, several questions remain: is this effect reliant on the recruitment of P300/CBP by MYB (investigated in Chapter 4), and is the effect dependent on the sequence context of the MYB^{TA} binding site?

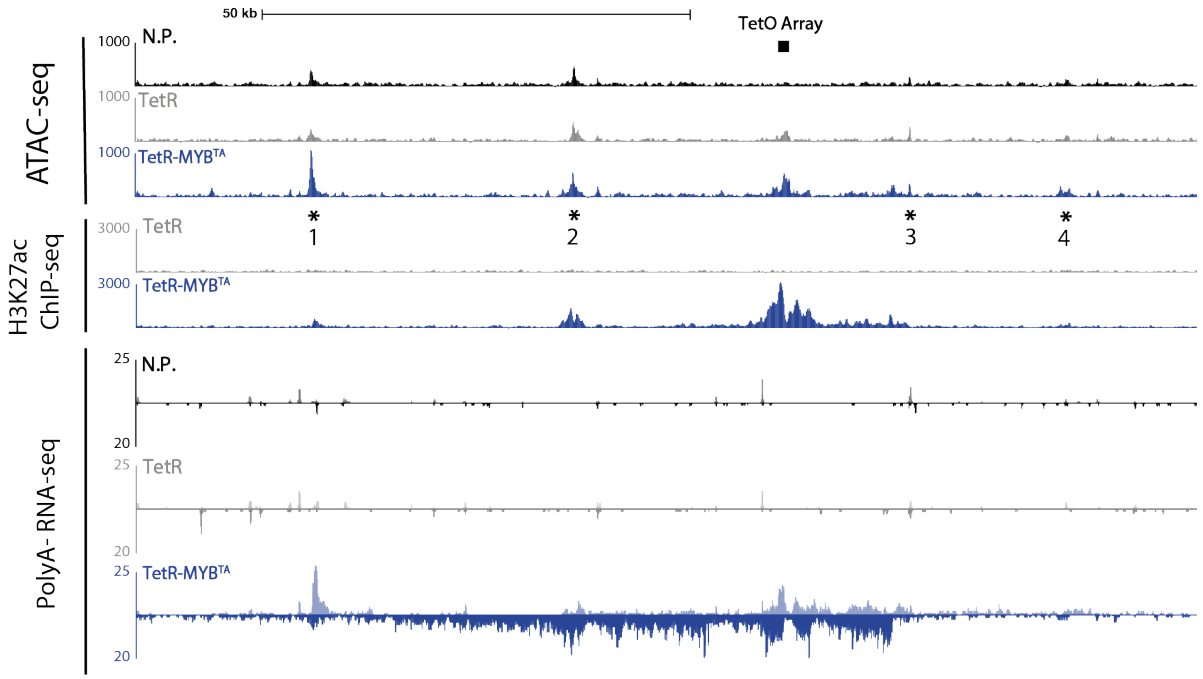


Figure 3.3 Binding of TetR-MYB^{TA} to the *TetO* array is able to increase chromatin accessibility (ATAC-seq), H3K27ac (ChIP-seq) and eRNA transcription (Poly(A)- RNA-seq) both locally and at distal regions.

Region shown is ~100kb surrounding the *TetO* array on Chromosome 8. Representative tracks for ATAC-seq (n=3), H3K27ac ChIP-seq (n=1) and Poly(A)- RNA-seq (n=3) are shown. Activated distal regions are marked with * and numbered 1-4. Data generated from Dr I-Jun Lau, Milne Group.

3.1.4 The need for an additional locus for investigating context-dependence.

In this chapter, I wanted to ask if the enhancer forming capabilities displayed by MYB^{TA} at the *TetO* locus inserted into chromosome 8 is a general property of MYB or if MYB is only able to establish enhancer activity in a certain genomic contexts. To explore this further, I planned to introduce a second *TetO* array elsewhere in the genome and investigated the enhancer activity of TetR-MYB^{TA} there. Previously, a neutral region of chromatin on chrX of mESCs (chrX:10,935,857-11,046,248 (mm9)) was characterised

by Dr Caz Harrold (Hughes, WIMM) and confirmed to be devoid of genes and functional elements (Appendix).

To confirm that this region is not transcriptionally inert and retains the potential to be activated, I reference findings from Dr Caz Harrold's thesis (Harrold, C. 2019). The mouse alpha globin locus is a well-studied gene regulated by a super enhancer of five enhancer elements (Hay *et al.*, 2016). In her work, she aimed to dissect the role of one single enhancer element of the alpha globin cluster to study its effect alone on chromatin and transcription. She isolated the R2 enhancer element, the strongest enhancer from the erythroid-specific super enhancer cluster and inserted it into this neutral region on chrX in mESCs. This allows us to investigate the effect of a single, well described enhancer element away from its native genomic context where the other alpha globin enhancers and erythroid TFs are likely contributing to alpha globin expression.

In mESCs, there is no activation of this region, as the R2 enhancer is not active in mESCs (Georgiades *et al.*, 2023). After differentiating the cells to embryoid bodies, a cell line in which the R2 enhancer is active, she showed that, within the neutral region, the R2 enhancer displayed characteristics comparable with an active promoter (Figure 3.4). Specifically, by ATAC-seq she observed an increase in chromatin accessibility at the R2 insertion site and distal elements. Moreover, there was an increase in H3K27ac and H3K4me1 levels at both the R2 insertion site and the distal elements, signifying a hallmark feature of enhancers. Intriguingly, a substantial rise in H3K4me3 at the R2 insertion site was noted, suggesting this site could be a cryptic promoter, which has now been activated. Excitingly, she even detected both eRNA and mRNA transcription by PolyA- and PolyA+ RNA-seq respectively, initiating from the R2 insertion site (Georgiades *et al.*, 2023).

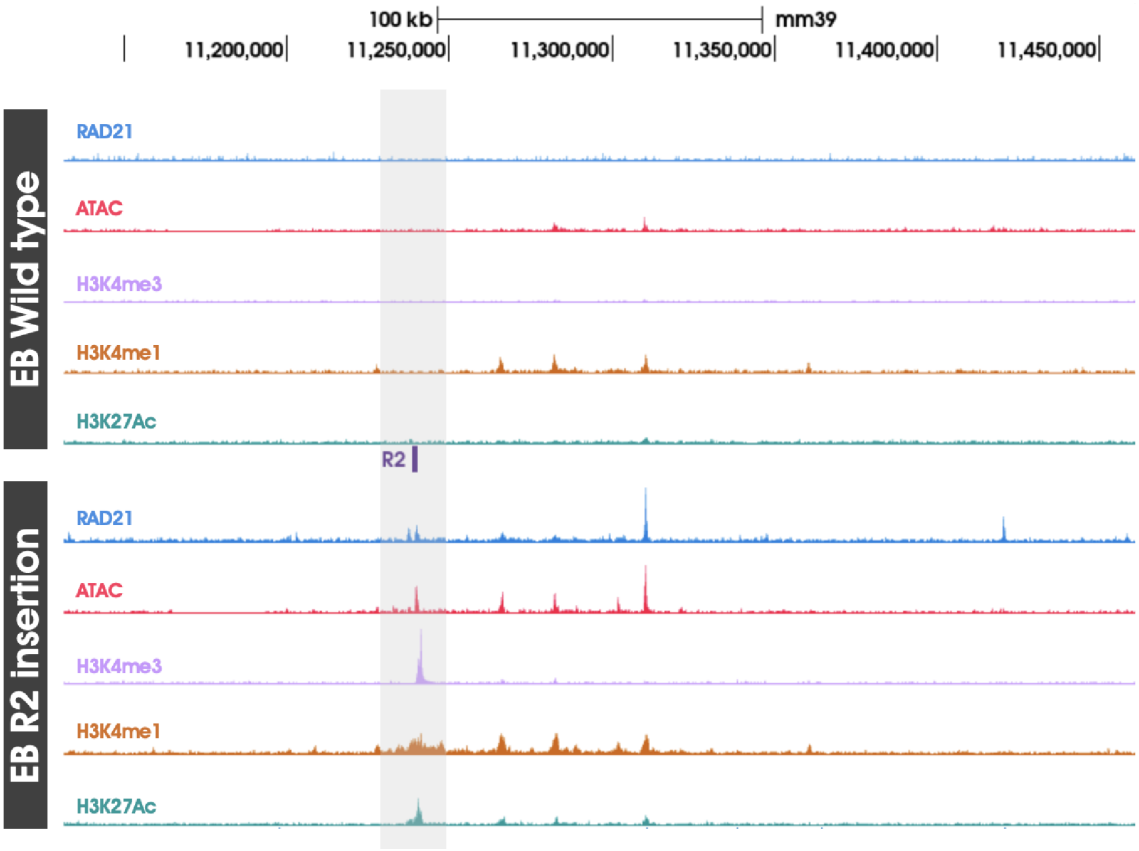


Figure 3.4 Chromatin marks in embryoid bodies after the insertion of the R2 element in a neutral region on chromosome X

Figure taken from Dr Emily Georgiades' thesis (Georgiades, E. 2022) and data generated from Caz Harrold's thesis (Harrold, C. 2019) where further details are available.

Representative tracks for ATAC-seq and ChIP-seq for H3K27ac, H3K4me1, H3K4me3 and cohesin from two/three biological replicates for WT (ChIP-seq/ATAC-seq data respectively) and three independently targeted clones for embryoid bodies (EB) R2 insertion. The R2 insertion site is denoted in purple.

Based on the work in Dr Harrold's thesis, we concluded that this locus on chromosome X is an appropriate system to further investigate the function of MYB, in a novel neutral region. By using a site on chromosome X in male mESCs, we avoided the complexity of having two copies in the genome. Utilising this system, in this chapter, I engineered

an additional *TetO* array in the same insertion site where the alpha globin enhancer was inserted, now referred to as 'chrX-*TetO*'.

The chrX-*TetO* model serves as an exemplary platform for investigating the role of MYB and, indirectly, the role of CBP and P300 within an alternative genomic context. Given that MYB is capable of activating distal regions at the chromosome 8 locus, and that the alpha globin R2 enhancer can activate distal regions at the chromosome X locus, I wanted to ask whether MYB targeted to chromosome X would be able to activate either the same elements as R2, or other distinct elements. To address this, I engineered the new chrX-*TetO* array in the same mESCs in which the chr8-*TetO* array already exists. Therefore, I could directly compare the transcriptional effects of TetR-MYB^{TA} at the two loci. I conducted chromatin profiling at the chrX-*TetO* locus to assess the impact of targeting these proteins within distinct sequence environments.

3.1.5 Aims

I set out to investigate whether the enhancer-forming property of MYB and subsequently CBP/P300, is a unique feature of the chromatin context. To do this I had three main aims:

1. Engineer a novel *TetO* locus in a distinct neutral region of chromatin.
2. Express TetR-MYB^{TA} and validate its binding and functionality at the new *TetO* locus.
3. Investigate if MYB^{TA} is sufficient to establish an enhancer-like elements at the new *TetO* array.

3.2 Results

3.2.1 Characterisation of an in vivo targeting strategy

Prior to engineering a new *TetO* array in the chrX neutral region, I set out to fully decipher the sequence and orientation of the surrounding human and bacterial sequence of the *TetO* array situated on chromosome 8. The primary objective was to generate a custom genome to allow next generation sequencing mapping. The *TetO* array had initially been incorporated into a bacterial artificial chromosome (BAC) containing around ~180kb of human chromatin before insertion into mouse chr8, which incorporated a NeoR marker for positive clone selection. While the Klose lab had successfully identified the insertion site of the BAC using splinkerette PCR, certain aspects remained unclear; the orientation in which the BAC was inserted as well as the orientation of the human region and the NeoR gene relative to this insertion.

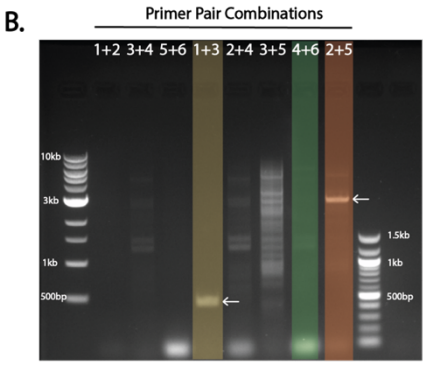
To address these issues, I designed primer sets flanking each of the junctions where the orientation remained undetermined (Figure 3.5A) and performed PCR amplification and gel electrophoresis to identify which pairs produced a product, from which I could determine the correct orientation (Figure 3.5B). Detection of a PCR product on an agarose gel indicated the tested junction exists. PCRs using primer pairs: mouse_A (1)/human_TetO_B (3) and human_TetO_A (2)/NeoR_B (5) were successful, allowing us to infer the correct orientation of BAC insertion.

Sanger sequencing was used to confirm the exact junctions and so produce a comprehensive sequence of the inserted region (Figure 3.5C). This was used to create a custom genome for the chr8-*TetO* mESCs, facilitating precise mapping of next generation sequencing data. The product between NeoR_A (4) and mouse_B (6), highlighted in green, was not detected on the agarose gel (Figure 3.5B). This is likely

because the amplicon is too large to be amplified by PCR, due to previously unidentified BAC sequence in blue (Figure 3.5C). Therefore, I conducted additional primer design and subsequent PCR experiments to sequence this region.

A.

Primer Name	Direction	Amplicon Size (bp)
mouse_A (1)	Forward	482
human_TetO_B (3)	Reverse	
human_TetO_A (2)	Forward	2871
NeoR_B (5)	Reverse	
NeoR_A (4)	Forward	6027
mouse_B (6)	Reverse	



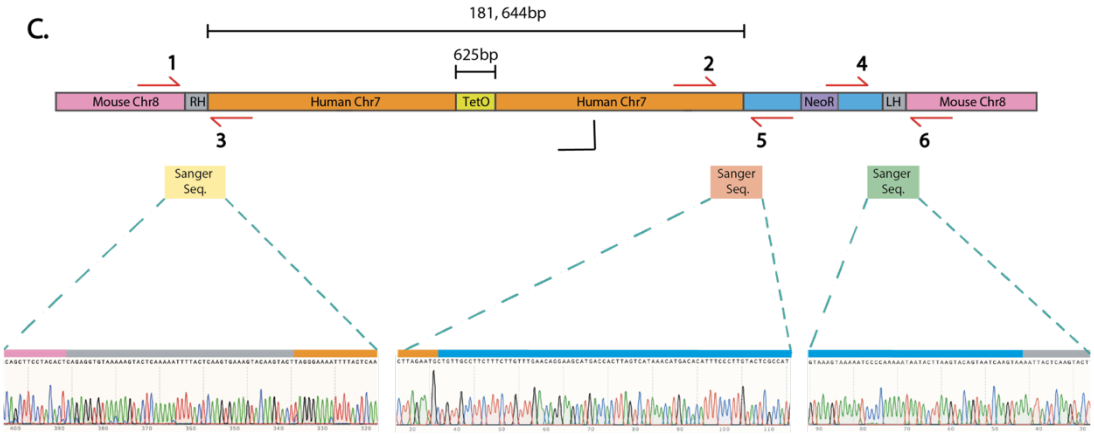


Figure 3.5 The orientation and sequence of edited mESCs harbouring *TetO* array, confirmed by PCR screening and validated by Sanger Sequencing

gDNA from E14 (TOT2N variant) mESCs was used as template for PCR.

A| Table with the primer pairs used for PCR amplification to identify the orientation of BAC insert.

B | Agarose gel image, each lane is the product of a PCR using the indicated primer pair combination. Highlighted lanes show the three primer pairs from the correct orientation. Arrows show the product generated.

C | Schematic showing the correct orientation of BAC insertion, determined by the PCRs. Using all primers in different combinations allows identification of the correct orientation based on which primer pairs give a PCR product on the agarose gel. Sanger sequencing traces for each

junction. Additional primer pairs were designed and used to sequence the intervening BAC sequence in blue which was previously unidentified.

3.2.2 Using CRISPR-Cas9 HDR to engineer an additional *TetO* array on chromosome X

The *TetO*-ChrX donor vector was designed in collaboration with Philip Hublitz of Genome Engineering and synthesised by GeneArt (ThermoFisher Scientific). The array includes two mutations (highlighted in red) compared to the *TetO* array on chromosome 8 (i) a TA to AT mutation at the *TetO* primer binding sites to distinguish the two loci in ChIP-qPCR experiments and (ii) an A deletion to remove a DpnII site, ensuring that the *TetO* array is contained within a single *TetO* fragment to allow for Capture C design (Figure 3.6).

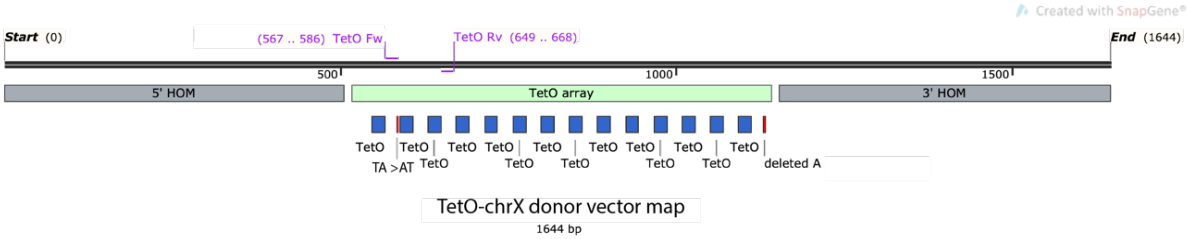


Figure 3.6 *TetO*-ChrX donor vector map

Total vector size is 1644bp containing the *TetO* array ~625bp, flanked by ~500bp homology arms targeting either side of the insertion site on chromosome X.
The vector also includes restriction enzyme sites for easy exchange of homology arms. Plasmid backbone including kanamycin resistance marker not shown. Figure created using SnapGene.

For the integration of the new *TetO* array into the chrX neutral region, we used Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 homology dependent repair (HDR). Chr8-*TetO* cells were transfected with the chrX-*TetO* donor

vector alongside the gRNA that had demonstrated the highest efficiency for targeting this specific genomic region, as previously identified in Dr Caz Harrold's thesis (Harrold, C. 2019). Cells were transfected using either lipofectamine LTX or lipofectamine 2000. Untransfected cells of the same origin were used to set a gate for GFP+ cells. Using lipofectamine LTX saw a GFP+ population of 12.1% and lipofectamine 2000 a population of 16.7% (Figure 3.7).

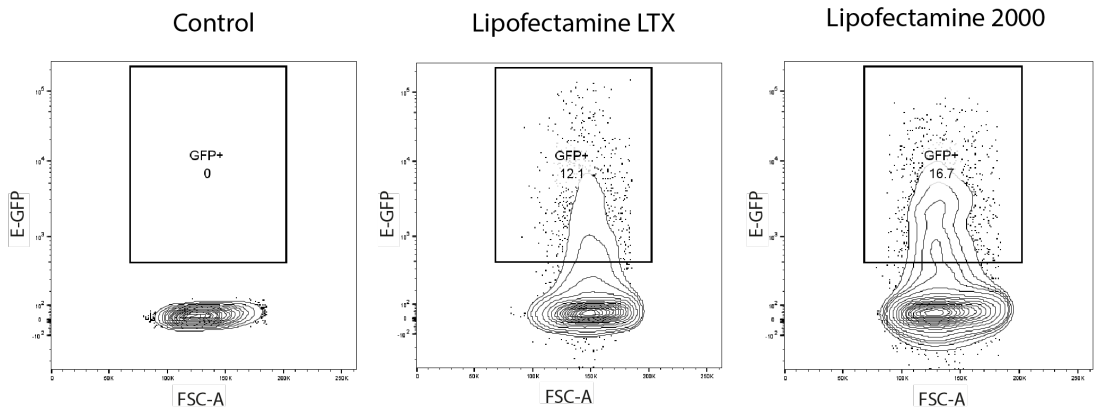


Figure 3.7 Gating strategy for fluorescence activated cell sorting (FACS) after transfection with *TetO*-ChrX donor plasmid and guide plasmid containing GFP

Control cells are wild type E14 (TOT2N variant) which were not transfected, used to gate for GFP+ cells (GFP+ = 0.0%).

Cells transfected with the donor and the guide plasmid using lipofectamine LTX displayed 12.1% GFP+ transfected cells.

Cells transfected with the donor and the guide plasmid using lipofectamine 2000 had a population of 16.7% GFP+ cells.

Analysis conducted on FlowJo.

Internal primers (which only amplify positive insertions) were used as an initial screen for successful editing in a pool of clones (Figure 3.8A). Clones within each positive pool were further screened using spanning primers to confirm insertion of the correct size

(Figure 3.8B). As this editing was on chromosome X of male mESCs there was no need to isolate homozygous edited clones. Sanger sequencing confirmed successful editing and ensured there were no unwanted mutations (Figure 3.8C). Clone B1 was selected for use in all subsequent experiments.

A.

Type	Primer	Name
Spanning	Forward	ChrX_TetO_LH_Fw
	Reverse	ChrX_TetO_RH_Rv
Internal (LH)	Forward	ChrX_TetO_LH_Fw
	Reverse	ChrX_TetO_LH_Rv
Internal (RH)	Forward	ChrX_TetO_RH_Fw
	Reverse	ChrX_TetO_RH_Rv

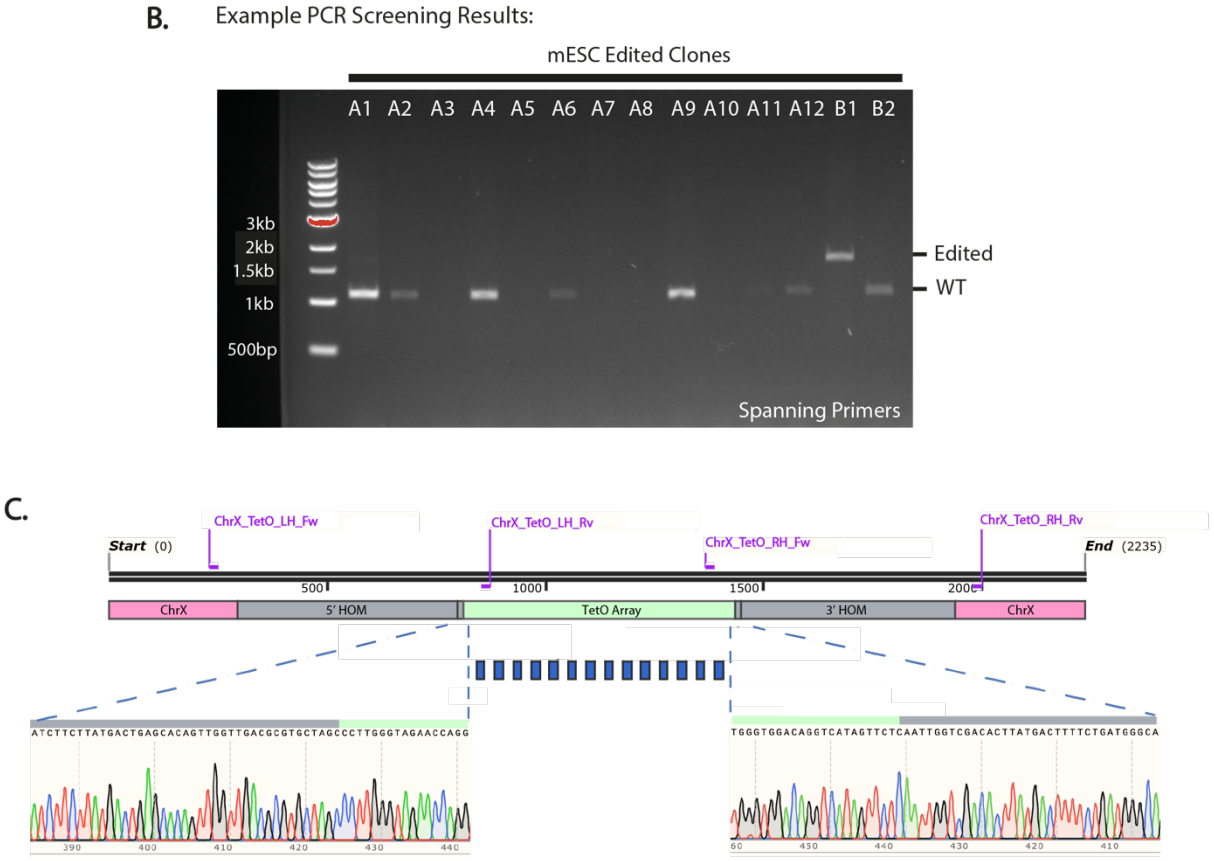


Figure 3.8 Identification of positive clone for *TetO* array in neutral region of ChrX

Clones were screened using the above screening strategy, details are in Materials and Methods.

A | Table with the primer pairs used for PCR amplification to identify positive clones. Internal primers were used for initial identification of successful editing in a pooled screen. Then gDNA

from individual clones was used as a template using spanning primers to further identify positive insertion clones.

B | Agarose gel using spanning primers, B1 is a positive, homozygous clone. 1kb ladder is on the left-hand side.

C | The edited band for clone B1 from the agarose gel was extracted, purified and sequenced by Sanger sequencing to confirm the clone was not part of a mixed population and that there were no unwanted mutations at the insertion site.

Off-target effects remain a disadvantage of CRISPR-Cas9 HDR. This occurs when dCas9 recognises a genomic site which differs from the sequence the guide RNA recognises and results in cleavages and the potential integration of the donor sequence into unintended locations. In such instances, the off-target site is distinct from the intended chrX-*TetO* insertion site targeted by the guide RNAs. Consequently, there is a risk of *TetO* array insertion elsewhere in the genome. Although this possibility was not fully explored due to focusing on the chrX-*TetO* and its adjacent ~200kb chromatin, any additional alterations outside of this region were deemed unlikely to impact my conclusions. However, if additional *TetO* arrays were inserted elsewhere in the genome, they could offer supplementary sites for investigation if inserted correctly. To identify if this may have occurred, digital droplet PCR could be employed to identify the copy number of the *TetO* sequence within the genome and splinkerette PCR to identify the precise genomic location. This would in theory expand the number of sites available for investigation beyond the two explored in this chapter (chr8-*TetO* and chrX-*TetO*).

3.2.3 Validation of TetR-MYB^{TA} System at chrX-*TetO*

Human and mouse MYB and its interacting partners CBP and P300 are very similar with >90% amino acid identity (Table 3.1). Therefore, findings from this chapter are thought to be applicable to both organisms.

Protein	% amino acid identity
P300	93.7
CBP	95.3
MYB	90.5

Table 3.1 Conservation of MYB and interacting proteins.

Percentage amino acid identity between human and mouse. Analysis conducted using UniProt Consortium 2023.

After successfully cloning a mESC line harbouring chr8-*TetO* and chrX-*TetO* (referred to as N.P. = No Protein), the cells were used as a base in which to express TetR fusion proteins. Two lines were engineered, 'TetR' which stably expresses TetR and the 'TetR-MYB^{TA}', which stably expresses TetR-MYB^{TA} (i.e. the TA domain of MYB). The transfected plasmid was randomly integrated into the genome and puromycin was used to select for positive clones. A schematic showing the TetR-MYB^{TA} binding at the *TetO* array is displayed in Figure 3.9A. Clones were screened by western blot using the FS2 (FLAG-StrepII) antibody which recognises the N-terminal FLAG-StrepII tag of the TetR portion of the fusion (Figure 3.9B).

To confirm that the TetR-MYB^{TA} fusion was binding as expected at both the chr8-*TetO* and the new chrX-*TetO* I performed ChIP-qPCR using the FS2 antibody. Strong signal was detectable at the *TetO* array (0Kb) and reduced as I use primers moving further from the *TetO* array (0.5 and 1 kb away). This shows that the binding of TetR-MYB^{TA} is restricted to the *TetO* array, at least in this 2kb region. FS2 ChIP-qPCR in the TetR cell line was used as a control, where TetR also binds specifically to the *TetO* array in both

chr8-*TetO* and chrX-*TetO*. Similar levels of binding were observed for the TetR and TetR-MYB^{TA}, indicating that the presence of the MYB^{TA} domain did not interfere with DNA binding.

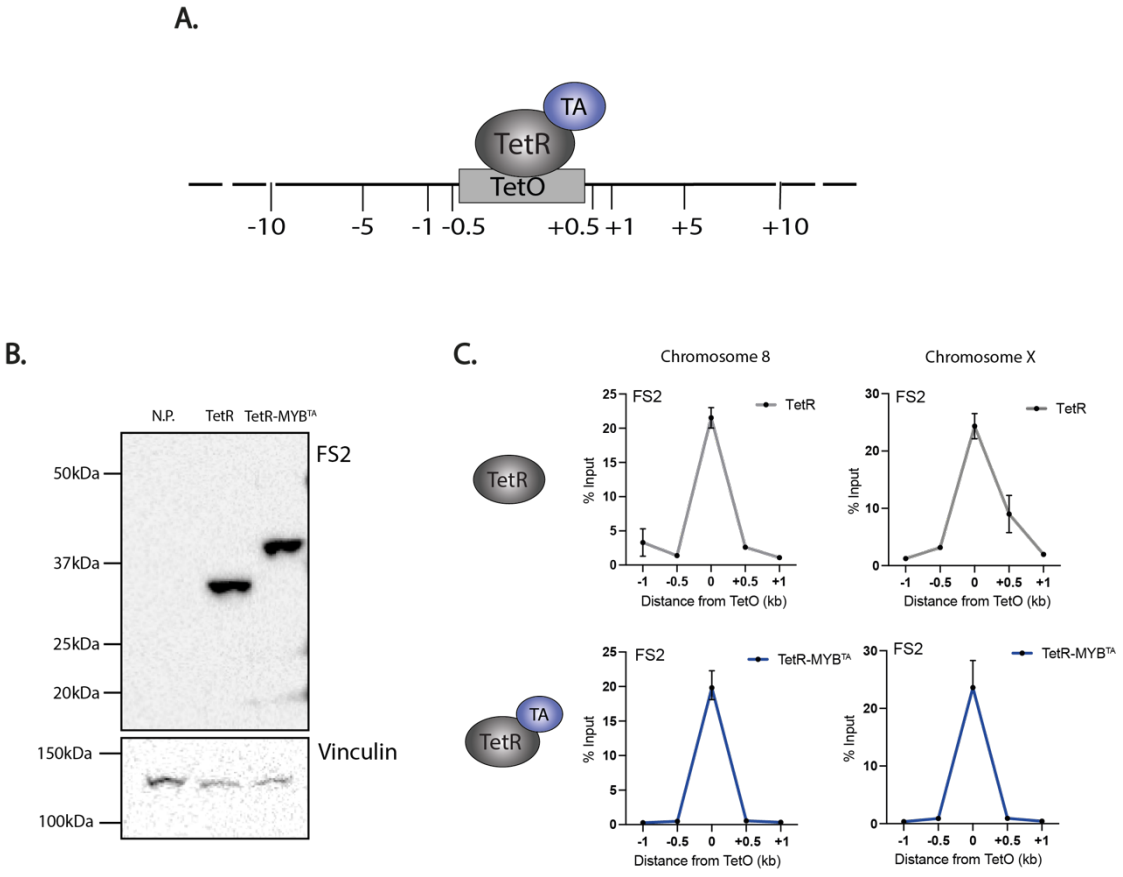


Figure 3.9 ChIP-qPCR and Western Blot show successful expression of TetR and TetR-MYB^{TA} in independent clones

A | Schematic of TetR-MYB^{TA} binding to the *TetO* array. Numbers represent approximate positions of qPCR primers in kb from the *TetO* array.

B | Western blot using protein extracted from N.P = (No TetR Protein), TetR stably expressed and TetR-MYB^{TA} stably expressed in E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO.* Top panel: FS2 (TetR) antibody which recognises TetR portion of the TetR fusion protein. Bottom panel: Vinculin antibody as a loading control. Precision plus ladder was used as a molecular weight marker.

C | ChIP-qPCR using the FS2 (TetR) antibody to detect chromatin binding of TetR. TetR: Data comes from two clones stably expressing TetR and error bars indicate the range (two datapoints). TetR-MYB^{TA} = data comes from three clones stably expressing TetR-MYB^{TA} and error bars indicate the range. % Input was calculated using the DDCt (ChIP versus input controls) method. X axis is comprised of primer pairs at the *TetO* array and +/- 0.5 and +/- 1 kb +/- 5kb (expect chrX +5kb) from the *TetO* array.

Next, I tested whether TetR-MYB^{TA} was able to recruit its functional partner, P300, at both chr8-*TetO* and the new chrX-*TetO.* Indeed, ChIP-qPCR for P300 in the TetR-MYB^{TA} cell line revealed the presence of P300 localised to *TetO* at both chr8-*TetO* and chrX-*TetO* (Figure 3.10). Furthermore, I detected H3K27ac by ChIP-qPCR, at both chr8-*TetO* and chrX-*TetO*, a mark that notably spreads over several kb. This indicated that TetR-MYB^{TA} is bringing in functional P300 at both loci, which is able to acetylate chromatin.

Notably there is a dip in H3K27ac at the *TetO* array itself. This is an observation previously noted with H3K27ac (Lau, I. 2019) but also H3K27me3 and H2AK199Ub (Blackledge *et al.*, 2014). This may be due the high density of TetR binding at the *TetO* array, resulting in nucleosome depletion, hence there is no substrate to be acetylated here.

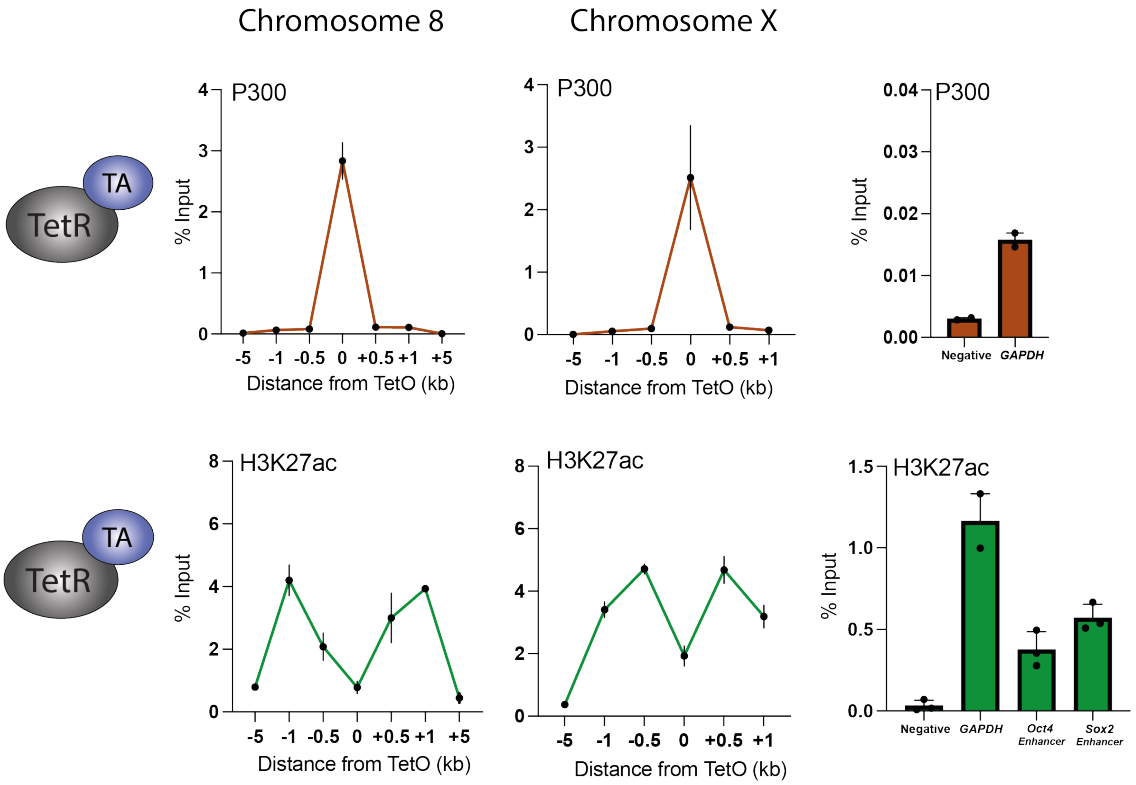


Figure 3.10 ChIP-qPCR using P300 and H3K27ac antibody at Chromosome 8 and Chromosome X

P300 ChIP (n=2) in the E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO* stably expressing TetR-MYB^{TA}, error bars indicate the range (two datapoints).

H3K27ac ChIP (n=2) in the E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO*

stably expressing TetR-MYB^{TA}, error bars indicate the range (two datapoints).

% Input was calculated using the DDCt (ChIP versus input controls) method. x axis is comprised of primer pairs at the *TetO* array and +/- 0.5 and +/- 1 kb +/- 5kb (expect chrX +5kb) from the *TetO* array.

3.2.4 MYB is still able to generate an enhancer at chr8-*TetO* in the new cell line.

To ensure that MYB^{TA} behaves the same in my newly engineered cell line as in the original chr8-*TetO* cells, I looked at chr8-*TetO* to see if MYB^{TA} binding still produced a novel enhancer element here. Consistent with previous results, TetR-MYB^{TA} increased chromatin accessibility, H3K27ac and transcription at the *TetO* locus and four distinct distal peaks, which accurately align to the same regions that Dr I-Jun Lau investigated (Figure 3.3). A novel finding was that the H3K27ac peak, but not chromatin accessibility, was diminished after a 3-hour doxycycline treatment, which disrupts the binding of TetR-MYB^{TA} (Figure 3.11). These findings strongly indicate that TetR-MYB^{TA} binding enhances accessibility at the *TetO* sites, a property that persists even upon the removal of TetR-MYB^{TA} through doxycycline treatment. In contrast, the presence of H3K27ac is abolished in the presence of doxycycline, highlighting the rapid removal rate of H3K27ac, which may be explained by a high turnover of this modification. I carried out TT-seq to investigate nascent transcription in addition to eRNA transcription detected by Poly(A)- RNA-seq (Lau, I. 2019). Using this orthologous technique, transcription was detected at the *TetO* array as well as initiating from the ATAC-seq peaks *1-4.

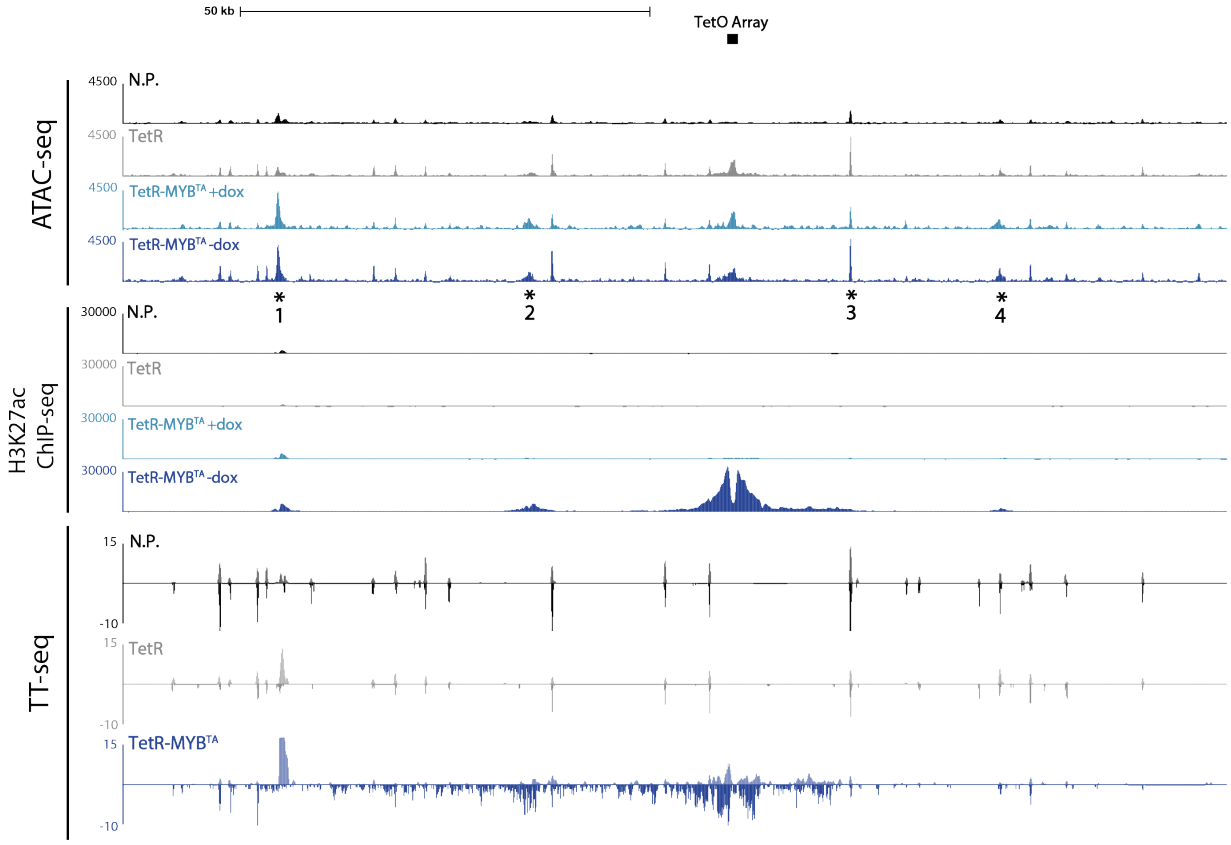


Figure 3.11 Binding of TetR-MYB^{TA} to the chr8-*TetO* array is able to increase chromatin accessibility (ATAC-seq), H3K27ac (ChIP-seq) and nascent transcription (TT-seq) both locally and at distal regions.

Representative tracks from UCSC of ATAC-seq (average of n=2), ChIP-seq (n=1) and TT-seq (n=1). Region shown is ~150kb surrounding the chr8-*TetO*, 50kb scale and TetO array shown. Cells are the E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO*. N.P. = No TetR Protein, TetR protein stably expressed, TetR-MYB^{TA} stably expressed. +dox = doxycycline treatment for 3 hours. Activated distal regions are marked with * and numbered 1-4.

3.2.5 MYB's enhancer forming capabilities are unable to be recapitulated at chrX-*TetO.*

To assess the enhancer-forming potential of MYB^{TA} at the newly generated novel chrX-*TetO* locus, an analysis of chromatin state in this region was carried out.
ATAC-seq analysis revealed that, similar to the observations at the chr8-*TetO*, the binding of TetR and TetR-MYB^{TA} at *TetO* resulted in a marked increase in chromatin accessibility at the *TetO* site. In N.P. cells, where TetR is not expressed, the chromatin remained in a closed configuration at the *TetO* array, consistent with previous observations (Figure 3.12). H3K27ac ChIP-seq profiling uncovered substantial spreading of H3K27ac at the *TetO* array, specifically to cells expressing TetR-MYB^{TA}. This acetylation was absent in both N.P. and TetR conditions, and, notably, dissipated after a 3-hour doxycycline treatment in TetR-MYB^{TA} cells, associated with a loss of TetR-MYB^{TA}. As observed at chr8-*TetO*, accessibility at the chrX-*TetO* array itself was maintained under these conditions.

In contrast to the chr8-*TetO* locus, the addition of TetR-MYB^{TA} was not associated with the appearance of any distal ATAC-seq or H3K27ac peaks at the chrX-*TetO* locus. Consistent with these findings, nascent transcription at the chrX-*TetO* locus was only detected in the TetR-MYB^{TA} cell line, apparently initiating at the *TetO* array, without any marked extension to distal regions within the approximately 100kb region under investigation. These collective observations indicate that MYB^{TA} binding at the chrX-*TetO* can generate transcription from the site at which it binds, similar to a promoter element. However, it is not sufficient to activate transcription at distal elements, in contrast to its binding at chr8-*TetO*. This suggest that the ability of MYB^{TA} to generate an enhancer element is contingent upon specific contextual properties, which may be linked to the underlying DNA sequence or the presence of other transcription factors at the locus.

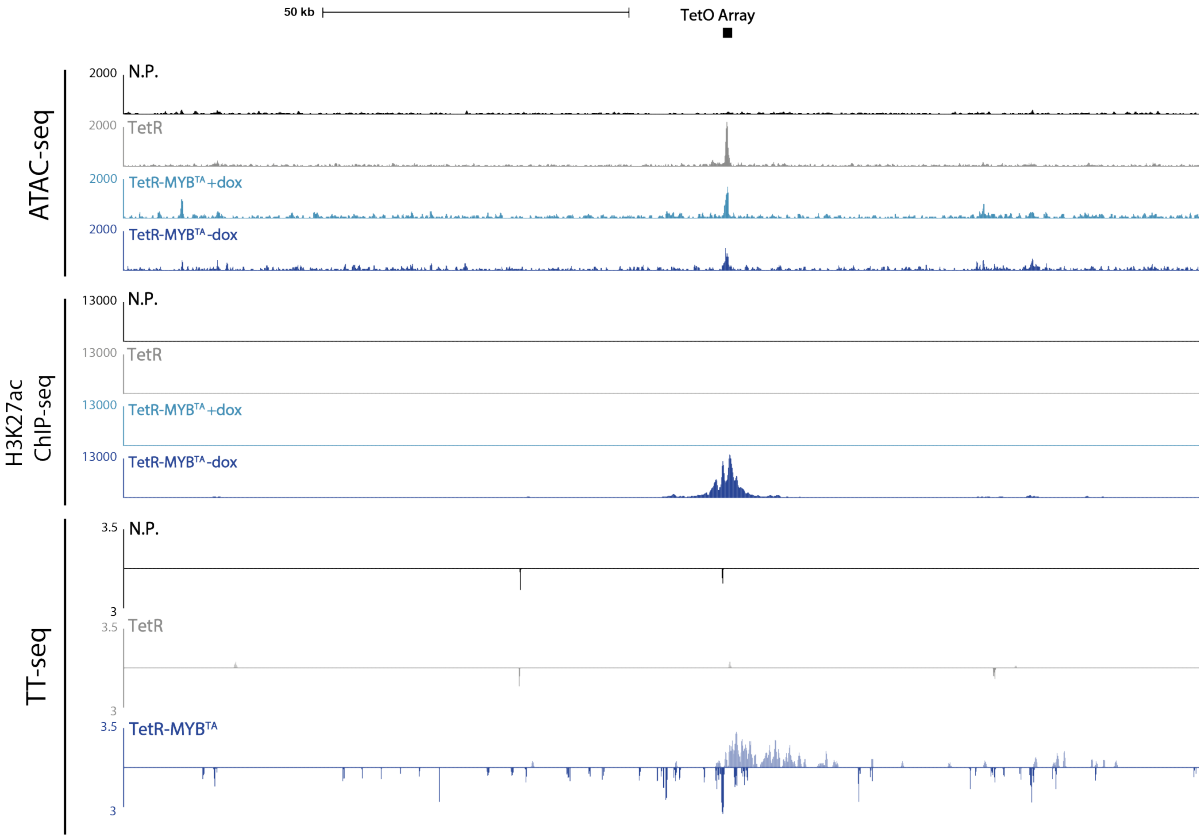


Figure 3.12 Binding of TetR-MYB^{TA} to the chrX-*TetO* array is associated with increased chromatin accessibility (ATAC-seq), H3K27ac (ChIP-seq) and nascent transcription (TT-seq), restricted to the locally *TetO* array

Representative tracks from UCSC of ATAC-seq (n=2), ChIP-seq (n=1) and TT-seq (n=1). Region shown is ~150kb surrounding the chrX-*TetO*, 50kb scale and *TetO* array shown. Cells are the E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO*. N.P. = No Protein, TetR protein stably expressed, TetR-MYB^{TA} stably expressed. +dox = doxycycline treatment for 3 hours.

3.3 Discussion

In this chapter, I demonstrated that the TA domain of MYB, a key haematopoietic TF, has the capability of generating an enhancer-like element in a neutral chromatin environment. However, the enhancer forming ability of MYB is contingent upon the genomic context within which it is bound, activating distal elements from one locus (chr8-*TetO*) but not at another (chrX-*TetO*). This implies that the surrounding DNA sequence plays a pivotal role in enhancer function.

We generated a fusion protein by combining the MYB TA domain with TetR DNA binding domain. The TetR-MYB^{TA} fusion protein exhibits robust binding in our initial tethering system, the TetR binding site on chromosome 8 (chr8-*TetO*). Strikingly, cryptic regulatory sites located up to 50kb away from where MYB^{TA} was bound were activated and transcription initiated here. To explore the potential context-dependency of this observation, I introduced a novel TetR binding site in a distinct chromatin region. This specific locus had been extensively characterised by the Hughes group, with previous results indicating that inserting an endogenous enhancer here resulted in robust activation of distal, cryptic regions. Despite comparable levels of MYB^{TA} binding to the newly engineered TetR binding site on chromosome X (chrX-*TetO*) as at chr8-*TetO*, the chromatin was only activated locally, and distal regions were not activated in this case. We conclude this may be due to specific TF binding motifs that are present at chr8-*TetO* that allow these sites to be activated, whilst chrX-*TetO* may not possess these.

Engineering the TetO *array on chrX*

The TetR system can be interrogated to explore *de novo* protein interactions within a chromatin context, while circumventing the complexities introduced by other transcription factors and endogenous regulatory elements.

The primary objective for introducing an additional *TetO* locus was to investigate whether the enhancer-forming ability of MYB was contingent upon the surrounding context. Simultaneously, there were certain limitations associated with the original *TetO* site on chromosome 8. Firstly, the *TetO* array had initially been incorporated into a BAC containing around ~180kb of 'neutral' human chromatin before insertion into mouse chr8. The potential issue arises from the observation that human chromatin may not behave in the same way in mouse cells and features such as abnormal DNA methylation might be present. The Klose lab, who originally engineered this locus, carried out bisulfite sequencing in the DNA surrounding the engineered *TetO* array and showed that the CpG dinucleotides maintain the methylation patterns observed in human tissue, upon stable integration into the mouse genome (Blackledge *et al.*, 2014). This suggests that normal chromatin features are maintained in the integrated DNA sequence. However, to circumvent this potential issue for my new TetR binding site, we identified a neutral chromatin region in chrX of mESCs and directly engineered the ~600bp *TetO* array into this site.

Secondly, the original *TetO* site on chr8 incorporated a NeoR resistance gene for positive clone selection. *NeoR* has its own promoter, which could potentially interact with the *TetO* array and influence transcription – emphasising the importance of using N.P and TetR only controls. When I engineered the new *TetO* array on chrX, I employed CRISPR-Cas9 HDR editing, eliminating the need for a resistance marker and providing a cleaner context to study MYB's isolated activity.

Engineering a new *TetO* array came with additional challenges. In ChIP-qPCR experiments, it would be impossible to distinguish the two *TetO* arrays using the same primer pair that covers the *TetO* array (Primer *TetO* Fw and Rv). Therefore, when designing the HDR donor plasmid, I made two point mutations within the *TetO* array, in the intervening sequence between *TetO* repeats, rather than the *TetO* sequence itself,

and designed a new primer pair specifically for the chrX *TetO* region. For ChIP-seq experiments, sonication fragments chromatin to ~200-300bp. Since the *TetO* array is ~600bp is it unlikely that a single read will contain information solely from the *TetO* array without at least some sequence spanning into the unique junctions. Even so, reads that map to multiple genomic loci will be filtered out during analysis.

Use of ES cells as a model system

My results indicate that tethering MYB to chromatin at the chrX-*TetO* was sufficient in activating the chromatin proximal to its binding site but insufficient to induce gene activation distally. Importantly, while MYB^{TA} was unable to activate distal elements, previous work (Harrold, C. 2019) has demonstrated that insertion of the R2 alpha globin enhancer was able to activate transcription at this locus. This argues that the lack of enhancer-like activity with MYB^{TA} was not attributed to the inherent transcriptional inactivity of the new locus. Importantly, however, the R2 enhancer only becomes active when the mESCs are differentiated into embryoid bodies (Georgiades *et al.*, 2023). This opens the possibility that the new chrX neutral region may be enriched for erythroid-specific TF motifs, such as GATA1 and KLF1 and their binding may be crucial for activation of the locus after R2 insertion. Therefore, this region may not be able to be activated in ES cells, and it is possible that MYB^{TA} may be able to activate transcription of this locus in an erythroid context. Further motif analysis in the new chrX-*TetO* region compared to the original chr8-*TetO* region could provide further insights into this hypothesis.

Is histone acetylation coupled with chromatin accessibility and transcription?

A novel and exciting result emerged when introducing a 3-hour doxycycline treatment, removing the TetR-MYB^{TA} from the *TetO* array. This is a major advantage to the TetR system, allowing us to investigate the role of MYB^{TA} in maintenance of enhancer activity by observing how long enhancer associated histone marks and chromatin

accessibility remain after TetR-MYB^{TA} dissociation from chromatin. After 3 hours of doxycycline treatment, we observed a decrease in H3K27ac levels while chromatin accessibility persisted. This finding agrees with recent research using a CBP/P300 bromodomain inhibitor which results in global H3K27ac loss but little change in chromatin accessibility (Raisner *et al.*, 2018). In agreement, mESCs that have a mutation in H3.3K27 to H3.3K27R and therefore do not display H3K27ac also show little change in chromatin accessibility (Zhang *et al.*, 2020). This suggests that once an active enhancer is established, high levels of H3K27ac are not required to maintain chromatin accessibility, at least in the short term. Chromatin accessibility may instead be maintained by TFs and co-activators or other histone marks in the absence of H3K27ac. Interestingly, regions that maintain high levels of chromatin accessibly after CBP/P300 depletion using shRNAs in mESCs are highly enriched for lineage-specific transcription factors (Martire *et al.*, 2020). Additionally, it has been previously suggested that genes with a high concentration and diversity of co-activators at their promoter may be more resistant to loss of CBP/P300 and H3K27ac (Kasper *et al.*, 2010).

An intriguing next step would be to repeat the TT-seq experiment including a doxycycline treatment as well, to examine if maintenance of transcription is dependent upon the continued presence of H3K27ac. By carrying out TT-seq, it should be possible to pinpoint which stage of transcription is first affected, whether it be initiation or elongation. This would indicate whether the enhancer element remains active in the absence of TetR-MYB^{TA}, or if the retention of chromatin accessibility is insufficient to drive transcription.

Further Characterisation of the TetO regions

A prospective avenue for further investigation involves characterising the chromatin profile at newly activated peaks. ATAC-seq showed that the distally activated peaks at chr8-*TetO* were already accessible, and the addition of TetR-MYB^{TA} further enhanced their activation. It is possible that this indicates that TFs are bound prior to activation of the locus, and that higher levels of/more stable TF binding at these sites is induced by MYB^{TA}. Previous work in the chrX region demonstrated that inserting the R2 erythroid enhancer into the chrX neutral region activated distal regions exclusively in embryoid bodies, which were already marked by H3K4me1 in the WT, unedited embryoid bodies (Figure 3.4). This aligns with the finding that H3K4me1 is observed at 'poised' developmental enhancers without H3K27ac and no detectable transcription (Creyghton *et al.*, 2010; Rada-Iglesias *et al.*, 2011; Zentner, Tesar and Scacheri, 2011; Bonn *et al.*, 2012; Won *et al.*, 2012; Malik *et al.*, 2014; Wang *et al.*, 2016; Bleckwehl *et al.*, 2021). This suggests that H3K4me1 may mark sites that are inherently primed for activation, often called 'cryptic' sites. Together, these results suggest that it may be possible to predict latent regulatory sites by the presence of chromatin accessibility, or the presence of H3K4me1.

Furthermore, as mentioned earlier in the discussion, motif analysis will be crucial to dissect why the chr8-*TetO* region was able to be activated after introducing TetR-MYB^{TA} while the surrounding chromatin at chrX-*TetO* remained silenced. As a starting point, motif-based sequence analysis tools including MEME suite could be employed to detect embryonic and erythroid transcription factor binding sites across a 100kb region surrounding the *TetO* array, comparing the chr8-*TetO* to the chrX-*TetO* regions.

Chapter 4: CBP/P300 are necessary for the maintenance of enhancer activity

4.1 Introduction

CBP and P300 are metazoan lysine acetyltransferases (KATs), widely recognised for their contributions to enhancer-mediated gene activation (Dancy and Cole, 2015; Long, Prescott and Wysocka, 2016; Weinert *et al.*, 2018). They acetylate histone and non-histone proteins (Zeng *et al.*, 2008; Tie *et al.*, 2009; Pasini *et al.*, 2010; Jin *et al.*, 2011; Dancy and Cole, 2015) and are ubiquitously expressed in nearly every tissue type in mammals, highlighting their central importance as key transcriptional regulators.

In addition to their catalytic role, CBP and P300 interact with over 400 protein partners and form central nodes in gene regulatory networks (Bedford *et al.*, 2010). In this chapter, I set out to explore the requirement of CBP and P300 for enhancer activity genome wide by depleting CBP and P300 protein levels in mESCs. Furthermore, I expand on results presented in chapter 3, to see if the ability of MYB to maintain enhancer activity is contingent upon the presence of CBP and P300.

4.1.1 CBP and P300 Structure and Function

CBP and P300 have highly conserved functional domains. There is overall approximately 60% sequence similarity between CBP and P300 in both mouse and human due to the sequence variability in the intervening intrinsically disordered regions (Liu *et al.*, 2008; Zeng *et al.*, 2008; Dancy and Cole, 2015; Rühlmann *et al.*, 2019). The overall structure and functional domains are summarised in Figure 4.1.

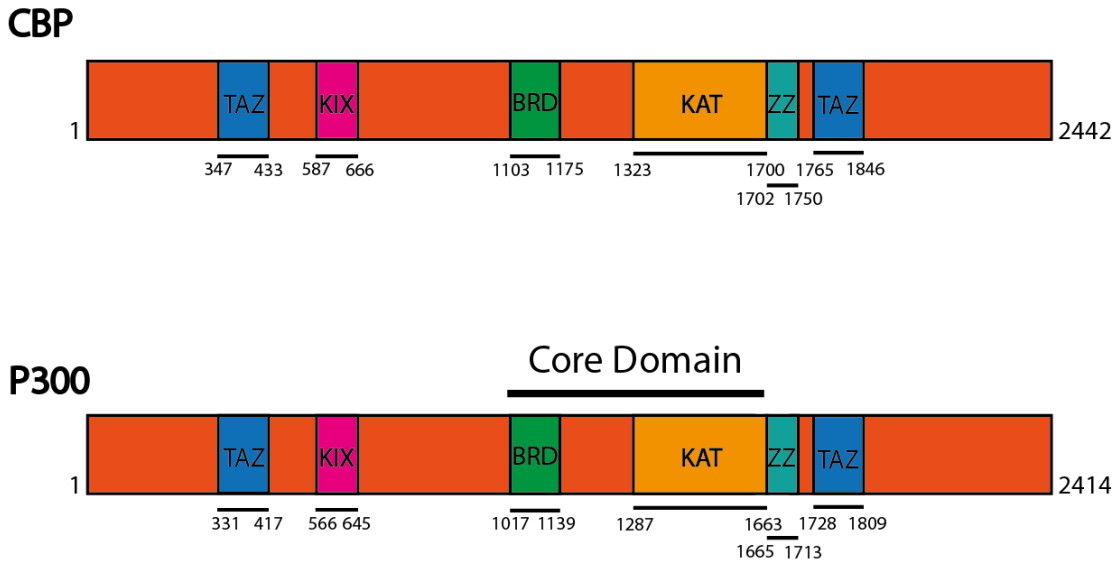


Figure 4.1 Schematic representation of the CBP and P300 protein

In mammalian cells, CBP and P300 have 7 globular domains: 2 TAZ and 1 ZZ zinc finger domains (blue and teal respectively), KIX domain (pink), bromodomain (BRD) (green), CH2 domain (contains PHD domain and RING finger domain (not shown) and lysine acetyltransferase domain (KAT) (yellow), which includes an autoinhibitory loop.

CBP – Total protein size is 2442 amino acids.

P300 - Total protein size is 2414 amino acids. The Core domain includes the BRD domain and the KAT domain (Hilton, 2015).

CBP and P300 include domains with three different functions:

1. Bromodomain: Enables recognition and binding to chromatin/proteins that have already undergone acetylation.

2. KAT domain and CH2: Responsible for acetylating both chromatin and non-histone proteins and self-acetylation.

3. KIX, TAZ and ZZ domains: Important for interacting with key transcriptional regulators.

Amino acid residues are shown for the human CBP and P300, information obtained from UniProt, 2023.

Both CBP and P300 harbour 7 globular domains which exhibit high sequence similarity (Dyson and Wright, 2016). The KAT domain is responsible for the catalytic acetyltransferase function and contains a disordered, autoinhibitory loop within it (Thompson *et al.*, 2004). The bromodomain is important for recognising and binding acetylated residues on histone and non-histone proteins (such as CREB and p53 (Zeng *et al.*, 2008). CH2 contains a PHD and RING finger domain, thought to be a negative regulator of auto-acetylation (Delvecchio *et al.*, 2013; Plotnikov *et al.*, 2014). The TAZ1 domain (also known as CH1), TAZ2 and ZZ domains (also known as CH3) and KIX domains are responsible for interactions with transcription factors, the transcription machinery and other co-factors. Hence, CBP/P300 harbours a role as a scaffold or transcription bridge as well as an acetyltransferase (further reviewed in section 1.3.4). In particular, the KIX domain of CBP/P300 is notably responsible for the interaction with the TA domain of MYB (Zor *et al*., 2004; Kasper *et al*., 2002), an interaction demonstrated *in vivo* in chapter 3. In this chapter, I aim to degrade the full length CBP and P300 proteins, abrogating both acetyltransferase and protein-protein interaction functions, in order to observe the impact on transcription.

4.1.2 CBP and P300 functional redundancy

CBP and P300 are paralogous lysine acetyltransferases (KATs) that are thought to have mainly overlapping activities in the cell. However, recent studies have revealed that despite high homology, they also harbour some unique roles. CBP and P300 have different specificities for lysine substrates *in vitro* (Henry, Kuo and Andrews, 2013; Henry *et al.*, 2015). They also act in distinct processes *in vivo* experiments (Oliveira *et al.*, 2006; Fauquier *et al.*, 2018) including myoblast differentiation, muscle integrity and motor skill learning. It is tempting to speculate that the sequence variability outside of the functional domains may be driving their differences in function.

In vivo experiments show that homozygous deletions of either CBP ($Cbp^{-/-}$) or P300 ($p300^{-/-}$) in mice is embryonic lethal, suggesting they have distinct functions in embryonic development (Yao *et al.*, 1998; Tanaka *et al.*, 2000). Conversely, by knocking down CBP and P300 individually in mouse ESCs, they have functionally redundant roles in maintaining self-renewal (Fang *et al.*, 2014), which may suggest key roles for the proteins in development.

In some cases, CBP and P300 mutations can be associated with distinct diseases. CBP is commonly mutated in cases of relapsed childhood acute lymphoblastic leukaemia (ALL) whereas P300 is less frequently affected (Waanders *et al.*, 2020). Furthermore, mice with homozygous mutations in the KIX domain of P300 develop anaemia, B-cell deficiency, thymic hypoplasia, megakaryocytosis and thrombocytosis, whereas mice with the same mutation in CBP are healthy (Kasper *et al.*, 2002) .

CBP and P300 appear to have common and unique transcriptional targets. Individual knockdown of CBP and P300 using shRNA in mouse ES cells causes transcriptional dysregulation of distinct gene subsets (Martire *et al.*, 2020). These distinct transcriptional targets may contribute to the different functions performed by each protein. Despite these observations, it is still unclear to what extent CBP and P300 have different functions at the molecular level and whether they act differently at chromatin.

4.1.3 The dTAG system

To investigate the function of CBP/P300 in enhancer activity, I planned to degrade full length CBP and P300, in single or combination degradation models, and assess the effects on enhancer features. In contrast to experiments specifically targeting the acetyltransferase domain (e.g. small molecule inhibitors), these experiments would

investigate the effect of eliminating not only histone acetylation deposited by their catalytic domain, but also the acetyl 'reader' and scaffolding functions exhibited by the other functional domains.

I employed the dTAG system (Nabet *et al.*, 2018) for inducible degradation of CBP. This involves engineering the endogenous copies of the gene of interest, so that the protein is tagged in-frame with an inactive, mutant version of $FKBP12^{F36V}$, to produce a degron cell line. dTAG-13 is then added, for inducible degron of the tagged protein. dTAG-13 is a heterobifunctional degrader, one part selectively recognising the $FKBP12^{F36V}$ tagged protein and a thalidomide moiety that induces cereblon-dependent degradation for subsequent ubiquitination of the target protein, leading to proteasomal degradation (Figure 4.2)

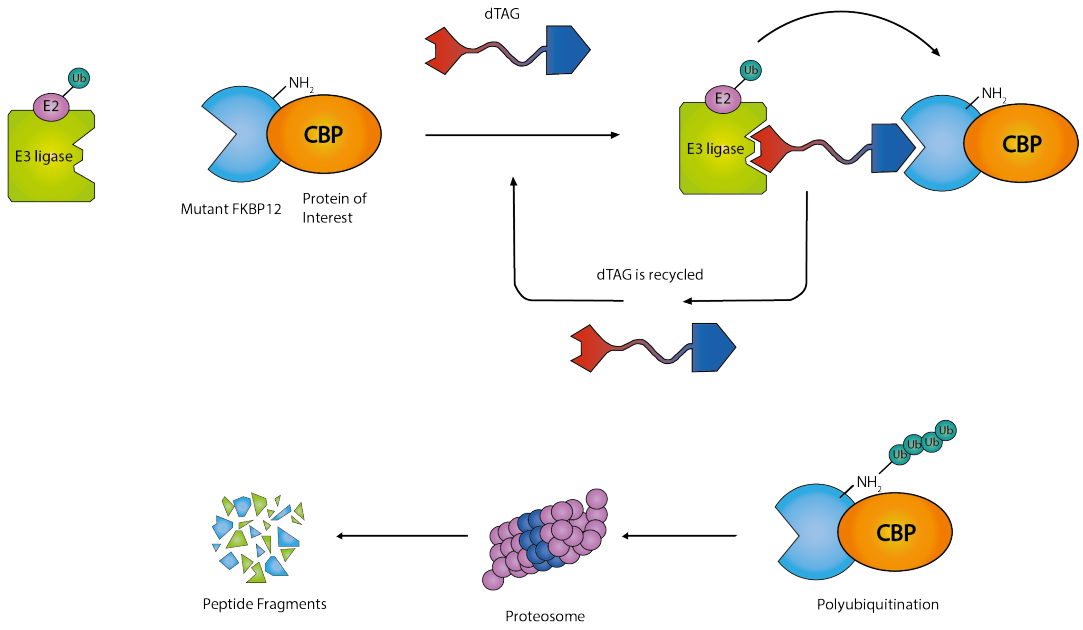


Figure 4.2 dTAG-13 mediated protein degradation

This schematic depicts the process of dTAG-13 mediated degradation and machinery involved. The protein of interest, in this case CBP, is tagged with $FKBP12^{F36V}$. dTAG-13 is added to the cells, which recognises the $FKBP12^{F36V}$ domain of the edited protein and also recruits the

endogenous E3 ubiquitin ligase, cereblon. This leads to polyubiquitination of the CBP-FKBP^{F36V}, targeting the protein for proteasomal degradation. Figure made using Illustrator and BioRender.

The dTAG system offers distinct advantages over traditional knockouts and knockdown techniques for essential proteins, where prolonged depletion affects cell viability (Zuber *et al.*, 2011). An additional advantage is that depletion of the target protein occurs rapidly in the dTAG system, as it operates at the protein level, in contrast to knockouts and knockdowns which rely on natural protein turnover to reduce the protein pool.

4.1.4 PROTAC system

The proteolysis-targeting chimera (PROTAC) system is analogous to the dTAG system however does not require genome engineering. The thalidomide component is retained but the FKBP12^{F36V} binding module is replaced with a domain that recognises the endogenous protein of interest (Figure 4.3).

A PROTAC which targets both CBP and P300, dCBP-1, was recently developed (Vannam *et al.*, 2021), based on the bromodomain inhibitor GNE-781. It binds the bromodomain of both proteins and directs them to cereblon for ubiquitination and proteasomal degradation via the thalidomide component (Figure 4.3). This degradation process results in a dramatic loss of oncogenic enhancer activity in multiple myeloma cells.

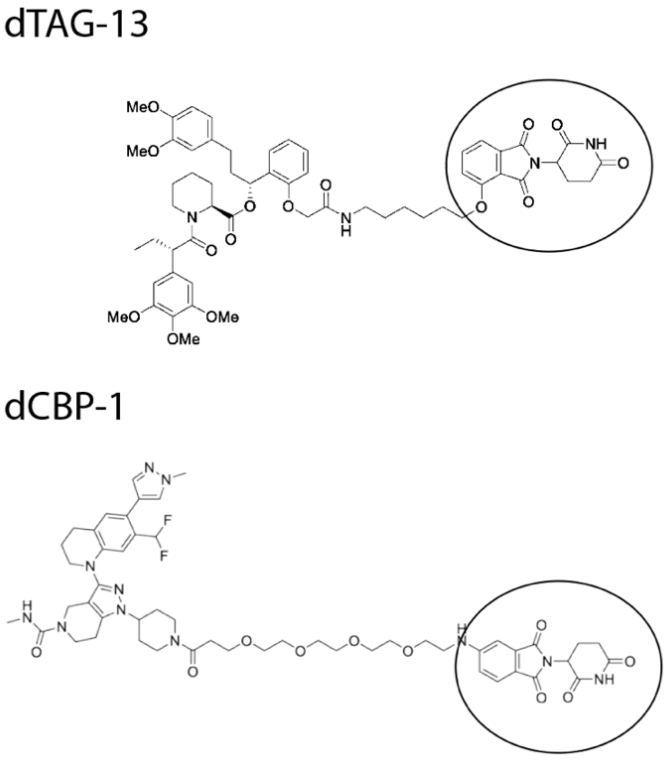


Figure 4.3 Comparison of the chemical structure of dTAG-13 and dCBP-1 (PROTAC)

Chemical structure of dTAG-13 (top) and dCBP-1 (bottom). Thalidomide moiety (circled) is the same for both molecules, this directs the protein of interest to cereblon for ubiquitination and proteasomal degradation. dTAG-13 contains an $FKBP12^{F39V}$ binding module while dCBP-1 contains a bromodomain binder based on the bromodomain inhibitor, GNE-781.

4.1.5 Aims

In this chapter, I directly test the necessity of CBP and P300 in controlling enhancer activity. To do this, I used two systems: a degron cell line that allowed me to degrade Cbp alone and a PROTAC that allowed me to degrade both Cbp and p300 simultaneously. Using these systems, I aim to:

1. Investigate the individual contribution of CBP to enhancer maintenance.
2. Investigate the combined contributions of CBP and P300 to enhancer maintenance genome wide.
3. Investigate the combined contributions of CBP and P300 to maintaining an enhancer formed by MYB.

4.2 Results

4.2.1 Using CRISPR-Cas9 HDR to generate Cbp-FKBP12^{F36V} expressing cells

To assess the requirement for CBP in enhancer function, I created a Cbp degron in mouse ES cells, so that I could degrade Cbp alone. The Cbp_FKBP12^{F36V}_HA donor plasmid (Appendix) was designed and synthesised in collaboration with Philip Hublitz of Genome Engineering, WIMM. The incorporation of an HA tag allowed me to use a high affinity HA antibody in place of the Cbp antibody to detect chromatin binding of the engineered Cbp-FKBP12^{F36V} protein in ChIP experiments.

I engineered the Cbp-FKBP^{F36V} protein using CRISPR-Cas9 homology dependent repair (HDR). I transfected the Cbp_FKBP12^{F36V}_HA donor plasmid in parallel with an gRNA incorporated into pX458, which expresses dCas9, the gRNA and eGFP. I

engineered the Cbp degron into cells that already harboured the *TetO* array on chr8 and chrX and expressed TetR-MYB^{TA}, introduced in chapter 3, so that I could assess the loss of Cbp on MYB^{TA} enhancer activity at the *TetO* regions, as well as endogenous targets. Cells were sorted on eGFP expression, with 32.3% of cells in the population displaying GFP positivity (Figure 4.4).

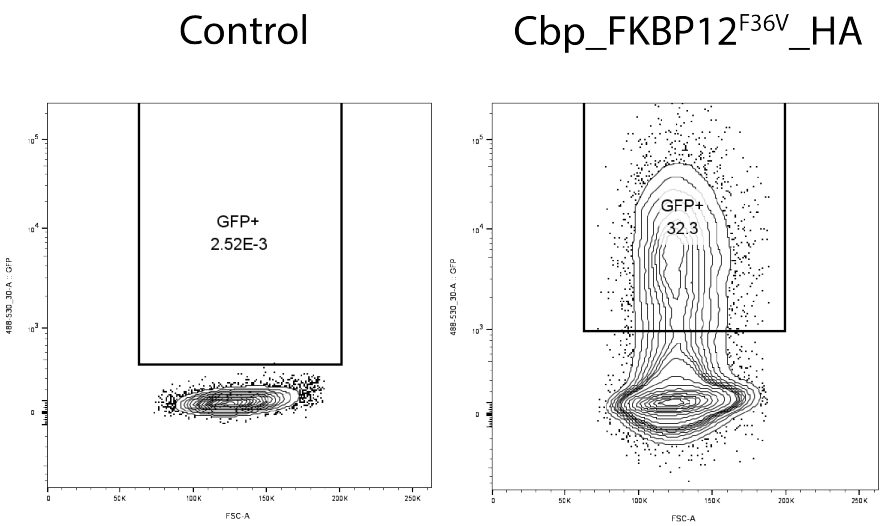


Figure 4.4 Representative data from cell sorting for eGFP+ fluorescence activated cell sorting (FACS) after transfection with Cbp_FKBP12^{F36V}_HA donor plasmid and guide plasmid containing eGFP.

Left: Control cells are wild type E14 (TOT2N variant) which were not transfected, this is used to gate for untransfected cells (GFP+ = 0.0%).

Right: Cells transfected with the donor (Cbp_FKBP12^{F36V}_HA) and the guide plasmid have a population of 32.3% GFP+ transfected cells.

Internal primers (which only amplify positive insertions) were used for an initial screen for successful editing in a pool of clones. Clones within each positive pool were further screened using spanning primers to confirm insertion of the correct size and to isolate homozygous edited clones. Sanger sequencing confirmed successful editing and ensured there were no unwanted mutations (Figure 4.5). Clone B2 was selected for use in all subsequent experiments.

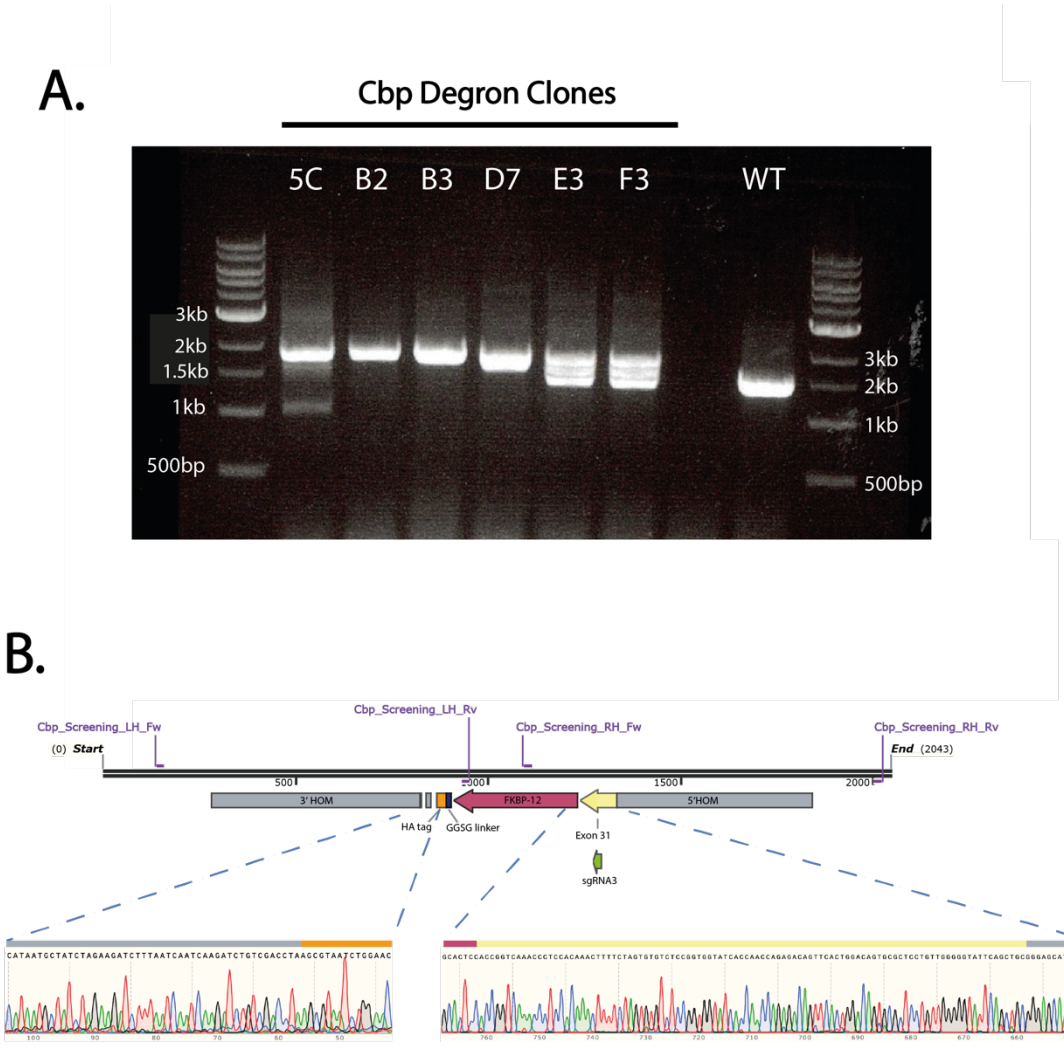


Figure 4.5 Identification of positive clone for Cbp_FKBP12^{F36V}_HA

Clones were screened using the above screening strategy, details are in Materials and Methods A | Agarose gel using spanning primers, 5C, B2, B3 and D7 appear to be positive, homozygous knock-in clones, while E3 and F3 appear to be heterozygous clones.

B | The band from clone B2 was extracted, purified, and sequenced using Sanger Sequencing to confirm the sequence of the insertion site and to avoid using a clone with any unwanted mutations. Representative sequencing traces at the insertion junctions.

4.2.2 Validation of Cbp-FKBP12^{F36V} Clones

Following confirmation of the successful integration of the FKBP12^{F36V} domain upstream of the stop codon of the endogenous *Cbp* gene, I subsequently assessed the size and expression level of the engineered protein. The tagged protein expressed in homozygous clones demonstrated a noticeable shift of the detected band towards a higher molecular weight, indictive of the increase in size following incorporation of the FKBP12^{F36V} domain.

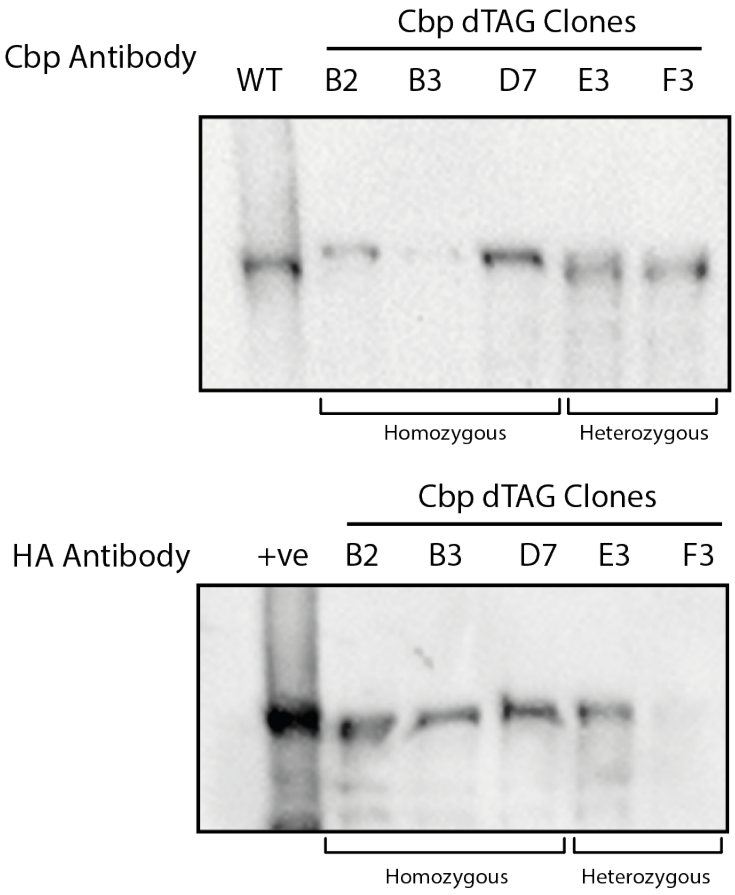


Figure 4.6 Western Blot using Cbp or HA antibody for further validation of Cbp-FKBP12 clones

Top panel: Wild type (WT) is unedited E14 TOT2N variant expressing TetR-MYB^{TA} and representative 3 homozygous clones and 2 heterozygous clones, where the wild type and the edited band are present.

Bottom panel: positive control is MLL-AF4 degron tagged SEM cell line. 3 homozygous clones and 2 heterozygous clones are represented. Ladder not shown.

As shown in Figure 4.6, the shift towards a larger protein size is evident in clones B2, B3 and D7 using the Cbp antibody. Meanwhile, clones E3 and F3 exhibit a wild type band and a larger edited band, indicating these are likely heterozygous clones, agreeing with Figure 4.5. Furthermore, I also conducted a western blot targeting the HA tag, to confirm that the protein could be detected with this antibody. A single band was present in all clones, albeit relatively weakly in clone F3. We would expect to see a single band for both homozygous and heterozygous clones as the untargeted allele in the heterozygous clones does not include the HA tag so is not detected in the western blot. As a positive control, I included lysate from an ALL cell line expressing the MLL-AF4 onco-fusion protein tagged with HA (provided by Dr Nick Crump). The ladder is not shown, as the conditions for this western blot were still being optimised and ladder did not transfer under the initial conditions. The lack of a loading control in this western blot means that I cannot infer anything about the relative expression levels of each clone or in comparison to the wild type Cbp. Clone B2 was used for subsequent experiments.

4.2.3 Degradation of Cbp at the protein level

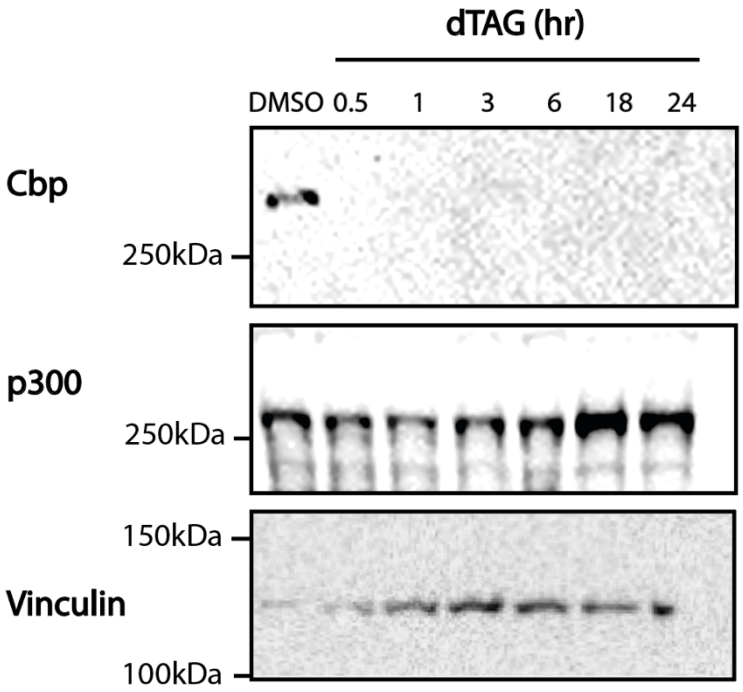


Figure 4.7 Western Blot using Cbp and p300 antibodies to assess the rate of degradation of Cbp in Cbp-FKBP12 clones

Cells are E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO*, engineered to contain a homozygous Cbp-FKBP12 (Clone B2). Cells were treated with 1μM dTAG-13 for 30 minutes, 1 hour, 3 hours, 6 hours, 18 hours and 24 hours and Cbp and p300 antibody was used to determine protein loss. Vinculin antibody is used as a loading control.

Next, I conducted a series of western blots to test the rate of Cbp protein degradation caused by dTAG-13 treatment at different time intervals. Concurrently, I investigated the possibility of changes in p300 protein levels in response to Cbp degradation, by conducting a p300 western blot. In these experiments, dTAG-13 was added at a concentration of 1μM for a range of timepoints: 30 minutes, 1 hour, 3 hours, 6 hours, 18 hours and 24 hours. Notably, a complete loss of detectable Cbp protein was observed

after 30 minutes of dTAG -13 treatment. Moreover, the protein levels of p300 appeared to increase at the 18- and 24-hour timepoints, suggesting a potential compensatory effect (Figure 4.7).

4.2.4 Degradation at the chromatin level

Having established a cell line in which Cbp can be inducibly degraded, I measured the presence of both proteins at chromatin by ChIP-qPCR following a time-course of dTAG-13 treatment. From this I could assess whether the global reduction in Cbp protein levels also resulted in a loss of Cbp binding at chromatin. I was also interested to determine whether the observed increase in p300 protein levels in response to Cbp protein degradation (Figure 4.7) led to an increase in chromatin binding.

As established in Chapter 3, there are elevated levels of Cbp and p300 at the *TetO* locus, recruited by TetR-MYB^{TA}. Consequently, this locus served an appropriate role for looking at changes in Cbp and p300 at chromatin.

At the *TetO* locus, a distinct peak of Cbp was evident at the *TetO* arrays on both chromosome 8 and X, which dissipated dramatically at the -0.5 and +0.5 positions (Figure 4.8). Notably, at the 1 hour and 3-hour dTAG treatment conditions, there was a substantial reduction in Cbp signal, consistent with the loss observed by western blotting (Figure 4.7). This was replicated with ChIP-qPCR using an antibody recognising the HA tag (Figure 4.8). At endogenous loci, Cbp and HA signal is also reduced.

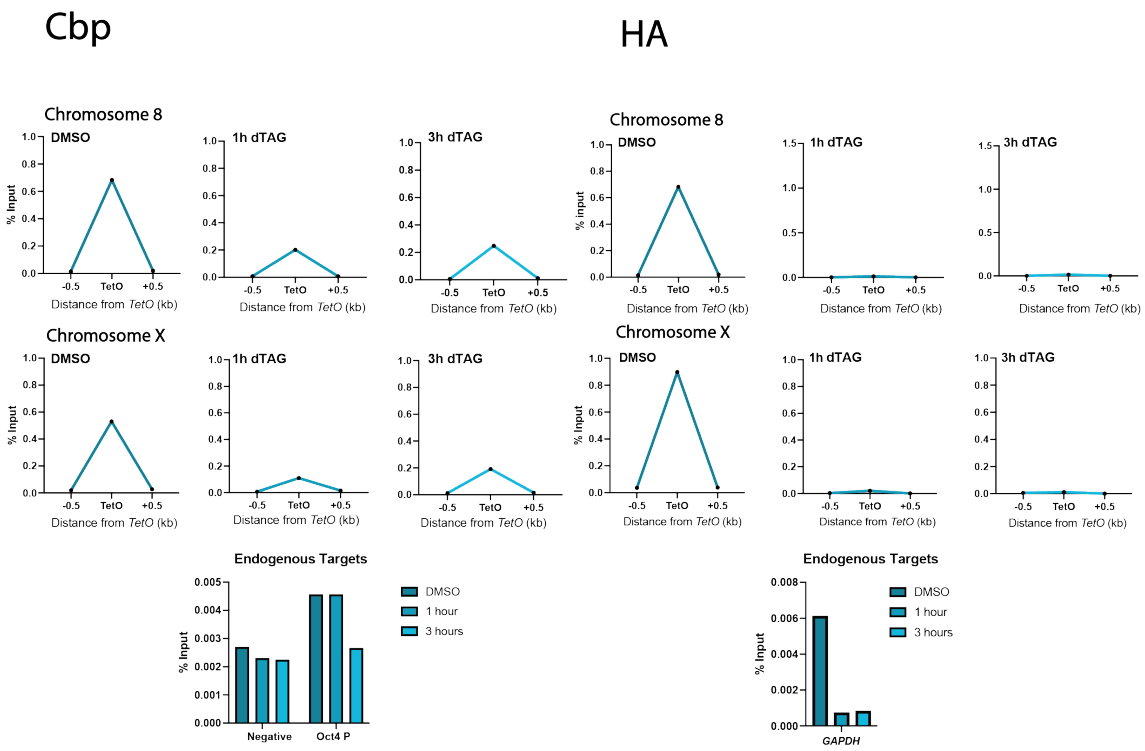


Figure 4.8: ChIP-qPCR using Cbp and HA antibodies at Chromosome 8 and Chromosome X *TetO* loci

% Input was calculated using the DDCt (ChIP versus input controls) method. X axis represents primer pairs at the *TetO* array and +/- 0.5kb from the *TetO* array.

Left: Cbp ChIP (n=1) after DMSO treatment for 3h or dTAG treatment (1μM) for 1h and 3hr in Cbp-FKBP12 E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO.*

Right: HA ChIP (n=1) after DMSO treatment for 3h or dTAG treatment (1μM) for 1hr and 3hr in Cbp-FKBP12 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO.*

As expected, ChIP-qPCR for p300 did not show a decrease following dTAG-13 treatment. Conversely, I observed an increase in p300 signal at the 3-hour timepoint. This finding aligns with the western blot observations that p300 protein levels increase at 18 and 24 hours of dTAG-13 treatment. The effect observed at 3h may be an early reflection of this increase in total protein levels, but it may also reflect increase binding of p300. It is possible that at this earlier timepoint, soluble p300 that is not binding

anywhere and is already in close proximity may bind to the chromatin, to compensate for the loss of Cbp. However, this is not reflected at endogenous targets, where p300 levels remain similar (Figure 4.9).

Next, I conducted ChIP-qPCR for H3K27ac to assess the effect of Cbp degradation on histone acetylation. There was no substantial decrease in H3K27ac signal after dTAG treatment, indicating that Cbp degradation does not cause a reduction in H3K27ac at the loci investigated; *TetO* and the *Gapdh* promoter. (Figure 4.9). The notable increase in p300 binding and the fact that H3K27ac does not change by ChIP-qPCR indicated that there may be no discernible effect on transcription, due to compensatory activity of p300 in the absence of Cbp. Consequently, I shifted my focus to investigate the simultaneous degradation of both Cbp and p300.

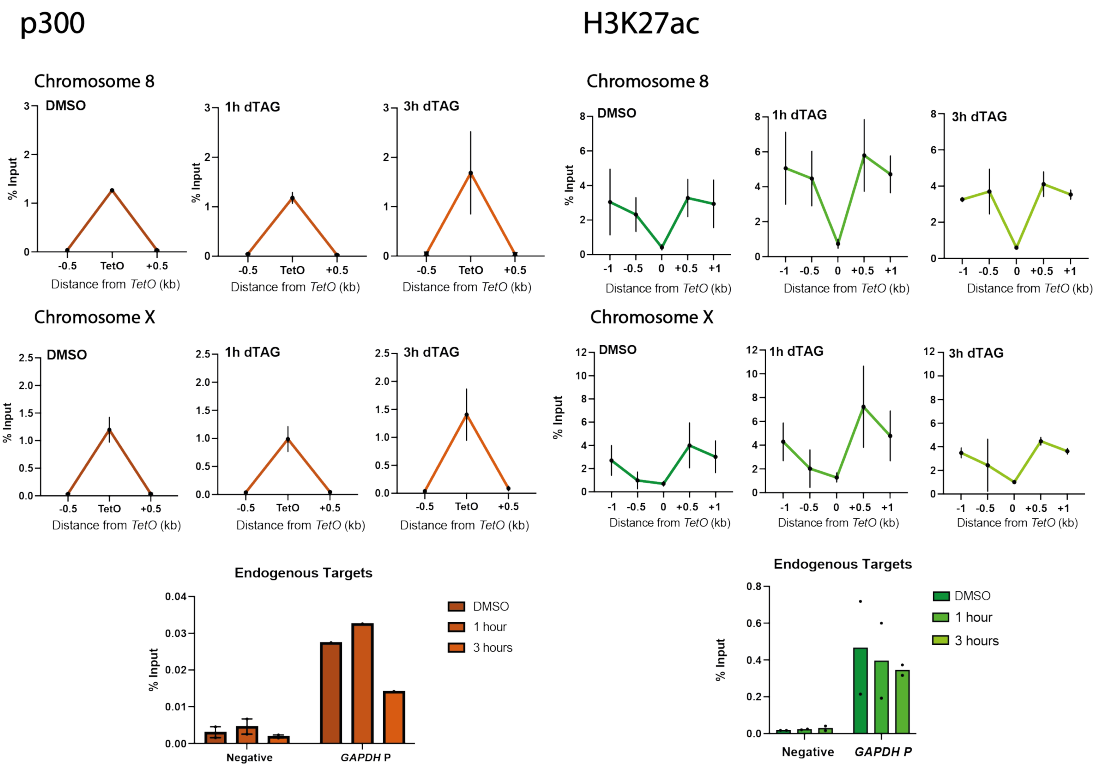


Figure 4.9 ChIP-qPCR using p300 and H3K27ac antibodies at Chromosome 8 and Chromosome X *TetO* loci

% Input was calculated using the DDCt (ChIP versus input controls) method. X axis represents of primer pairs at the *TetO* array and +/- 0.5kb from the *TetO* array.

Left: p300 ChIP (n=2), error bars indicate the range, after DMSO treatment for 3h or dTAG treatment (1μM) for 1h and 3hr in Cbp-FKBP12 E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO.*

Right: H3K27ac ChIP (n=2), error bars indicate the range, after DMSO treatment for 3h or dTAG treatment (1μM) for 1h and 3hr in Cbp-FKBP12 E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO.*

4.2.5 Optimisation of a PROTAC for degrading Cbp and p300

We employed a PROTAC, dCBP-1, that was specifically developed to target Cbp and p300 for proteasomal degradation (Vannam *et al.*, 2021). While dCBP-1 has been successfully validated in various cancer cell lines, its efficacy in mESCs has not yet been elucidated. Treatment with dCBP-1 in multiple myeloma cells (MM.1S) causes a marked decrease in both Cbp and p300 at the protein and chromatin level (Vannam *et al.*, 2021). This depletion was accompanied by a substantial decrease in gene expression, including a reduction in expression of the oncogene *MYC*, indicating that CBP and P300 play crucial roles in cancer-specific transcription in these cells.

4.2.6 Loss of Cbp and p300 at the protein level

Initially, I conducted western blot experiments to determine the optimal concentration and time point of dCBP-1 treatment required for complete loss of Cbp and p300 in the E14 mESC TOT2N variant cell line expressing TetR-MYB^{TA}. I tested the effect of dCBP-1 at two timepoints, 6 and 24 hours and two concentrations, 0.5μM and 1μM.

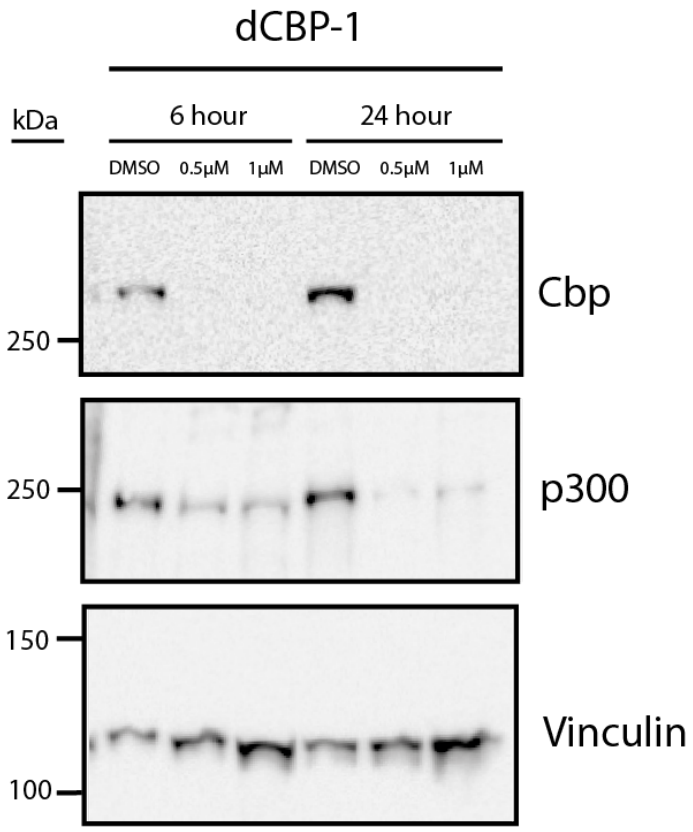


Figure 4.10 Western Blot for Cbp and p300 at various timepoints and concentrations dCBP-1 (PROTAC) treatment

Western Blot using antibodies recognizing Cbp and p300 after DMSO or dCBP-1 treatment for 6 and 24 hours and at 0.5μM and 1μM concentration of dCBP-1 (PROTAC)

There was a substantial loss of detectable Cbp at both 6 and 24 hours in both 0.5μM and 1μM dCBP-1. In the case of p300, there was a less pronounced but still noticeable reduction at both 6 and 24 hours using in both 0.5μM and 1μM (Figure 4.10). Longer treatment was associated with a slight further decrease in protein levels.

Based on these results, for subsequent experiments, I maintained the dCBP-1 concentration at 0.5μM. Given the dramatic reduction in Cbp at 6 hours, I decided to test how quickly levels would drop by using shorter time intervals: 30 min, 1 hour, 3-hour, 6 hour and 24 hours. This revealed a substantial loss of Cbp after only 1 hour of

PROTAC treatment. In contrast, p300 exhibited a delayed response, with a substantial reduction only evident at 6 and 24 hours (Figure 4.11).

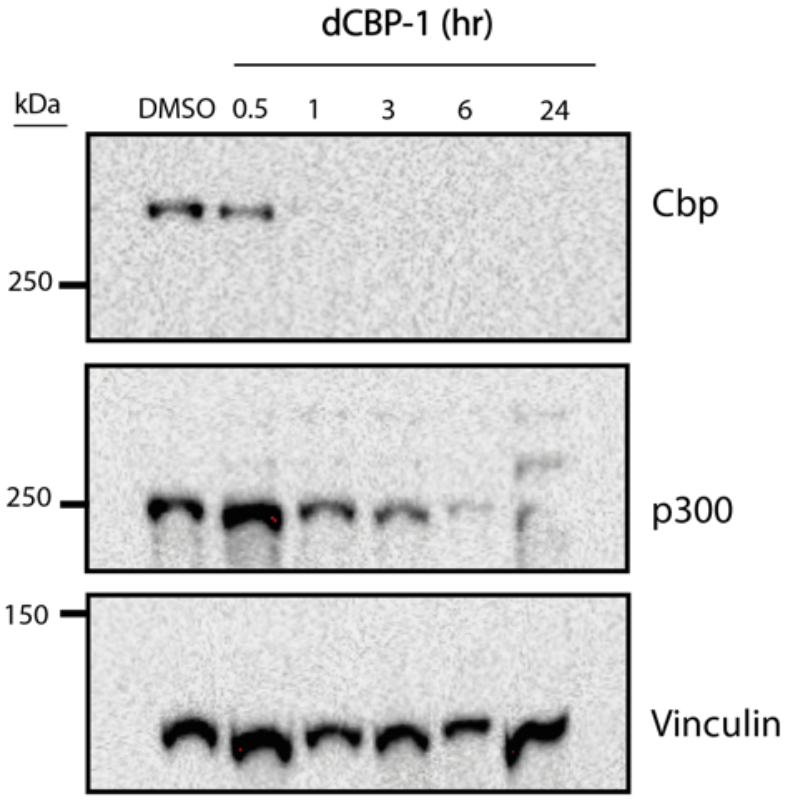


Figure 4.11 Western Blot for Cbp and p300 at different timepoints of dCBP-1 (PROTAC) treatment

Western Blot using Cbp and p300 antibodies after DMSO or dCBP-1 treatment at 30 minutes, 1 hour, 3-hour, 6-hour and 24-hour treatments at concentration of 0.5μM of d-CBP-1 (PROTAC).

Next, I used the H3K27ac antibody in a western blot experiment to assess whether acetylation levels were affected by degradation of Cbp and p300. This result revealed a substantial reduction in H3K27ac at 1, 3, 6 and 24-hour dCBP-1 treatment (Figure 4.12), consistent with the timepoints at which Cbp and p300 levels had dropped. This

demonstrated that I had a functional loss of Cbp and p300 activity at chromatin. Next, I wanted to ask what effect this had on transcription.

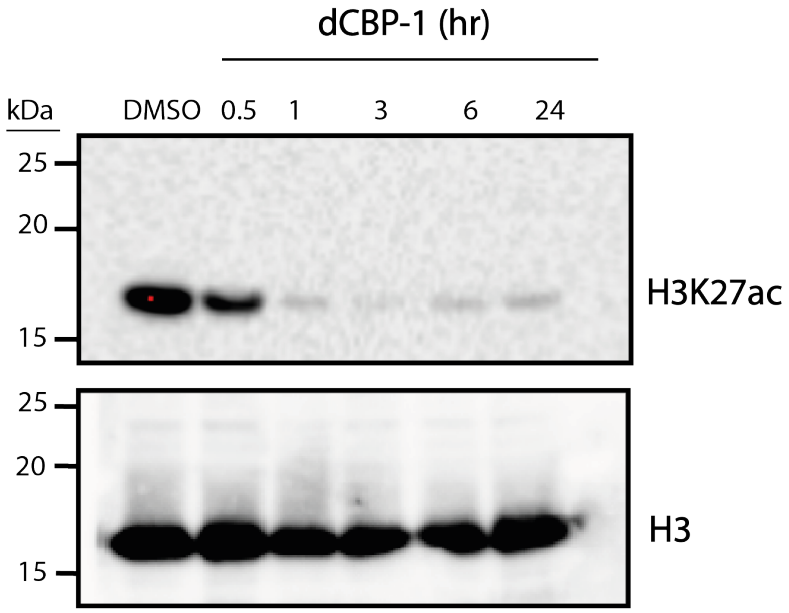


Figure 4.12 Western Blot for Histone H3 and H3K27ac at different timepoints of dCBP-1 (PROTAC) treatment

Western Blot using H3K27ac antibodies after dCBP-1 treatment at 30 minutes, 1 hour, 3-hour, 6-hour and 24-hour treatments at concentration of 0.5μM of dCBP-1 (PROTAC). H3 is used as a loading control.

4.2.7 Loss of Cbp and p300 at the chromatin level

In order to assess whether the reduction in global Cbp/p300 protein levels was associated with a loss of binding at chromatin, I conducted ChIP-qPCR with the p300 antibody. Unfortunately, I was not able to assess Cbp binding due to lack of availability of an effective Cbp antibody for ChIP. Given the fact that the two proteins co-localise at the *TetO* array (Figures 4,8, 4.9) and genome-wide (Ma *et al.*, 2021), we can reasonably infer that the Cbp would exhibit a similar response to PROTAC treatment as p300.

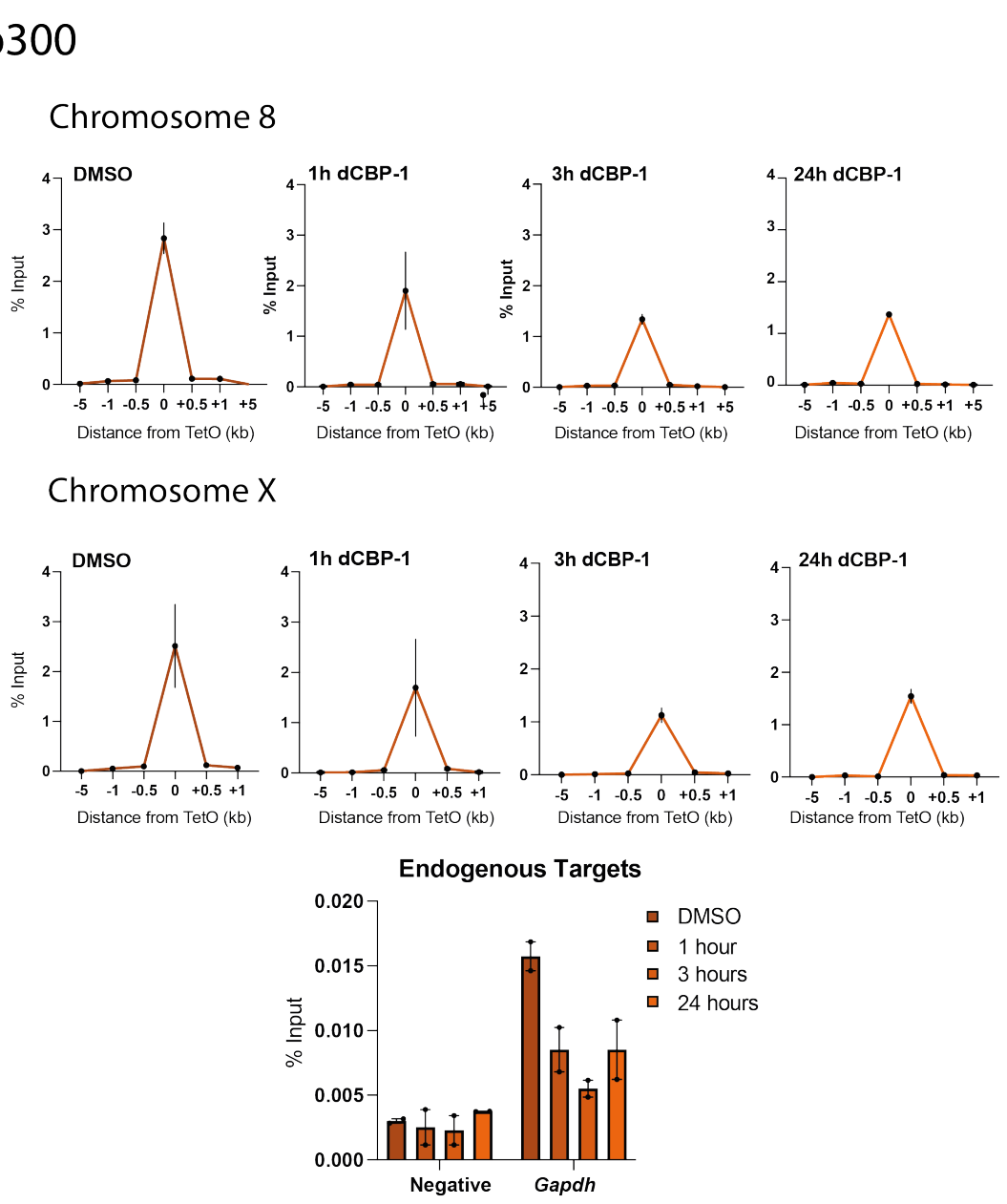


Figure 4.13 ChIP-qPCR using p300 antibody at Chromosome 8 and Chromosome X

% Input was calculated using the DDCt (ChIP versus input controls) method. X axis is comprised of primer pairs at the *TetO* array and +/-0.5kb, +/-1kb and +/-5kb from the *TetO* array, a negative control and endogenous loci.

ChIP-qPCR using p300 antibody (n=2), error bars indicate the range, after DMSO (after 24hr) or dCBP-1 treatment after 1hr, 3hr and 24hr.

Cells are E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO*, expressing TetR-MYB^{TA}

I observed a distinct peak of p300 binding at the *TetO* arrays at chromosome 8 and chromosome X. After 1- and 3-hour treatment with dCBP-1, there was a reduction, but not a complete loss, in p300 binding (Figure 4.13). This result was mirrored at the endogenous *Gapdh*.

Next, I asked whether the decrease in p300 binding at the *TetO* array was associated with a reduction in H3K27ac. However, despite the genome-wide reduction in protein level (Figure 4.12), there was a less pronounced effect on acetylation at the two *TetO* loci. There was a reduction in H3K27ac after 1-hour dCBP-1 treatment, followed by a slight increase at 3 hours, although levels were still high in comparison to endogenous loci (Figure 4.14). In contrast, there was a more prominent decrease in H3K27ac at endogenous loci, the *Oct4* and *Sox2* enhancers (Figure 4.14). This suggests that there may be selective retention of H3K27ac at the *TetO* loci following Cbp/p300 degradation, possibly because of the particular high levels of these proteins bound here even in the presence of dCBP-1 (Figure 4.13).

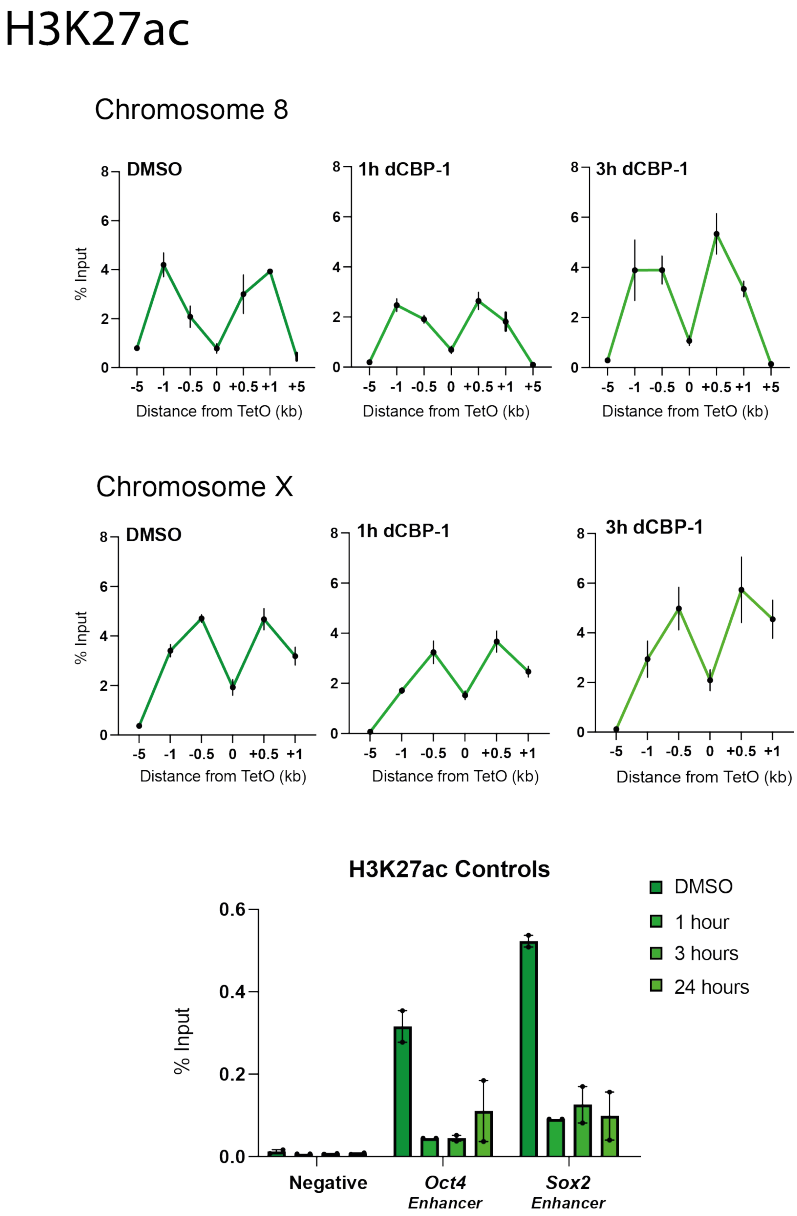


Figure 4.14 ChIP-qPCR using H3K27ac antibody at Chromosome 8 and Chromosome X

% Input was calculated using the DDCt (ChIP versus input controls) method. X axis is comprised of primer pairs at the *TetO* array and +/-0.5kb, +/-1kb and +/-5kb from the *TetO* array, a negative control and endogenous loci.

ChIP-qPCR using H3K27ac antibody (n=2), error bars indicate the range, after DMSO (after 24hr) or dCBP-1 treatment after 1hr, 3hr and 24hr.

Cells are E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO*, expressing TetR-MYB^{TA}.

4.2.8 PROTAC-mediated degradation of Cbp and p300 caused a reduction in H3K27ac genome-wide

After observing a dCBP-1 induced decrease in acetylation at the *Oct4* and *Sox2* enhancers by ChIP-qPCR, I decided to carry out referenced normalised ChIP-seq (nChIP-seq) (Orlando *et al.,* 2014) at 1hr, 6hr and 24hr dCBP-1 treatment compared to a DMSO control to investigate the genome-wide effect on H3K27ac levels across chromatin. At promoters, there was a global decrease in H3K27ac at 1 hour and a more modest decrease at 6 and 24 hours compared to DMSO (Figure 4.15). This pattern was similar to that observed at the *TetO* array by ChIP-qPCR (Figure 4.14). At enhancers, there was a global decrease in H3K27ac at all three timepoints, which looked similar to the pattern observed at *Oct4* and *Sox2* enhancers by ChIP-qPCR. This data suggests that enhancers are more sensitive to loss of Cbp and p300, displaying a stronger reduction in H3K27ac. This is in agreement with a study in mESCs using p300 shRNA, where promoters maintained H3K27ac enrichment to a greater extent than enhancers (Martire *et al.*, 2020).

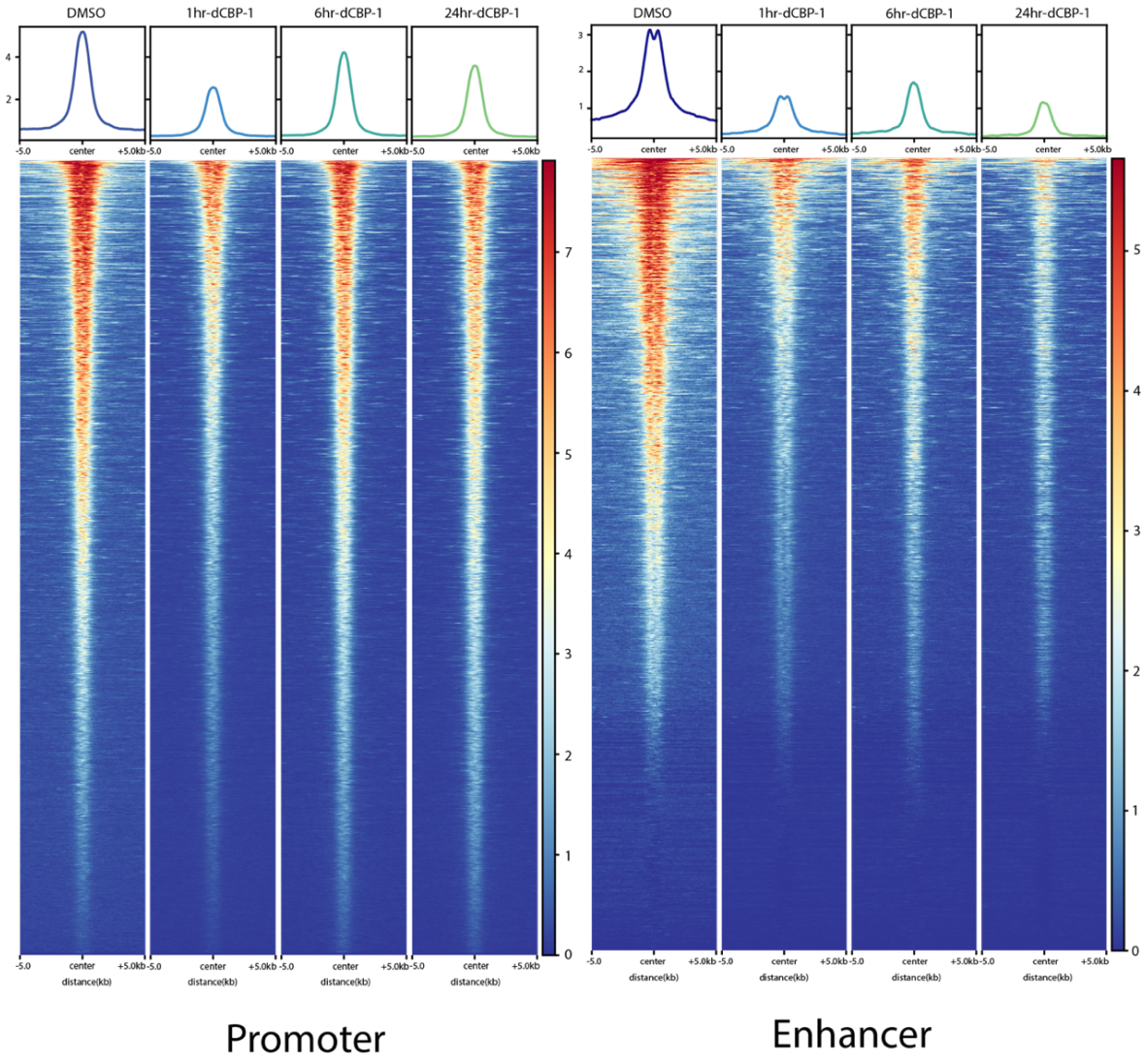


Figure 4.15 Heatmap of genome-wide H3K27ac nChIP-seq signal in a time-course experiment of dCBP-1 treatment

Left: Average H3K27ac nChIP-seq (n=1) peak profile and heatmap for each dCBP-1 treatment timepoint over a 10kb window around the promoter, defined by an ATAC-seq peak overlapping with a H3K27ac peak and less than 2kb from a TSS.

Right: Average H3K27ac nChIP-seq (n=1) peak profile and heatmap for each dCBP-1 treatment timepoint over a 10kb window around enhancers, defined by an ATAC-seq peak overlapping with a H3K27ac peak and more than 2kb from any TSS.

Heatmaps and metaplots generated with computeMatrix, plotHeatmap and plotProfile functions of deeptools.

4.2.9 PROTAC-mediated degradation of Cbp and p300 and loss of H3K27ac leads to a disruption of nascent transcription of *Sox2*

It has been reported that regions most sensitive to Cbp and p300 loss include those highly bound by ES-cell specific transcription factors by an shRNA assay (Martire *et al.*, 2020). Therefore, I chose to further investigate *Sox2*, a pluripotency specific gene driven by a super-enhancer ~100kb downstream of the promoter in mESCs (Li *et al.*, 2014).

I chose a 1-hour treatment with dCBP-1 to investigate the immediate effects of Cbp/p300 loss on transcription. This caused a reduction in H3K27ac at the promoter and enhancer of *Sox2*, an ES-cell specific gene. This was coupled with a reduction in nascent transcription from the *Sox2* promoter and eRNA transcription from the *Sox2* super-enhancer (Figure 4.16). This provides novel evidence that PROTAC-mediated depletion of H3K27ac at an endogenous ES-cell specific super-enhancer causes a loss of both eRNA transcription and transcription of its target gene.

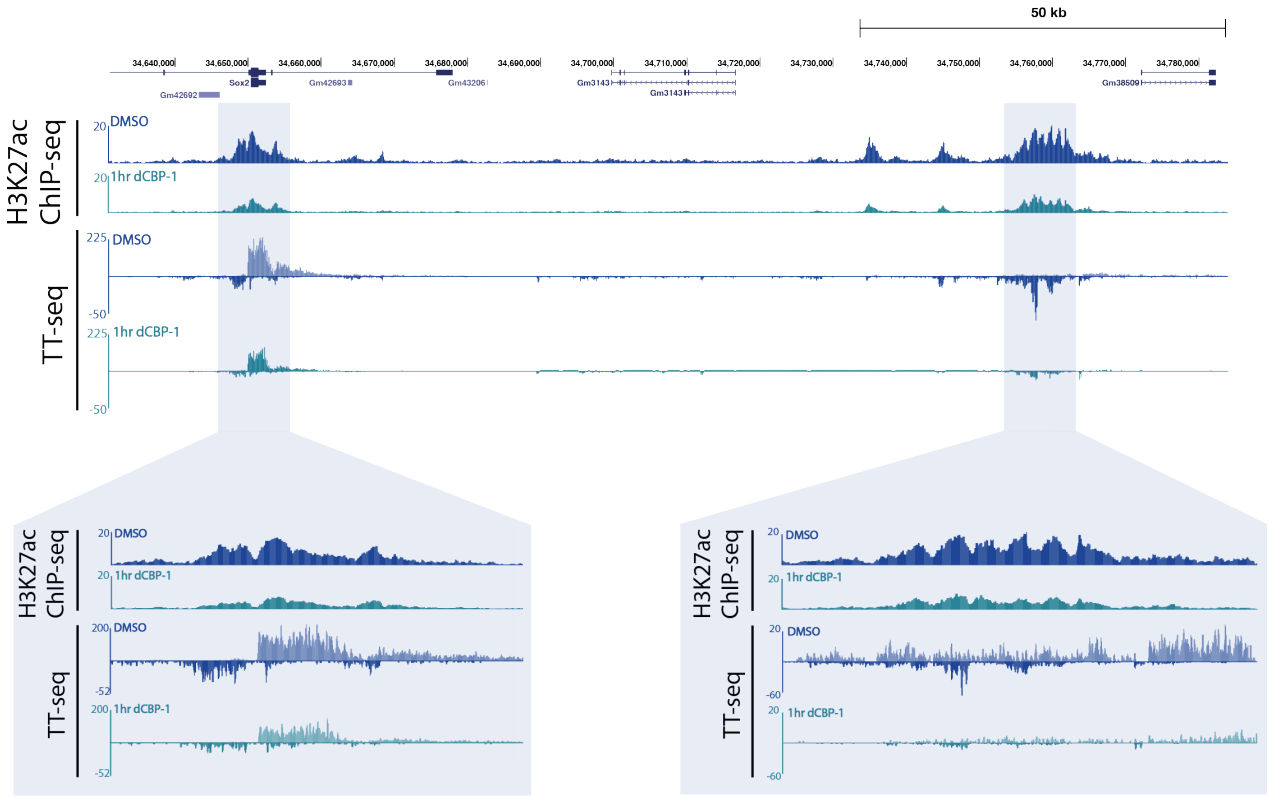


Figure 4.16 PROTAC-mediated p300 and Cbp degradation causes reduction in H3K27ac at the *Sox2* promoter and enhancer and a concurrent reduction in nascent transcription

Representative tracks from UCSC of H3K27ac nChIP-seq (n=1), referenced normalised TT-seq (n=3) after 1 hour of 0.5μM PROTAC treatment. Region shown is ~150kb surrounding the *Sox2* gene, including the *Sox2* distal super-enhancer (~100kb from *Sox2* gene). 50Kb scale shown. Highlighted are *Sox2* promoter and *Sox2* distal super-enhancer.

Cells are E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO* expressing TetR-MYB^{TA}.

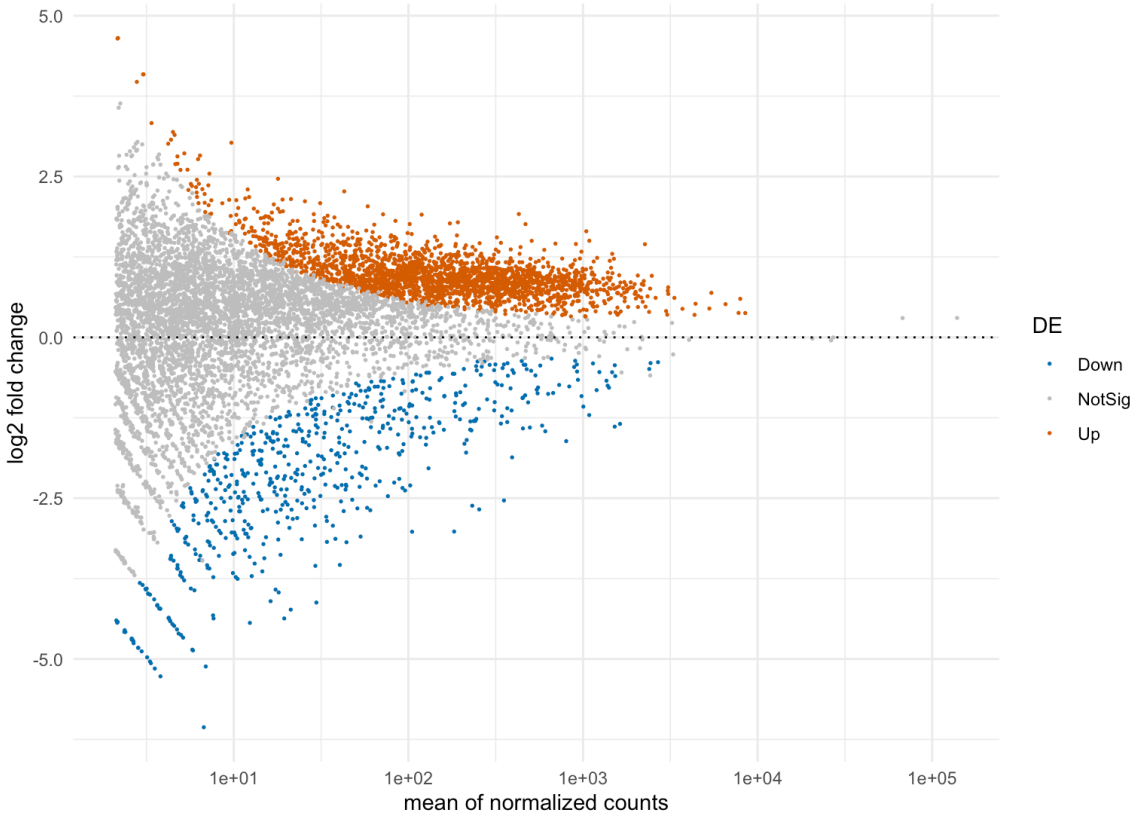


Figure 4.17 PROTAC-mediated Cbp and p300 degradation causes global gene expression changes

MA plot of referenced normalised TT-seq data (n=3) 1-hour PROTAC treatment v DMSO. Orange datapoints indicate upregulated genes, blue datapoints indicate downregulated genes and grey datapoints show no significant changes in transcription (p-adj < 0.05).
Cells are E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO* expressing TetR-MYB^{TA}.

dCBP-1 globally decreases H3K27ac, a histone mark highly correlated with gene activation. Therefore, there is a high chance treating cells with dCBP-1 would result in global transcriptional downregulation. Since DESeq2 normalisation assumes that there are no global differences across conditions, it is possible using DESeq2 may wrongly

identify gene expression changes. Therefore, to mitigate this issue, I performed referenced normalised TT-seq, adjusting the DESeq2 scaling factors by the relative proportions of spike-in exogenous RNA to normalise for global transcriptional changes between DMSO and dCBP-1 condition.

On a genome wide level, there is both upregulation and downregulation of global transcription (Figure 4.17). There were 2245 upregulated genes, but these were upregulated to a lesser degree (log-fold change) than the 615 downregulated genes. 5071 genes were not significantly affected by the PROTAC treatment. This argues that not all genes are dependent upon Cbp/p300 for transcription.

4.2.10 PROTAC-mediated degradation of Cbp and p300 and loss of H3K27ac leads to an impairment of MYB TA to act as an enhancer

In chapter 3, I showed that the MYB TA domain, fused to the TetR DNA binding domain (TetR-MYB^{TA}), can activate loci distal to the chr8-*TetO* array, analogous to an endogenous enhancer. One explanation is that this is due to its ability to recruit Cbp and p300, leading to H3K27ac deposition, a histone modification classically used to identify and mark enhancers (Weinert *et al.*, 2018). Using the H3K27ac nChIP-seq and referenced normalised TT-seq performed after 1hr PROTAC treatment in our E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO* and expressing TetR-MYB^{TA}, I was able to assess the effect H3K27ac loss not only genome-wide at endogenous ES cell targets but also at the *TetO* array (Figure 4.18).

Consistent with the ChIP-qPCR (Figure 4.14), I found that 1 hour of dCBP-1 treatment results in a modest decrease in H3K27ac by nChIP-seq at the *TetO* array. Surprisingly, compared to the reduction seen at the *TetO* array, which was minimal, this effect was

more evident at the distal region highlighted in Figure 4.18, which corresponds to ATAC peak 3 in Figure 3.3 and Figure 3.11. Concurrently, there was a modest reduction of nascent transcription from these distal sites, suggesting that recruitment of Cbp/p300 and H3K27ac, at least in part, contributes to MYB^{TA}-driven enhancer activity (Figure 4.18). However, Cbp/p300 degradation does not revert the *TetO* region back to its original state, where there are no clear peaks of H3K27ac and no evidence of transcription (Figure 3.11). This suggests there are other transcription factors or co-factors interacting with MYB^{TA} to allow MYB to maintain an enhancer-like element or that the residual Cbp/p300 and H3K27ac observed at the *TetO* array is sufficient to retain some of the enhancer activity.

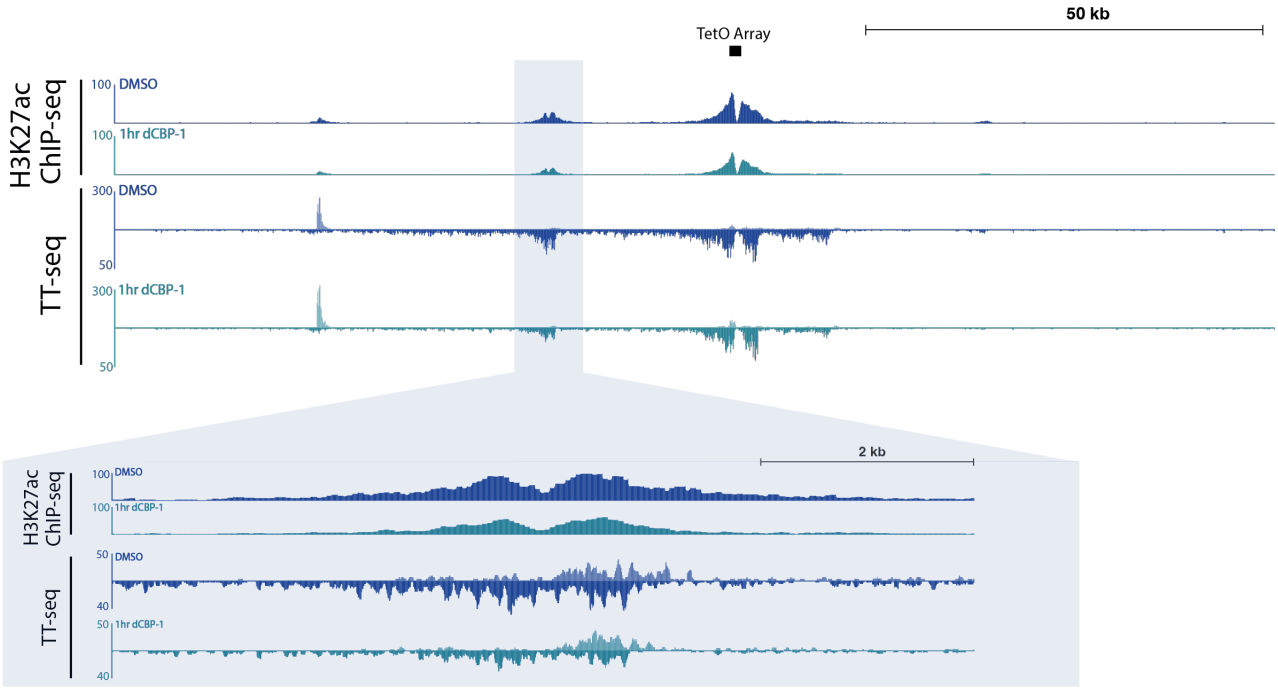


Figure 4.18 Loss of H3K27ac by PROTAC-mediated degradation of Cbp and p300 leads to a reduction in MYB^{TA}-driven enhancer activity, thereby reducing nascent transcription at distal loci

Region shown is ~150kb surrounding the *TetO* array at Chromosome 8. Representative tracks from UCSC of H3K27ac nChIP-seq (n=1), referenced normalised TT-seq (n=3) after 1 hour of DMSO or 0.5μM PROTAC treatment.
Cells are E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO*, expressing TetR-MYB^{TA}.

4.3 Discussion

In this chapter, I aimed to investigate the roles of CBP and P300 in enhancer-mediated transcription. I engineered a degron cell line tagging Cbp for dTAG-mediated degradation, to dissect the role of Cbp alone in transcription. After obtaining evidence that this may lead to compensatory effects i.e. an increase in p300 activity, I made use of a PROTAC for the degradation of both Cbp and p300 (dCBP-1). Treatment of ES cells with dCBP-1 lead to a reduction in histone acetylation, as evident by H3K27ac western blot and ChIP experiments and distinct global gene expression changes. Among those genes that were downregulated, there was a reduction in transcription from the enhancer and promoter of the ES-cell specific gene *Sox2*, as well as from its enhancer. Concurrently, I built on results from chapter 3, investigating whether Cbp/p300, and subsequent deposition of histone acetylation, are required to maintain an enhancer formed by MYB binding. When I deplete Cbp/p300, reducing H3K27ac by nChIP-seq, I observed a moderate reduction in transcription from the activated regions. One interpretation of this is that there are other proteins involved in maintaining an enhancer formed by MYB binding, besides Cbp/p300. However, the high levels of residual Cbp/p300 and H3K27ac at the *TetO* array after PROTAC treatment prevent drawing a strong conclusion here.

Use of E3 ubiquitin ligase mediated degradation

We made use of the dTAG and PROTAC systems for inducible degradation of Cbp and p300. These methods direct the proteins to the cereblon E3 ubiquitin ligase for

proteasomal degradation and have been shown to be rapid and highly potent (Nabet *et al.*, 2018; Vannam *et al.*, 2021). In comparison, inhibition of CBP and P300 activity through the use of therapeutic inhibitors targeting the acetyltransferase or bromodomain *in vivo* often causes a less pronounced reduction in H3K27ac compared to protein degradation (Vannam *et al.*, 2021). Therefore, I employed an inducible system for the controlled depletion of Cbp and p300 to study the early transcriptional effects of H3K27ac loss with precision. Using the dTAG system, I engineered a Cbp degron cell line, which rapidly degraded Cbp after 1 hour of dTAG addition. However, I observed a consequential increase in p300, leading to steady state H3K27ac. I attempted to engineer a p300 degron cell line, to compare the loss of Cbp alone to the loss of p300 alone. However, this was not possible within the time frame of this thesis but is currently in progress.

I employed a PROTAC, dCBP-1, to degrade both Cbp and p300 in order to cause a more potent loss of histone acetylation. However, it is still evident that we have not achieved substantial loss of H3K27ac by nChIP-seq, as compared to the loss observed in multiple myeloma cells (Vannam *et al.*, 2021). Instead, I observe a modest reduction in H3K27ac by nChIP-seq at the *Sox2* enhancer and promoter in our ES cell model, in combination with a modest reduction in transcription by TT-seq. Therefore, we may have not reached the threshold of histone acetylation loss for dramatic transcriptional changes. On a global scale, we observed distinct gene expression changes, rather than global downregulation which we might have expected due to the strong correlation between histone acetylation and transcription. This may be because different enhancers have different dependencies on co-factors (Neumayr *et al.*, 2022) and those with a greater local concentration and diversity of coactivators are less reliant on CBP/P300 to maintain steady state gene expression (Kasper *et al.*, 2010).

Acetylation from other KAT enzymes

It is also very possible that the modest reduction in H3K27ac and subsequently transcription is due to functional redundancy with other KAT enzymes. Although CBP/P300 have been shown to be the major KAT enzymes for H3K27ac (Jin *et al.*, 2011) it is highly likely that when these enzymes are depleted, other KAT enzymes including those from the GNAT and MYST family, may act in a functionally redundant manner, taking the roles of CBP and P300 to sustain a level of H3K27ac (Martire *et al.*, 2020). Furthermore, KATs from the GNAT and MYST family will not be depleted by the PROTAC and therefore will continue to acetylate their cognate lysine residues, maintaining a diverse repertoire of active chromatin marks. There may also be a global redistribution of KAT enzymes upon the loss of CBP/P300. For example, in cells treated with the CBP/P300 bromodomain inhibitor CCS1477, CBP/P300 is itself redistributed from MYB-bound enhancers to RFX-enhancers in AML cells and from IRF4-bound enhancers to E2A-bound enhancers in myeloma (Nicosia *et al.*, 2023).

The role of CBP/P300 in chromatin structure

An area of future work includes carrying out ATAC-seq after PROTAC-mediated depletion of Cbp/p300, which was not possible in this thesis due to time constraints. In chapter 3, I showed that loss of binding of TetR-MYB^{TA} after doxycycline treatment caused a loss of H3K27ac with no change in chromatin accessibility. In agreement, other studies show that there is often very subtle or no change in chromatin accessibility after depletion of H3K27ac through CBP/P300 inhibitor treatment. Multiple myeloma cells treated with A-485, a CBP/P300 inhibitor targeting the acetyltransferase domain, saw a pronounced decrease in transcription of cell-type specific gene sets but no change in chromatin accessibility by ATAC-seq (Hogg *et al.*, 2021). Treatment of MOLM-16 cells with GNE-049 had a similar effect as with the KAT domain inhibitor; a reduction in enhancer-mediated oncogenic expression but no alteration in the chromatin accessibility by ATAC-seq (Raisner *et al.*, 2018). This is in agreement with

studies in ES cells; depleting H3K27ac leads to little to no change in chromatin accessibility (Martire *et al.*, 2020; Zhang *et al.*, 2020).

This evidence favours a model whereby H3K27ac is involved in sustaining transcription, without directly modifying chromatin structure, at least when considered independently. This may be explained by the myriad of other acetylation marks deposited by other KATs e.g. H3K9ac and H4K16ac (Martire *et al.*, 2020). Together, these marks could be contributing to maintaining a fixed, open chromatin state. Conversely, an example where chromatin accessibility is disrupted upon H3K27ac loss is with PROTAC-mediated degradation of CBP and P300 in multiple myeloma cells. This led to a reduction in chromatin accessibility by ATAC-seq and disruption of oncogenic transcription by RNA-seq. This suggests a threshold of acetylation loss may need to be reached or that loss of the full length protein is necessary to see an effect on chromatin structure (Vannam *et al.*, 2021).

There is also evidence to suggest that hyperacetylation can cause changes in higher order chromatin structure. Entinostat is a KDAC inhibitor that causes an increase in acetylation spreading, leading to an increase in chromatin accessibility and redistribution of 3D chromatin loops at super-enhancers by Hi-ChIP in rhabdomyosarcoma cells (Gryder, Pomella, *et al.*, 2019; Gryder, Wu, *et al.*, 2019). This suggests that acetylation does have a role in higher order chromatin structure, but it may be hyperacetylation rather than hypoacetylation that causes a detectable change.

Furthermore, an exciting avenue to explore would be the change in 3D chromatin structure following PROTAC-mediated depletion of H3K27ac using the 3C technique Micro Capture-C. Recent work has revealed CBP/P300 catalytic inhibition with A-485 leads to a reduction in transcription concurrent with a decrease in frequency and spreading of enhancer-promoter interaction by Hi-C in lymphoma cells, albeit at low

resolution (Sungalee *et al.*, 2021). Conversely, treatment of MOLM-16 AML cells with the CBP/P300 bromodomain inhibitor GNE-049 caused a reduction in enhancer-mediated oncogenic expression but no alteration of 3D chromatin structure by 3C (Raisner *et al.*, 2018). This suggests that the cell type and the 3C method used leads to different conclusions regarding the involvement of CBP/P300 in maintaining 3D chromatin contacts. Having already observed a loss of transcription at the enhancer and promoter of *Sox2* after PROTAC-mediated H3K27ac loss, this is an ideal system to conduct MCC (Hua *et al.*, 2021) and see what happens to the enhancer-promoter contacts of an ESC-specific gene driven by a super-enhancer with a high resolution 3C method. If there is a loss of enhancer-promoter contacts due to H3K27ac reduction, this could be because H3K27ac is directly involved in mediating these interactions or is indirectly involved via recruiting TFs and bromodomain-containing co-activators. It is possible transcription of *Sox2* may occur independently of physical contact with the *Sox2* promoter, as shown in mESCs using live-cell imaging (Alexander *et al.*, 2019).

Cbp/p300 harbour structural roles in addition their catalytic activity

The depletion of full length Cbp and p300 is likely to cause a multitude of effects beyond the loss of H3K27ac, owing to its role as a transcriptional scaffold and bridge (see section 1.3.4). It has previously been observed that poised developmental enhancers are enriched for p300 with no presence of H3K27ac (Rada-Iglesias *et al.*, 2011; Zentner, Tesar and Scacheri, 2011), suggesting that p300 plays a structural role at a subset of enhancers, potentially those that are primed or latent (see section 3.3). Genetic and PROTAC-mediated depletion of *CBP/P300* simultaneously abrogates its catalytic and scaffolding functions. To investigate the non-catalytic function of CBP/P300, a catalytically dead version of the full length CBP/P300 could be engineered, which retains its interactions with other proteins. Conversely, to investigate the catalytic function of CBP/P300, an attempt was made to dissect the role of H3K27ac enhancer activity. ES cells were engineered with a substitution of H3.3 lysine

to arginine, so that H3.3 could no longer be acetylated. This caused a loss of H3K27ac at enhancers but no changes in chromatin accessibility (ATAC-seq) and transcription (RNA-seq) (Zhang *et al.*, 2020). However, H3.3K27R mutations will also deplete H3K27me3, a repressive transcriptional mark, so it is difficult to separate the different effects of eliminating these two modifications. Therefore, there is a need for isolating the acetylation activity in a physiologically relevant model, without altering the chromatin landscape. I attempt this in Chapter 5, by using a truncated version of CBP/P300 only containing core catalytic units (Hilton *et al.*, 2015)

Conclusion

In conclusion, after depleting histone acetylation using PROTAC-mediated depletion of Cbp/p300, I observe a reduction in H3K27ac and transcription at the ES cell specific *Sox2* enhancer and promoter. In addition, I see a modest effect at the *TetO* array, where MYB^{TA} activity is high. I hypothesise that these loci have high levels of transcription factors and co-activators meaning they are less sensitive to histone acetylation loss. However, on a global scale, Cbp/p300 depletion causes a genome-wide loss of H3K27ac and extensive transcriptional dysregulation. This not only enhances our understanding of the function of these important epigenetic regulators but also provides novel insights into targeting CBP and P300 for future drug development, helping us to better understand relapse and resistance in the context of cancer therapy.

Chapter 5: P300-mediated acetylation is sufficient to establish an enhancer-like element, distinct from MYB

5.1 Introduction

Projects such as ENCODE (Dunham *et al.*, 2012) and Roadmap Epigenomics Project (Bernstein *et al.*, 2010) have established a strong correlation between H3K27ac and enhancers and promoters. However, the complete functional role of this epigenetic mark in enhancer activity and transcription remains to be elucidated. Previous evidence, including the results presented in chapter 4, shows that removing P300 and CBP decreases transcription globally (Raisner *et al.*, 2018; Hogg *et al.*, 2021; Pelham-Webb *et al.*, 2021; Vannam *et al.*, 2021). This suggests that P300 and CBP may be necessary to maximise enhancer activity. In this chapter, I test the sufficiency of these epigenetic enzymes to generate enhancer activity, by directly targeting P300 to neutral genomic loci, instead of removing it from existing enhancers. Furthermore, I take a reductionist approach, by isolating the acetyltransferase domain of P300, with the aim of directly testing the role of the acetyltransferase activity, independently of the other domains of P300.

Recent research testing the sufficiency of histone acetylation in gene activation has benefitted from the advancement of CRISPR-Cas9 technology, enabling the development of epigenome editing. In this technique, a nuclease deactivated version of the Cas9 protein (dCas9) is fused to a specific protein or domain of interest. The dCas9 retains its ability to be directed to a predetermined genomic region through co-transfection with a guide RNA. However, crucially, it no longer has the ability to cut DNA, thereby preventing any genetic modifications. For example, the P300 core

domain, which contains the bromodomain and the KAT domain (Figure 5.1), has been fused to dCas9 and targeted to promoters and enhancers to investigate whether the presence of acetylation influences transcription. Targeting dCas9-P300^{Core} to the enhancer cluster (HS2) of the beta globin locus caused a significant increase in transcription from the beta globin genes (*HBE, HBG and HBD*) by RT-qPCR in HEK293T cells where these genes are not usually expressed (Hilton *et al.*, 2015). This work established a causal relationship between acetylation at regulatory elements and transcription from the target gene. This approach allows for proteins to be investigated in their native context, ensuring physiological relevance. However, this is also a limitation because active transcription processes and pre-existing regulatory elements are already operating at these genomic loci, complicating the determination of whether P300 in isolation is responsible for any observed transcriptional effects. To improve upon this approach, and fully elucidate the function of H3K27ac in isolation from its pre-existing enhancer elements, for this chapter, I again employed the TetR system introduced in Chapter 3, to target P300 to a neutral chromatin region.

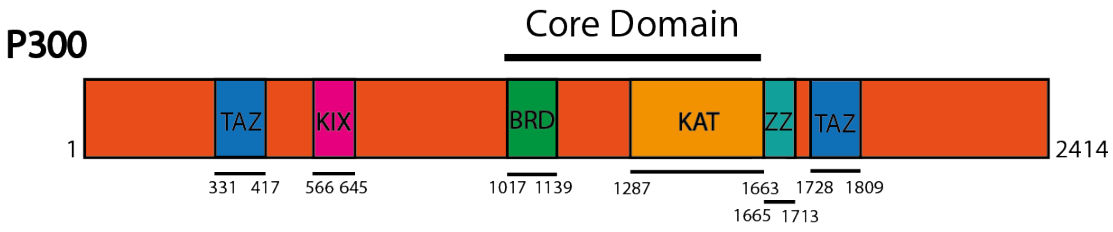


Figure 5.1 Schematic representation of the P300 protein

In mammalian cells, P300 has 6 main domains: 2 TAZ and 1 ZZ zinc finger domains (blue and teal respectively), KIX domain (pink), bromodomain (BRD) (green) CH2 domain (contains PHD domain and RING finger domain, not shown) and lysine acetyltransferase domain (KAT) (yellow), which includes an autoinhibitory loop. The core domain contains the BRD domain and the KAT domain (Hilton, 2015). Total protein size 2414 amino acids. Amino acid residues are shown for the human P300, information obtained from UniProt, 2023.

The findings presented in Chapter 3 demonstrated that TetR-MYB^{TA} has the ability to establish an enhancer-like element at one genomic locus (chr8-*TetO*), but apparently not at another (chrX-*TetO*). I hypothesise that this is due to the recruitment of CBP/P300 and subsequent deposition of histone acetylation. In this chapter, I aimed to build upon this observation and assess whether the deposition of acetylation by the P300 core domain alone, could independently create an enhancer-like element at the same genomic locus. Since I am excluding other domains of P300, including the KIX domain, this will also simultaneously test the necessity of the additional domains with scaffolding and bridging roles for enhancer function.

The P300 core domain has been effectively used in epigenome editing, displaying high levels of H3K27ac in a targeted manner (Hilton *et al.*, 2015). The crystal structure of the core domain has been determined (Delvecchio *et al.*, 2013) which demonstrates that this region forms a stably folded structure. Prompted by this, I fused the P300 core domain (using the same sequence as Hilton *et al*) to the TetR protein to investigate its potential to activate transcription in a neutral region. Additionally, a specific point mutation (D1399Y) within the KAT domain has been identified which completely abrogates the acetyltransferase activity of the P300 core domain (Hilton *et al.*, 2015), which I used as a negative control in our system.

5.1.1 Aims

I set out to investigate whether the P300-mediated acetylation plays a role in enhancer function. To do this I had three main aims:

1. Create a novel TetR-P300^{Core} model system for investigating enhancer function.
2. Validate TetR-P300^{Core} binding and functionality at the *TetO* locus.
3. Investigate if acetylation by P300 is sufficient to establish an enhancer-like element at the *TetO* array.

5.2 Results

5.2.1 Transient Transfection of TetR-P300^{Core}

Plasmids were designed and engineered in collaboration with Philip Hublitz of Genome Engineering at the WIMM. The TetR-P300^{Core} D1399Y contains a point mutation in the core domain which abolishes its acetyltransferase activity (Hilton *et al.*, 2015).

To initially validate the functionality of these plasmids, I transiently transfected TetR, TetR-MYB^{TA}, TetR-P300^{Core} wild type (WT) and TetR-P300^{Core} D1399Y into the chrX-*TetO* engineered mESCs (see Chapter 3) for 24hr and then fixed chromatin for ChIP-qPCR. This was to confirm successful binding of the TetR-P300^{Core} at *TetO* and deposition of histone acetylation, which would indicate expression of a functional lysine acetyltransferase domain.

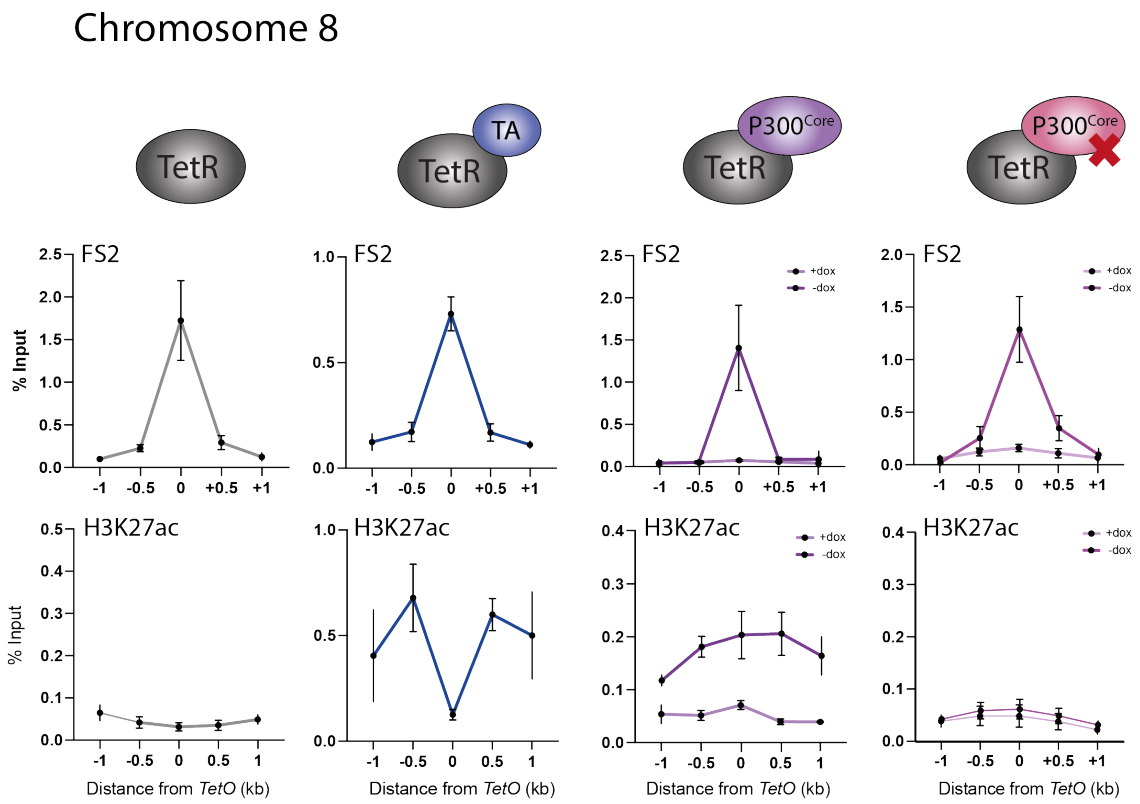


Figure 5.2 ChIP-qPCR using FS2 (TetR) and H3K27ac antibodies at chr8-*TetO*

Transient transfections (24h) of constructs expressing the indicated TetR fusion proteins in the E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO*. +dox = 3 hours of doxycycline treatment.

Top panel: FS2 (TetR) ChIP-qPCR (n=3) for -0.5, 0 and +0.5kb, error bars indicate the SEM, (n=2) for -1 and +1, error bars indicate the range.

Bottom panel: H3K27ac ChIP-qPCR (n=3) for -0.5, 0 and +0.5kb, error bars indicate the SEM, (n=2) for -1 and +1, error bars indicate the range.

% Input was calculated using the DDCt (ChIP versus input controls) method. X axis is comprised of primer pairs at the *TetO* array and +/- 0.5kb and +/- 1Kb from the chr8-*TetO* array.

To detect TetR binding, I performed ChIP-qPCR using the FS2 antibody. The results demonstrate a clear and specific binding of both TetR-P300^{Core} WT and D1399Y fusion proteins to the *TetO* array, at chromosome 8. Binding is diminished with the addition of doxycycline for 3h. This level of binding is comparable to TetR and TetR-MYB^{TA} (Figure 5.2).

We used H3K27ac as a proxy for P300-mediated acetylation, as this antibody is potent and has previously been validated. TetR-P300^{Core} WT showed enrichment for H3K27ac across the *TetO* region, which notably decreased after 3 hours of doxycycline treatment (Figure 5.2). Conversely, the TetR-P300^{Core} D1399Y displayed relatively low levels of H3K27ac in both the + and – doxycycline conditions, consistent with the inactivating mutation in this protein. TetR-MYB^{TA} serves as a positive control, as we have previously observed H3K27ac at the *TetO* array in response to TetR-MYB^{TA}, due to recruitment of CBP/P300 (Refer to Chapter 3). In contrast to the TetR-P300^{Core} constructs, TetR-MYB^{TA} displayed a noticeably stronger signal accompanied by a characteristic dip in signal at the *TetO* array, despite binding more weakly at the *TetO* array compared to TetR-P300^{Core}. This suggests that the recruitment of full length P300 by MYB results in more robust H3K27ac deposition compared to the P300^{Core} alone. However, despite this distinction, the observed phenotype was sufficient for further investigation.

These findings were replicated at the second *TetO* array on chromosome X (Figure 5.3). H3K27ac was observed with TetR-P300^{Core} WT fusion protein and its presence was not detected with the addition of doxycycline or with the TetR-P300^{Core} D1399Y fusion protein. As at chr8-*TetO*, no dip in H3K27ac was detected at the *TetO* array.

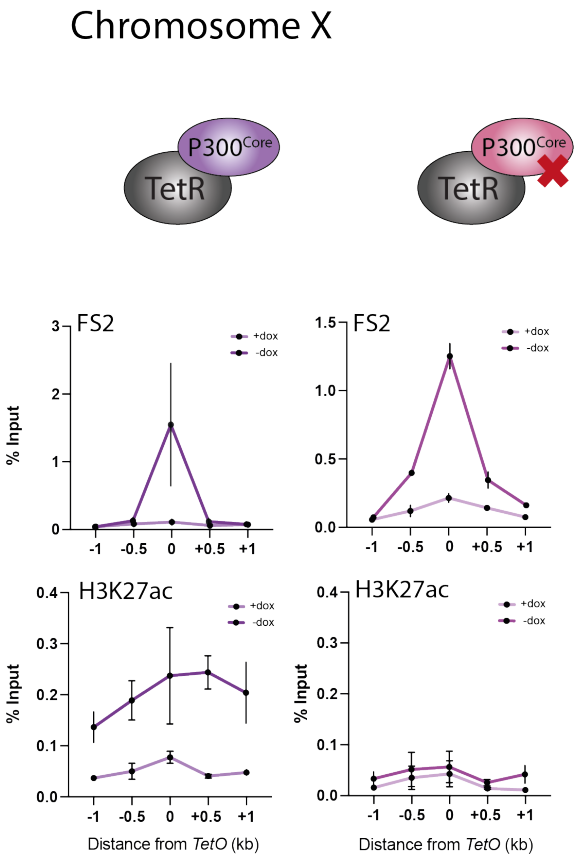


Figure 5.3 ChIP-qPCR using FS2 (TetR) and H3K27ac antibodies at ChrX-*TetO*

Transient transfections (24h) of constructs expressing the indicated TetR fusion proteins in the E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO*.

Top panel: FS2 (TetR) ChIP (n=2) error bars indicate the range.

Bottom panel: H3K27ac ChIP (n=3) for -0.5, 0 and +0.5, error bars indicate the SEM, (n=2) -1 and +1, error bars indicate the range.

% Input was calculated using the DDCt (ChIP versus input controls) method. X axis is comprised of primer pairs at the *TetO* array and +/- 0.5 and +/- 1 kb from the *TetO* array.

Subsequently, I conducted RT-qPCR to assess potential changes in transcription associated with P300^{Core}-mediated acetylation at the *TetO* array. Utilising the same primer pairs as for the ChIP-qPCR experiment, I extracted RNA and conducted cDNA synthesis using random hexamer primers to detect nascent transcription. qPCR revealed a modest increase in transcription across the chr8-*TetO* region when the TetR-P300^{Core} WT was introduced, as compared to TetR alone and TetR-P300^{Core} D1399Y (Figure 5.4). Transcription appeared to be dependent on the continued binding of TetR-P300^{Core} WT, as levels were reduced in the presence of doxycycline. Building upon this observation, I aimed to generate stable cell lines expressing either TetR-P300^{Core} WT and D1399Y separately.

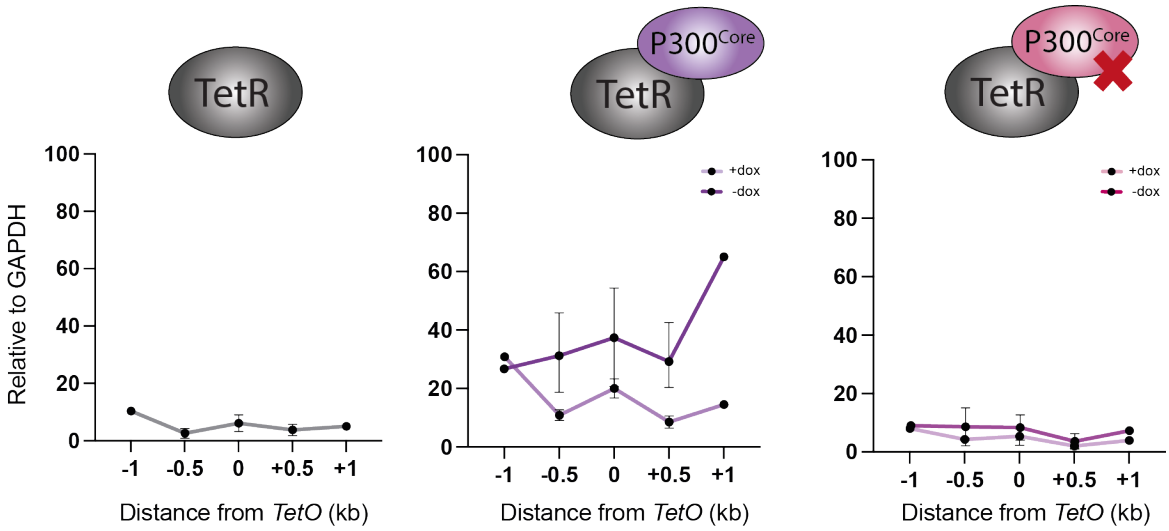


Figure 5.4 RT-qPCR at chr8-*TetO*

RT-qPCR (n=3) for -0.5, 0 and +0.5, error bars indicate the SEM, (n=2) for -1 and +1, error bars indicate the range. X axis is comprised of primer pairs at the *TetO* array and +/- 0.5 and +/- 1 kb from the chr8-*TetO* array. Transient transfections (24h) of constructs expressing the indicated TetR fusion proteins in the E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO*. +/- dox = 3 hours of doxycycline treatment.

Signal was normalised to DDCt (sample versus *Gapdh* Ct) and scaled (arbitrary units).

5.2.2 Making stable cell lines expressing TetR-P300^{Core}

I aimed to express the TetR-P300^{Core} WT and TetR-P300^{Core} D1399Y in two distinct cell lines, harbouring the chr8-*TetO* and chrX-*TetO*, by random integration and puromycin selection. However, despite multiple attempts, there were very few colonies that grew in puromycin-supplemented media post-transfection. This observation suggested that the integration efficiency of the TetR-P300^{Core} plasmid was sub-optimal, or that the stable expression of the TetR-P300^{Core} may be toxic to the cells, resulting in cell death.

One attempt led to the isolation of 6 clones for TetR-P300^{Core} WT and 6 clones for TetR-P300^{Core} D1399Y. I extracted protein and screened by western blot using the FS2 antibody, which detects the tag at the N-terminus of the fusion protein. However, it was evident that only TetR-P300^{Core} D1399Y clones expressed the fusion protein, whilst there was no detection of a TetR fusion protein of the expected size (105kDa) in the suspected TetR-P300^{Core} WT clones (Figure 5.5).

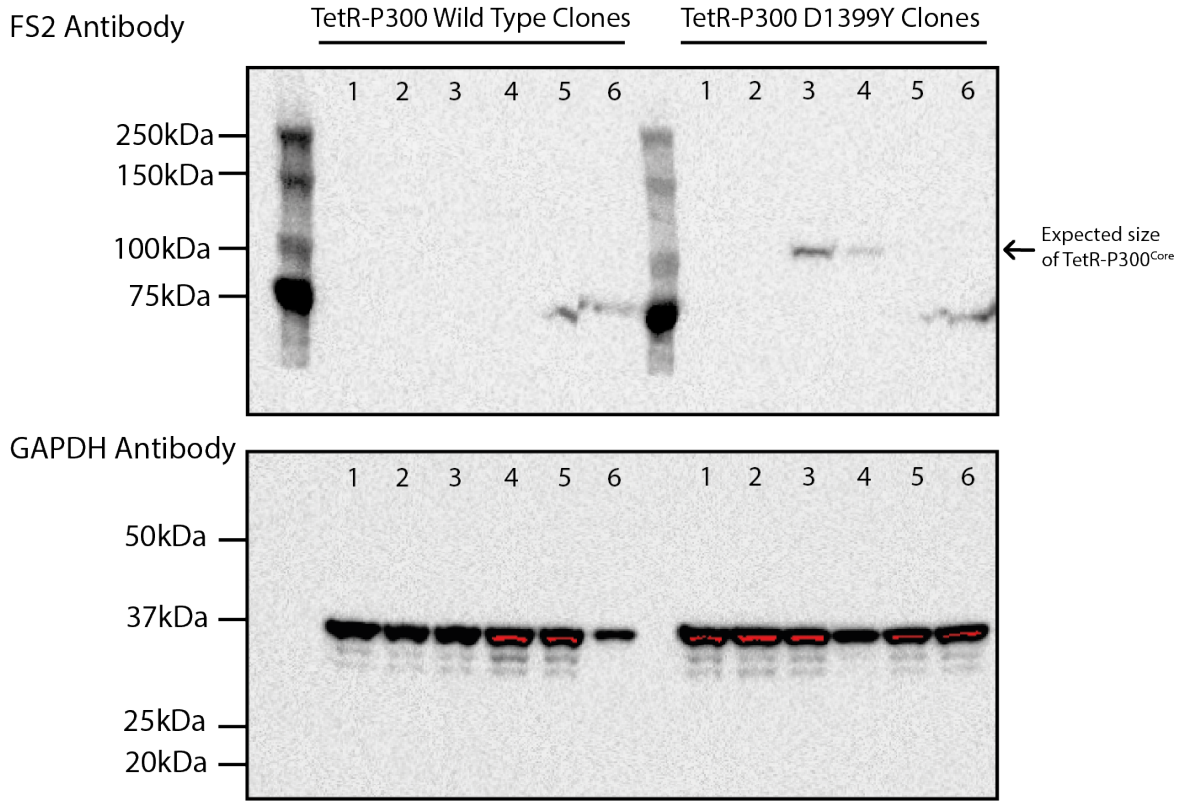


Figure 5.5 Western Blot using FS2 (TetR) antibody to screen for cells stably expressing TetR-P300^{Core}

Western blot using protein extracted from puromycin-resistant clones of the E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO*, transfected with TetR-P300^{Core} WT or TetR-P300^{Core} D1399Y. Top panel: six puromycin-resistant clones each for TetR-P300^{Core} WT and TetR-P300^{Core} D1399Y. The expected size of TetR-P300^{Core} is 105kDa. Precision plus ladder is displayed. Bottom panel: GAPDH as a loading control.

I hypothesised that overexpression of P300^{Core} may be toxic due to excessively high levels of histone acetylation perturbing chromatin structure or gene expression. To address this possibility, the construct was further engineered by PhilIp Hublitz of Genome Engineering, WIMM, to include a mutant oestrogen receptor (ER^{T2}) ligand-binding domain fused to the TetR-P300^{Core}. This ER^{T2} domain retains the TetR-P300^{Core} in the cytoplasm until the addition of tamoxifen, at which point the construct is released and translocates to the nucleus. Unfortunately, even with this modified strategy, I was still unable to generate clones that stably expressed TetR-P300^{Core} WT. This challenge may result from potential system leakiness, resulting in TetR-P300^{Core} prematurely migrating to the nucleus in the absence of tamoxifen, thereby inducing similar toxic effects.

5.2.3 Transient transfection with sorting strategy

Due to the challenges in generating a cell line stably expressing the TetR-P300^{Core} WT fusion protein, I opted to continue with a transient transfection approach. One disadvantage of this approach is that not all cells within a transfection experiment will take up the plasmid and express the fusion protein. Therefore, I modified this transfection strategy by implementing a sorting step to enrich for cells expressing the TetR-P300^{Core}, thereby augmenting the signal for subsequent next generation sequencing experiments.

To allow sorting of TetR-P300^{Core} expressing cells, Philip Hublitz of Genome Engineering, WIMM, engineered two new constructs, replacing the puromycin resistance marker with eGFP. (Figure 5.6). Therefore, I was able to sort the cells for eGFP-positivity i.e. those expressing TetR-P300^{Core}.

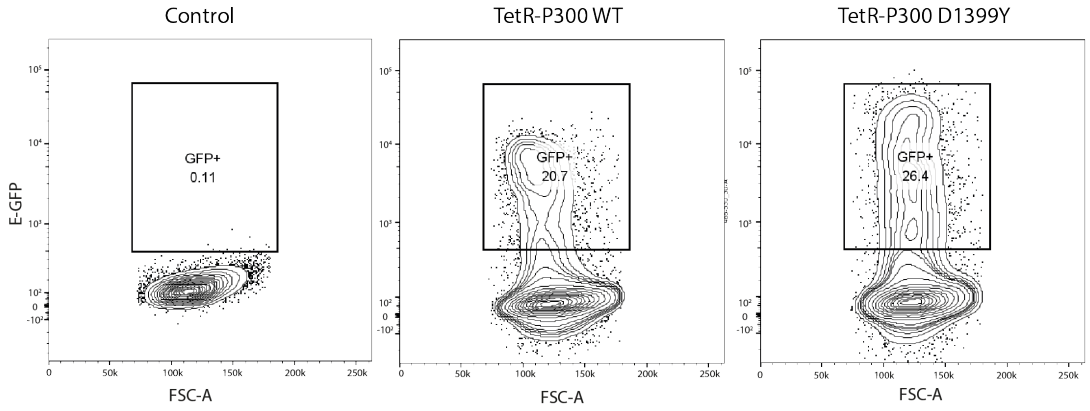


Figure 5.6 Representative gating strategy for fluorescence activated cell sorting (FACS) after transfection with TetR-P300^{Core} containing GFP

Control cells are wild type E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO,* this is used to gate for untransfected cells (GFP+ = 0.11%).

Cells transfected with TetR-P300 WT have a population of 20.7% GFP+ transfected cells.

Cells transfected with TetR-P300 D1399Y have a population of 26.4% GFP+ transfected cells.

Analysis conducted on FlowJo.

The TetR-P300^{Core} WT and TetR-P300^{Core} D1399Y plasmids were independently transfected into E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO*. Following a 24-hour incubation, cells were sorted based on eGFP expression. Representative FACS results are shown, the transfection efficiency for TetR-P300^{Core} WT was 20.7%, while TetR-P300^{Core} D1399Y reached 26.4% (Figure 5.6).

Untransfected cells originating from the same cell line were used as a control to establish the gating criteria for GFP+ cells.

After transfection, cells were immediately processed for downstream analysis (e.g. fixation for ChIP). Initially I conducted ChIP-qPCR to confirm that the sorting step did not affect the binding profile of the TetR-P300^{Core}, or the pattern of H3K27ac at the *TetO* loci. There was no change in the binding profile of TetR-P300^{Core} after the sorting step, but the signal strength had improved, with a 7-8 fold higher percentage input compared to unsorted transiently transfected cells (compare Figure 5.7 with Figure 5.2 and Figure 5.3).

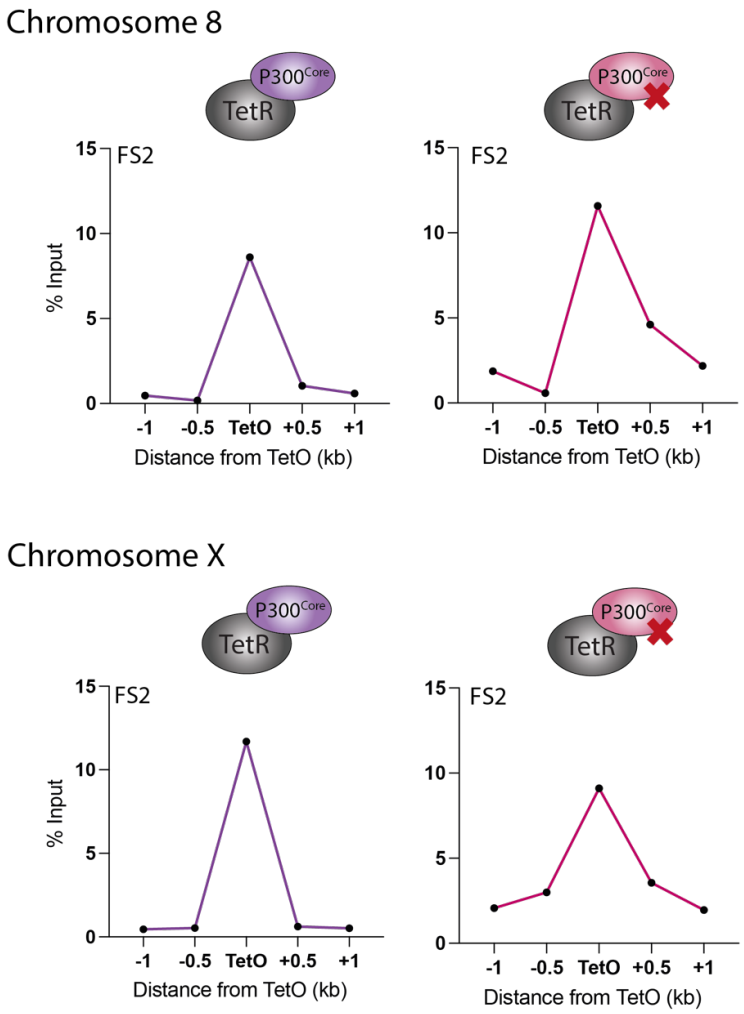


Figure 5.7 ChIP-qPCR using FS2 (TetR) antibodies at Chr8-*TetO* and ChrX-*TetO*

FS2 (TetR) ChIP (n=1) after sorting of transient transfection of the TetR-P300^{Core} WT plasmid in the E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO.*

% Input was calculated using the DDCt (ChIP versus input controls) method. X axis is comprised of primer pairs at the *TetO* array and +/- 0.5 and +/- 1 kb from the *TetO* array.

Remarkably, there was approximately a 25-35 fold increase in H3K27ac signal in the TetR-P300^{Core} WT compared to unsorted cells (compare Figure 5.8 with Figure 5.2 and Figure 5.3) and the classical dip in signal at the *TetO* array was now notable, as seen in the TetR-MYB^{TA} condition (Figure 5.8). There was also a 2-3 fold higher signal at the

TetO loci compared to the positive control, *Gapdh*, as observed for the TetR-MYB^{TA} construct (Figure 5.8).

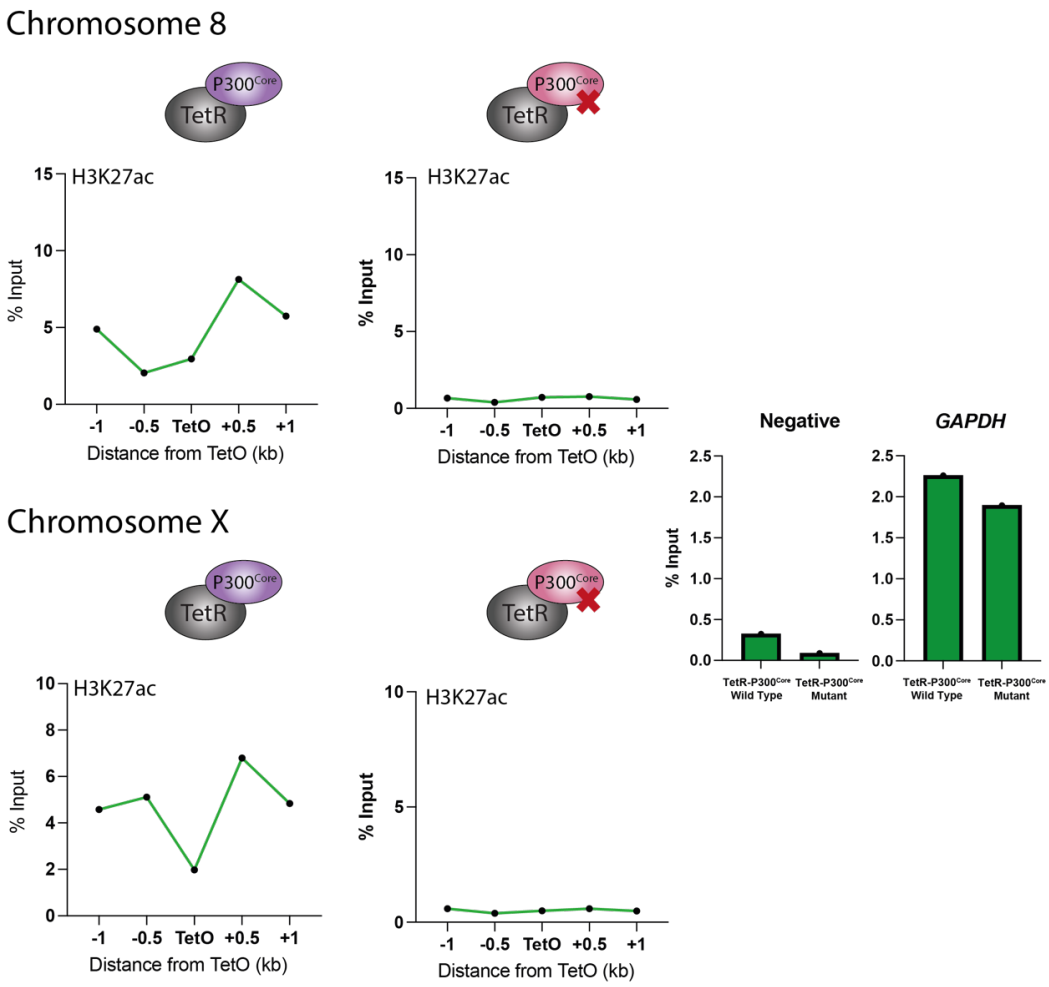


Figure 5.8 ChIP-qPCR using H3K27ac antibody at Chr8-*TetO* and ChrX-*TetO*

Transient transfections in the E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO*

H3K27ac ChIP (n=1) after sorting at chr8-*TetO* and chrX-*TetO.* qPCR at endogenous mouse loci is shown for comparison.

% Input was calculated using the DDCt (ChIP versus input controls) method. X axis is comprised of primer pairs at the *TetO* array and +/- 0.5 and +/- 1 kb from the *TetO* array.

5.2.4 TetR-P300^{Core} WT causes global hyperacetylation at chromatin.

The fact that I was not able generate TetR-P300^{Core} WT stably expressing cell lines suggested that overexpressing the P300 core domain may be toxic, and I hypothesised that this might be caused by hyperacetylation of chromatin or other proteins, resulting in cell death. To investigate whether TetR-P300^{Core} expression affected the global acetylation levels, I performed western blot after sorting TetR-P300^{Core} WT and D1399Y plasmids (Figure 5.9A). I also analysed cells stably expressing TetR-MYB^{TA} as a reference. These experiments revealed an elevated level of H3K27ac in cells transfected with TetR-P300^{Core} WT in comparison to cells expressing TetR-MYB^{TA} (Figure 5.9A). Conversely, cells transfected with TetR-P300^{Core} D1399Y displayed reduced levels of H3K27ac in comparison to the TetR-MYB^{TA}. This observation strengthens the conclusion that the TetR-P300^{Core} WT construct causes detrimental effects to the cell owing to hyperacetylation.

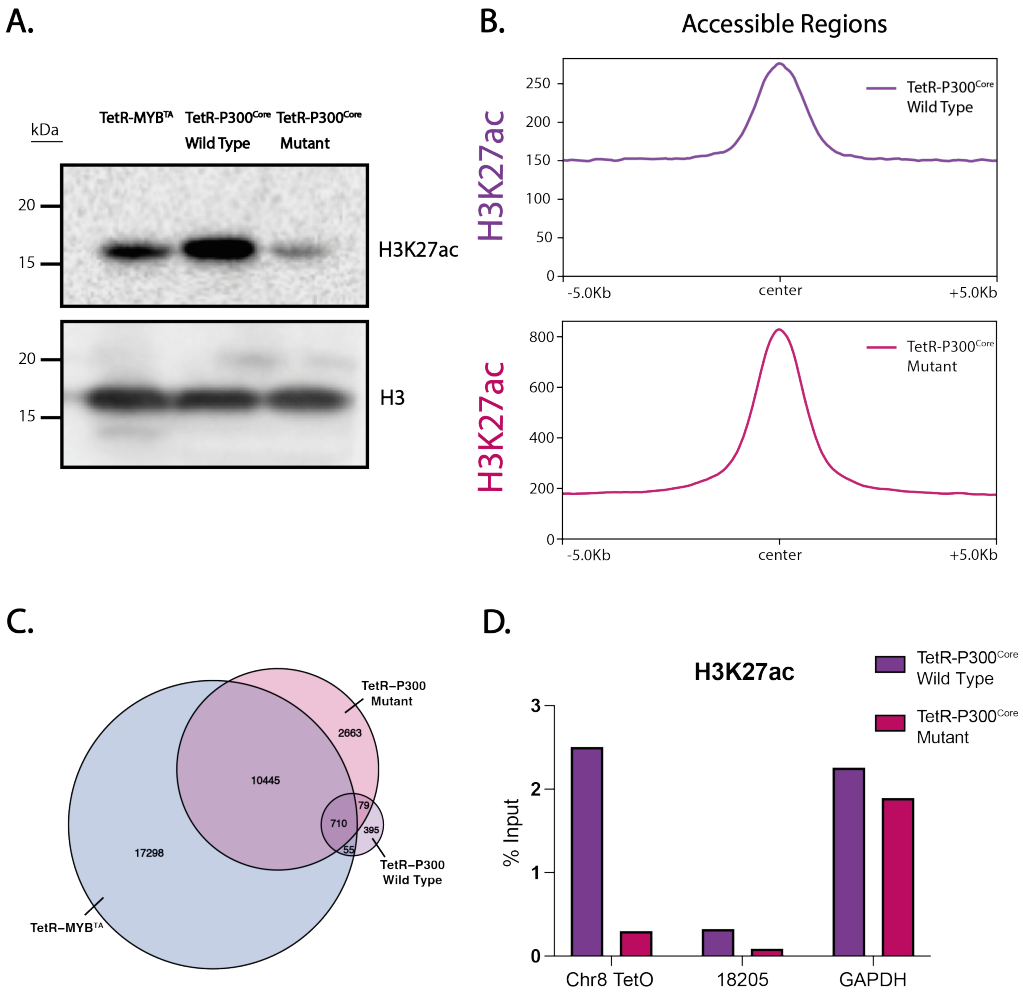


Figure 5.9: An increase in global H3K27ac in TetR-P300^{Core} WT

A | Western Blot using protein extracted from cells stably expressing TetR-MYB^{TA} or TetR-P300^{Core} WT and D1399Y 24hr transient transfections with sorting in in the E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO*. Top panel: H3K27ac and bottom panel H3 as a loading control. Precision plus ladder was used as a molecular weight marker.

B | Metaplots from H3K27ac ChIP-seq (n=1), centered on ATAC-seq peaks (+/- 5Kb) from the E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO.*

C | Venn diagram representing the overlap of macs2 H3K27ac ChIP-seq peak calls from macs2 in TetR-MYB^{TA} stably expressed cell line, TetR-P300^{Core} WT and D1399Y transient transfections with sorting.

D | H3K27ac ChIP-qPCR (n=1) from the same transfection as B/C, TetR-P300^{Core} WT and D1399Y transient transfections with sorting. % Input was calculated using the DDCt (ChIP versus input controls) method.

In an independent experiment, I conducted H3K27ac ChIP-seq to see if the higher levels of H3K27ac for TetR-P300^{Core} WT observed in the western blot impacted acetylation levels at specific chromatin loci. Surprisingly, considering the elevated acetylation levels, I observed a drastically reduced number of H3K27ac peaks in the TetR-P300^{Core} WT (1239) compared to the TetR-P300^{Core} D1399Y (13897) and the TetR-MYB^{TA} (27743) (Figure 5.9B). Further investigation revealed that the background signal was higher relative to the H3K27ac peak in the TetR-P300^{Core} WT, in comparison to the TetR-P300^{Core} D1399Y where the background signal was lower relative to the H3K27ac peak (Figure 5.9C), meaning that the relative enrichment of signal at H3K27ac peaks was reduced in the TetR-P300^{Core} WT. This indicates that many of the H3K27ac peaks may have been missed by peak callers because of a lack of statistical significance. So even though it looks as though there are fewer H3K27ac peaks in the TetR-P300^{Core} WT condition, this may be an artefact of increased global levels of acetylation, rather than reduced acetylation at H3K27ac peaks.

This finding was corroborated by ChIP-qPCR from the same experiment. The chr8-*TetO* array showed a strong and specific increase in H3K27ac in the WT, whereas levels of acetylation at *Gapdh* were comparable in the WT and D1399Y (Figure 5.9D). Therefore, the global reduction in H3K27ac signal in the TetR-P300^{Core} WT condition observed by ChIP-seq is likely due to a normalisation issue with ChIP-seq owing to increased background H3K27ac.

5.2.5. P300-mediated acetylation activates distinct regions from the MYB transactivation domain.

I transiently transfected cells with TetR, TetR-MYB^{TA}, TetR-P300^{Core} WT and TetR-P300^{Core} D1399Y and carried out FS2 (TetR) ChIP-seq to confirm the binding profile of the fusion proteins. This was to ensure specific binding of the TetR-P300^{Core} to the *TetO* array given that the core domain contains the bromodomain which potentially could cause genome-wide binding to acetylated lysine residues. This was possible without sorting as the antibody binds with high affinity and specificity, meaning there is very low background, to show sufficient signal in a mixed population. I saw a sharp peak of TetR binding at the *TetO* array in all four conditions (Figure 5.10).

Next, I carried out next generation sequencing applications in the TetR-P300^{Core} wild type and mutant conditions after sorting, to investigate their individual contributions to enhancer activity at the *TetO* locus. Sequencing traces from stably expressing TetR and TetR-MYB^{TA} cell lines, generated in chapter 3, are shown as a reference.

ATAC-seq in the TetR-P300^{Core} expressing cells showed a similar profile of chromatin accessibility around chr8-*TetO* to TetR and TetR-MYB^{TA} (Figure 5.10). For the H3K27ac ChIP-seq data, it was necessary to scale these tracks to distal peaks independently for each condition, due to the normalisation issue explained earlier. Zooming in on the *TetO* array does reveal that TetR-P300^{Core} WT exhibits higher levels of H3K27ac at the *TetO* array as compared to the TetR-P300^{Core} D1399Y (Figure 5.10). Consistent with ChIP-qPCR (Figure 5.9D), H3K27ac ChIP-seq showed an enrichment of H3K27ac at the *TetO* array in the TetR-P300^{Core} WT compared to the TetR-P300^{Core} D1399Y.

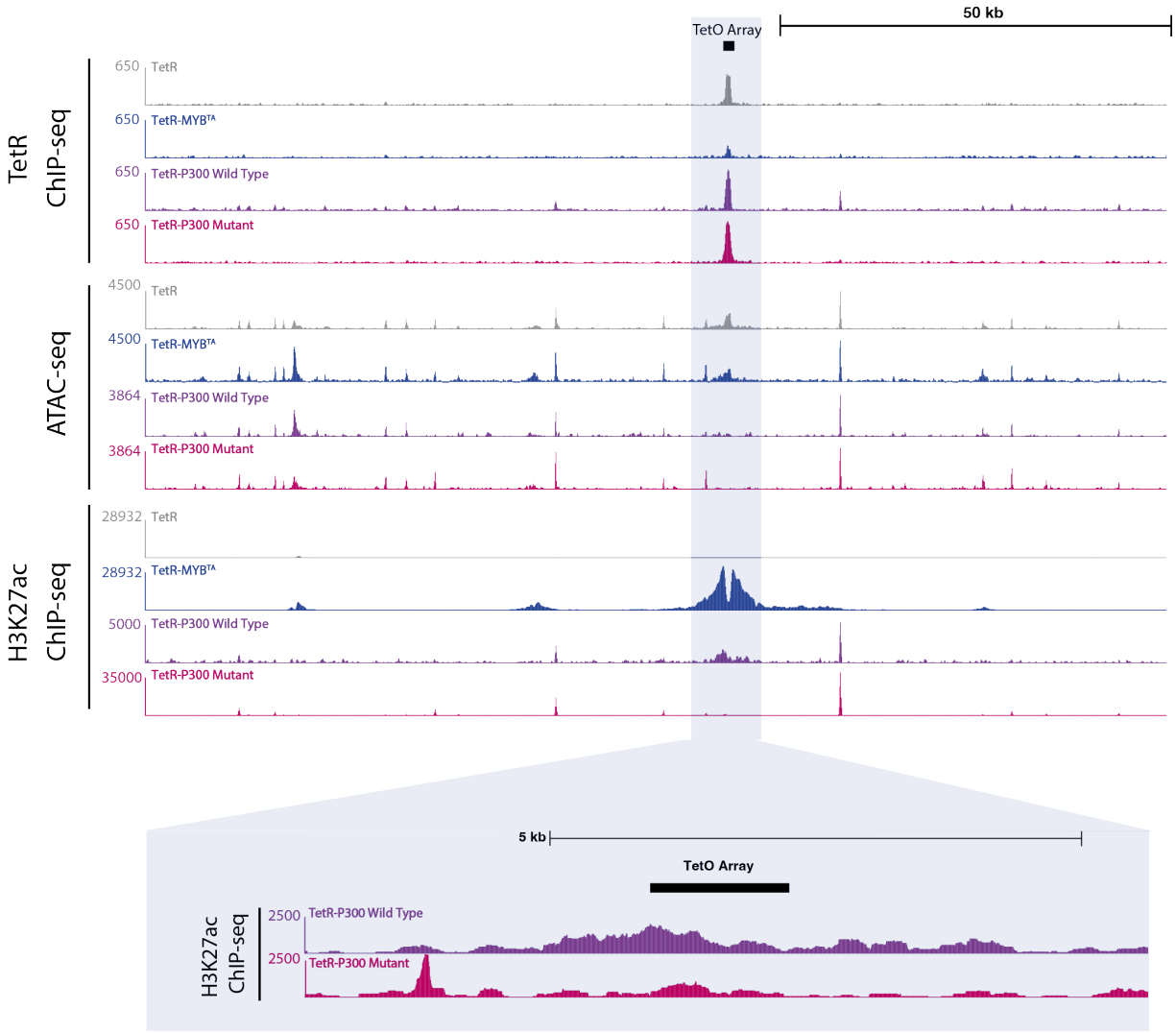


Figure 5.10 Binding of TetR-P300^{Core} WT to the *TetO* array causes an increase in H3K27ac at chr8-*TetO*

Representative tracks from UCSC of ATAC-seq (n=1), ChIP-seq for FS2 (TetR) and H3K27ac (n=1). Region shown is ~100kb surrounding the chr8-*TetO*, 50kb scale and *TetO* array shown. Cells are E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO* transiently transfected with TetR-P300 WT or TetR-P300 D1399Y. Asterisk highlight regions activated by TetR-MYB^{TA} or TetR-P300^{Core} WT.

Zoomed in: Representative tracks from UCSC of H3K27ac ChIP-seq (n=1). Region shown is ~10kb surrounding the chr8-*TetO*, 5kb scale and *TetO* array shown. E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO* transiently transfected with TetR-P300 WT or TetR-P300 D1399Y. Both conditions are group scaled.

PolyA+ RNA-seq enriches for polyadenylated mRNA. PolyA+ RNA-seq did not show signal in either TetR-P300^{Core} wild type or mutant conditions. PolyA- RNA-seq was used to capture non-coding RNAs including, specifically of interest here, enhancer RNAs (eRNAs). PolyA- RNA-seq did not reveal any changes in transcription at the *TetO* array (Figure 5.11A). In contrast, RT-qPCR detected transcription at the *TetO* array in the TetR-P300^{Core} WT, compared to the TetR-P300^{Core} mutant (Figure 5.4). This suggests the PolyA+/- RNA-seq did not capture transcription from the *TetO* array that RT-qPCR could detect, which may become visible with increased sequencing depth. However, despite this limit of detection, at distal regions, PolyA- RNA-seq revealed a modest increase in transcription following transfection with TetR-P300^{Core} WT, which may represent eRNA transcription (Figure 5.11A). This transcription is less pronounced than observed with TetR-MYB^{TA}.

It is difficult to make direct comparisons between TetR-MYB^{TA} (stable cell line) TT-seq and TetR-P300 (transient transfection) PolyA- RNA-seq, since these are different techniques, and the locus may behave differently following selection and stable protein expression compared to a transient transfection. Therefore, to remove these variables from the comparison, I conducted an independent experiment where I transfected cells with plasmids for TetR, TetR-MYB^{TA}, TetR-P300^{Core} WT and TetR-P300^{Core} D1399Y in parallel, and conducted RT-qPCR on unsorted cells, to compare transcription levels more directly between these conditions. I tested three loci where the RNA-seq experiments showed potential activation (highlighted regions in Figure 5.11A). One was specifically activated with TetR-MYB^{TA}, one specifically activated with TetR-P300^{Core}, and one activated in both conditions (Figure 5.11B). This confirmed that MYB^{TA} and P300^{Core} activate transcription at distinct and overlapping loci.

These data recapitulate the findings of the RNA-seq (Figure 5.11A), suggesting that TetR-P300^{Core} WT binding at the *TetO* array can establish an enhancer-like element, however it activates a mix of unique and common regions compared to the MYB^{TA} domain. This suggests that TetR-MYB^{TA} does not solely act through recruitment of P300 and histone acetylation deposition. Instead, other factors may be involved in the enhancer-forming capabilities of MYB, aligning with results presented in Chapter 4 where I showed degrading Cbp/p300 reduces MYB^{TA}-driven enhancer activity but does not completely abrogate it. It is notable that the P300^{Core} domain activates expression at a locus that is unaffected by TetR-MYB^{TA}, suggesting that the full length P300 may lack some activity of the Core domain alone, and/or is modulated by the presence of MYB or its other binding partners. It is possible that if TetR-P300^{Core} WT was stably expressed in this system, we might observe similar levels of transcription to that observed with TetR-MYB^{TA} and that the limit of detection has not been reached in the PolyA+/- RNA-seq experiments. However, we cannot draw this conclusion without a stable cell line expressing TetR-P300^{Core} WT.

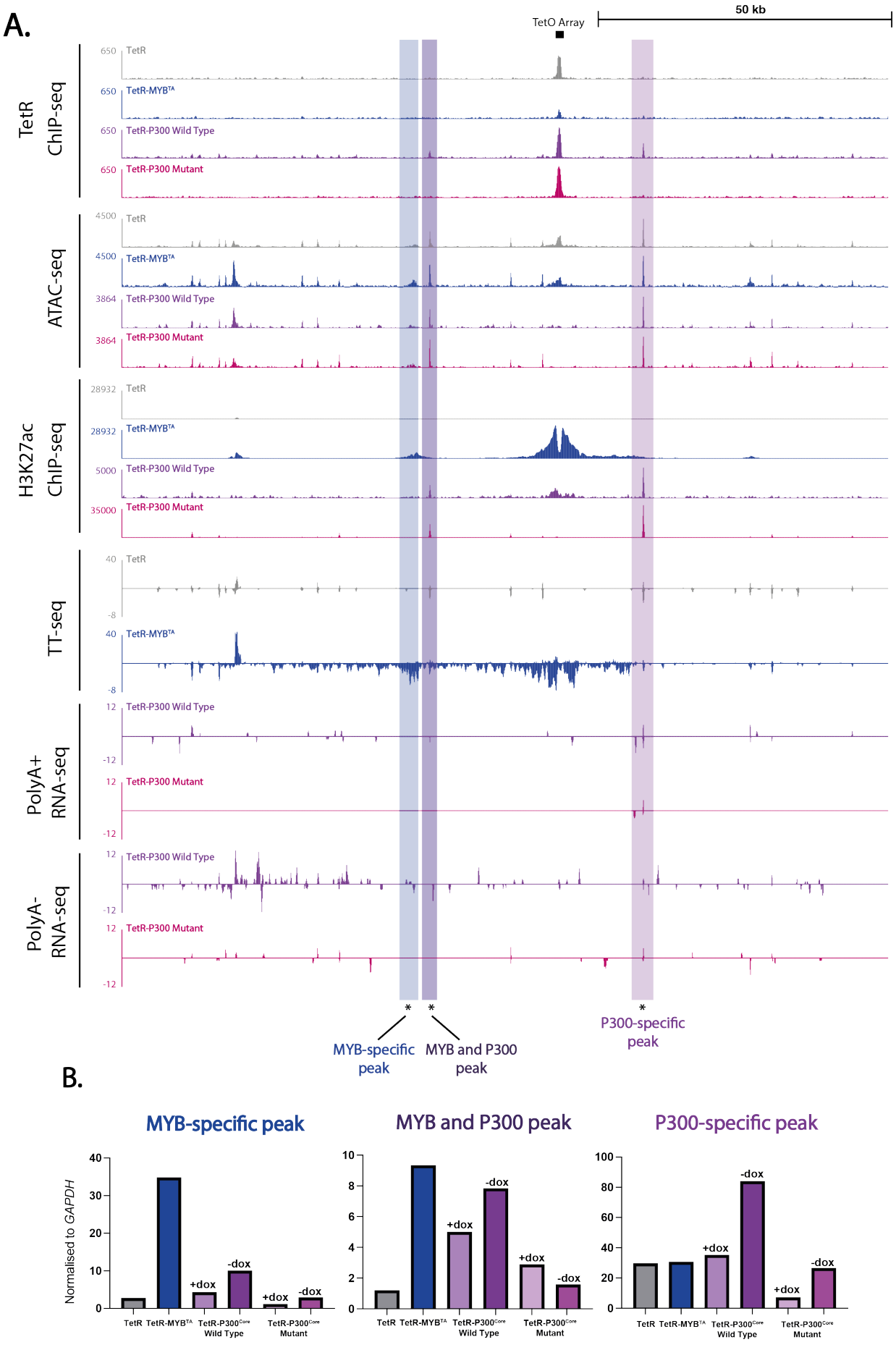

A.
TetO Array
50 kb
TetR ChIP-seq
ATAC-seq
H3K27ac ChIP-seq
TT-seq
PolyA+ RNA-seq
PolyA- RNA-seq
TetR
TetR-MYB^TA
TetR-P300 Wild Type
TetR-P300 Mutant
MYB-specific peak
MYB and P300 peak
P300-specific peak
B.
MYB-specific peak
MYB and P300 peak
P300-specific peak
Normalised to GAPDH
+dox
-dox
TetR
TetR-MYB^TA
TetR-P300^Core Wild Type
TetR-P300^Core Mutant

Figure 5.11 Binding of TetR-P300^{Core} WT to the *TetO* array causes an increase in H3K27ac at chr8-*TetO* and an increase in distal regions, distinct and unique to TetR-MYB^{TA}

A | Representative tracks from UCSC of ATAC-seq (n=1), ChIP-seq for FS2 (TetR) and H3K27ac (n=1), TT-seq and strand-specific PolyA+ and PolyA- RNA-seq (n=1). Region shown is ~100kb surrounding the chr8-*TetO*, 50kb scale and *TetO* array shown. Cells are E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO* transiently transfected with TetR-P300 WT or TetR-P300 D1399Y. Asterisk highlight regions activated by TetR-MYB^{TA} or TetR-P300^{Core} WT.

B | RT-qPCR (n=1). Regions shown correspond to regions highlighted in cells are E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO* transiently transfected for 24hr with TetR, TetR-MYB^{TA}, TetR-P300^{Core} WT and TetR-P300^{Core} D1399Y, then RNA extracted without cell sorting. Signal was normalised to DDCt (sample versus *Gapdh* Ct) and scaled (arbitrary units). +dox = doxycycline treatment for 3 hours.

5.2.6 P300-mediated acetylation is unable to activate chromatin at chrX-*TetO.*

Finally, we analysed the same set of experiments at chrX-*TetO* (Figure 5.12). As shown in chapter 3, binding of TetR-MYB^{TA} at the chrX-*TetO* did not cause activation of distal elements. I observed an increase in H3K27ac at the *TetO* array in the TetR-P300 WT condition, without adjusting for the normalisation artefact (see Section 5.2.4). There is a subtle increase in transcription observed in the TetR-P300 WT condition for polyA-RNA-seq. However, this appears to be initiated from the *TetO* array and does not extend further. This data aligns with results from chapter 3, indicating that neither TetR-MYB^{TA} nor TetR-P300^{Core} can activate distal regions at chrX-*TetO*, suggesting other factors or DNA elements are necessary for activating this locus.

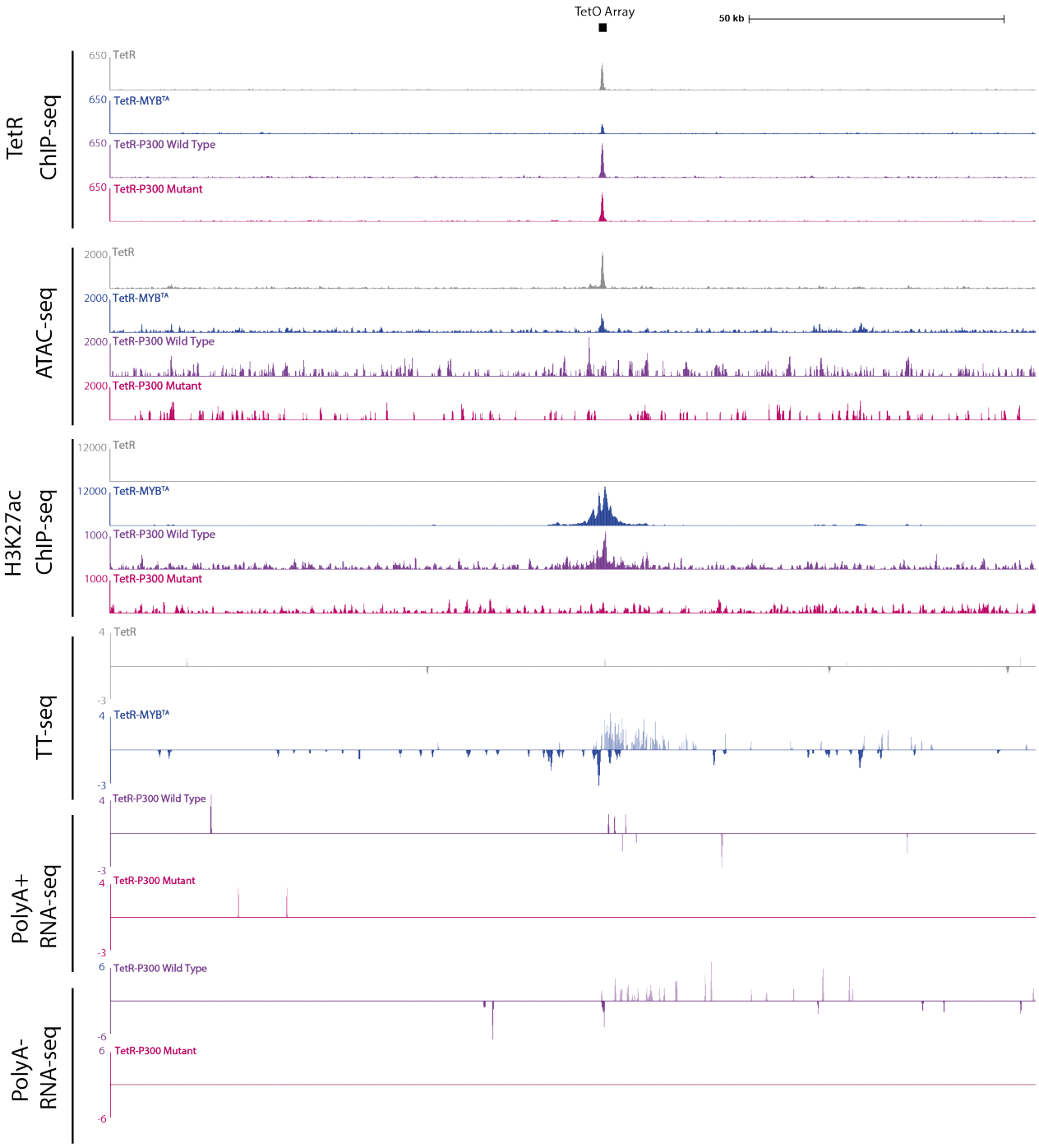


Figure 5.12 TetR-P300^{Core} WT binding causes an increase in H3K27ac at chrX-*TetO,* but no distal regions are activated

Representative tracks from UCSC of ATAC-seq (n=1), TetR ChIP-seq (n=1), H3K27ac ChIP-seq (n=1), strand-specific PolyA+ and PolyA- RNA-seq (n=1). Region shown is ~150kb surrounding the chrX-*TetO*, 50kb scale and *TetO* array shown. Cells are E14 mESC TOT2N variant cell line harbouring chr8-*TetO* and chrX-*TetO* transiently transfected with TetR-P300 WT or TetR-P300 D1399Y.

5.3 Discussion

When CBP/P300 are lost, H3K27ac reduction is more prominent at enhancers than at promoters (Raisner *et al.*, 2018; Martire *et al.*, 2020; Hogg *et al.*, 2021) as shown in Chapter 4. This indicates H3K27ac may, in part, drive enhancer-mediated transcription. However, it is currently unclear whether the recruitment of CBP and P300 to enhancers and the subsequent deposition of H3K27ac is a functional requirement for transcription or rather a bystander at active promoters and enhancers. In this chapter, I aimed to isolate the acetylation component of P300 ($P300^{Core}$) to understand if this was sufficient to establish an enhancer-like element, analogous to the effect of MYB^{TA} in Chapter 3. This might suggest that MYB^{TA} works through recruiting P300 and subsequent deposition of histone acetylation for its enhancer forming activity. We find that the $P300^{Core}$ can establish an enhancer-like element at chr8-*TetO,* but the signal achieved is diminished, potentially due to different biological techniques used. However, the enhancer-forming capability of $P300^{Core}$ is distinct from MYB^{TA} as a mix of common and unique distal elements are activated.

Impact of other acetylated residues

We employed the TetR tethering system in a neutral chromatin environment, aiming to determine whether P300-mediated acetylation in isolation is responsible for transcription driven by an enhancer-like element. By using this system, we eliminate many confounding factors that might also be contributing to enhancer function including endogenous regulatory elements, transcription factors bound to these elements and other KAT enzymes. P300 is known to acetylate many histone residues (Weinert *et al.*, 2018) but I chose to use H3K27ac as a proxy for P300-mediated acetylation. This is because the H3K27ac mark is the most well-studied and the antibody is very potent and effective in ChIP-seq experiments. However, it is well established that P300 can acetylate enhancer-associated transcription regulators including the cohesin complex,

MLL3/4 complex, mediator complex, super elongation complex and component of the general transcription machinery (Weinert *et al.*, 2018). Therefore, while I use H3K27ac as a mark of P300 activity, we cannot conclude that H3K27ac is the direct mediator of $P300^{Core}$-associated events. Instead, we can only conclude that the transcriptional effects observed are dependent on the presence of TetR-$P300^{Core}$ mediated acetylation, by comparison to the catalytically dead TetR-$P300^{Core}$ mutant (D1399Y).

Presence of the bromodomain in the P300 core

I attempted to isolate the acetylation component of P300 by fusing the P300 core domain to the TetR DNA binding domain. It was not possible to use the KAT domain alone, as both the KAT and bromodomains are required for stable protein folding (Delvecchio *et al.*, 2013). Furthermore, the core domain on its own is not sufficient to bind chromatin (Ferrie *et al.*, 2023), therefore fusing it to the TetR DNA binding domain allows association with chromatin.

The presence of the bromodomain in our TetR fusion protein was potentially problematic considering that the bromodomain possess its own functions distinct from acetyltransferase activity, including binding to acetylated residues. However, analysis of TetR-$P300^{Core}$ ChIP-seq revealed only a small number of binding sites genome-wide (data not shown) arguing that the bromodomain was not acting to direct the fusion protein to peaks of acetylated chromatin. In pursuit of generating cell lines expressing TetR-$P300^{Core}$, I was only able to successfully engineer cells stably expressing TetR-$P300^{Core}$ mutant and not those expressing TetR-$P300^{Core}$ wild type, suggesting the acetyltransferase component was causing cell death. P300 overexpression decreased proliferation of lung cancer cells, which was associated with increased apoptosis (Wang, S.-A. *et al.*, 2014). Consistent with this, transiently transfected TetR-$P300^{Core}$ cells showed highly elevated H3K27ac genome-wide (Figure 5.9.A).

Stable vs Transient Cell Lines

Due to the inability to create stable cell lines expressing TetR-P300^{Core} wild type, it was necessary to implement a transient transfection strategy coupled with cell sorting to harvest enough cells for next-generation sequencing experiments. A limitation of transient transfections is that in each condition a different number of cells will take up the plasmid making it difficult to compare within and between experiments. Additionally, fewer cells will tolerate toxic constructs, e.g. TetR-P300^{Core} wild type, compared to the mutant. By implementing a sorting step, I was able to enrich for cells expressing the fusion protein, by detection of eGFP, and harvest a similar number of cells per condition.

I found that the TetR-P300^{Core} wild type plasmid resulted in a high background of H3K27ac in ChIP-seq experiments meaning that the relative enrichment of signal at H3K27ac peaks was reduced. The increase in the number of reads at background loci results in a decrease in the proportion of reads mapping to H3K27ac peaks, so normalising the ChIP-seq data to total read count has the effect of reducing the height of these peaks artificially. Therefore, it was necessary to scale the H3K27ac ChIP-seq traces in the wild type and mutant conditions independently. Furthermore, it was also difficult to compare the transcriptional effect between the TetR and TetR-MYB^{TA} stable cell lines and TetR-P300^{Core} transient transfections due to higher overall signal in the stable cell lines compared to the transient transfections.

Conclusion

Despite having to use a mix of transient transfections and stable cell lines, we were able to make comparisons about which locations were activated rather than the relative strength of the peaks in these locations. To make a more compelling comparison I conducted transient transfections in all conditions (TetR, TetR-MYB^{TA}, TetR-P300^{Core}

wild type and TetR-P300^{Core} mutant) and conducted RT-qPCR at distally activated regions of interest. This led us to the conclusion that MYB^{TA} and P300^{Core} wild type can both activate the chromatin surrounding the chr8-*TetO*, but they activate both different and common regions, suggesting they act at least in part through some distinct mechanisms. It would be intriguing to investigate which other transcription related proteins may instead by required for MYB^{TA} driven enhancer activity. TAF12, a member of the basal transcription machinery, is known to interact with MYB^{TA}. TAF12 squelching has been shown to effectively disrupt the interaction between MYB and TFIID (Xu *et al.*, 2018), which could be employed in the TetR-MYB^{TA} cells to determine the effect on transcription at chr8-*TetO*.

The ability of TetR-P300^{Core} to activate transcription at distal elements agrees with data from Hilton *et al.*, 2015, using the same P300^{Core} domain in a dCas9 construct (dCas9-P300^{Core}). They observed that directly targeting dCas9-P300^{Core} to the enhancer of the beta globin locus using gRNAs in HEK293T cells caused a significant increase in expression of the beta globin genes *HBE, HBG and HBD* by RT-qPCR. It is possible that the increased in transcription observed here was more dramatic because the construct was targeted to endogenous enhancer/promoter pairings rather than a genomic element in neutral context, as with the TetR system, therefore allowing P300 to cooperate with other factors bound at this site.

Chapter 6: Discussion

Enhancers lie 10-100kb from target gene(s) and physically interact with the promoter to upregulate gene transcription. Histone acetylation is highly correlated with active regulatory elements but the functional role of histone acetylation in enhancer function is yet to be fully elucidated. CBP and P300 are highly similar lysine acetyltransferases (KATs) which deposit H3K27ac. It is thought that histone acetylation reduces the positive charge of histones, creating a permissive environment for transcription, along with acting as a binding site for transcription co-activators.

Using an *in vivo* chromatin targeting system (Blackledge *et al.*, 2014), previous results from our group have demonstrated that DNA binding by MYB, a transcription factor that interacts with several proteins including the KATs CBP and P300, establishes enhancer activity. This was achieved by fusing the transactivation domain of MYB to the TetR DNA binding domain, creating a TetR-MYB^{TA} fusion protein which can be directly targeted to an array of *TetO* sequences incorporated into the genome of mouse embryonic stem cells. We hypothesised that the recruitment of CBP/P300 by MYB and the subsequent deposition of histone acetylation may be sufficient to generate an enhancer.

To investigate whether the transcriptional function of MYB and CBP/P300 is a unique feature of the chromatin context or if these activities are maintained at different genomic loci, I engineered an additional *TetO* locus on chromosome X in a neutral region of chromatin, devoid of genes and functional elements. MYB^{TA} was no longer able to generate an enhancer-like element in this context but was able to promote transcription initiating from the *TetO* array. This suggests that the underlying DNA sequence of the target locus is key for MYB^{TA} to activate distal enhancers. This finding is consistent with existing research, which suggests that binding of MYB in different

genomic locations can have variable impacts on transcription based on the presence of cooperating TFs. In the case of T-ALL, for example, somatic mutations upstream of the TAL1 gene creates a MYB binding site, forming a novel enhancer consequently enhancing TAL1 gene expression (Mansour *et al.*, 2014). Motif searches revealed potential transcription factor binding sites for RUNX1, GATA3 and ETS1 in close proximity to this newly identified MYB binding site. This suggests that the presence of these cooperating TFs may play a crucial role in modulating MYB's ability to activate gene expression in this context.

To investigate if CBP/P300 are necessary for enhancer function, I generated a degron cell line for inducible Cbp degradation to investigate if Cbp was necessary for maintenance of enhancer activity established by MYB^{TA} at the *TetO* array. After observing compensatory activity of its paralogue, p300, I used a PROTAC for the inducible degradation of both Cbp and p300 and analysed the transcriptional effects at endogenous Cbp/p300 targets in a mESC model. The PROTAC, dCBP-1, which binds the bromodomain of Cbp/p300, caused a decrease in H3K27ac detected by western blot. At chromatin, there was global loss of H3K27ac, with a stronger loss at enhancers than promoters, coupled with genome-wide transcriptional dysregulation. This agrees with a multitude of research demonstrating that inhibition of CBP and P300 leads to pronounced gene expression changes (Raisner et al., 2018; Martire et al., 2020; Hogg et al., 2021; Vannan et al., 2021; Nicosia et al., 2023). The observed upregulation of genes is somewhat unexpected, given the strong association of CBP/P300 and histone acetylation with gene activation. However, one explanation for this could be the deacetylation of non-histone proteins causing secondary transcriptional effects. There was a reduction in H3K27ac at the promoter and enhancer of the ES cell specific gene *Sox2* and a reduction in nascent transcription from the *Sox2* gene. Strikingly, there was also a reduction in eRNA transcription from the super-enhancer driving *Sox2* expression. At the *TetO* array, there was a more modest effect on H3K27ac and

transcription, potentially due to the concentration and diversity of TF and co-activators present at this artificial locus (Kasper *et al.,* 2010). This may include a myriad of other HAT enzymes including those in the GNAT and MYST family that help to sustain a moderate level of H3K27ac in the absence of CBP and P300 (Martire *et al*., 2020). Despite this, there was a still a reduction in H3K27ac and transcription from distal elements originally activated by MYB^{TA}, suggesting that CBP/P300 are partially important for MYB^{TA} enhancer activity at the *TetO* array on chromosome 8.

The degradation of Cbp and p300 will likely have pronounced effects on the chromatin profile and transcription, considering that Cbp and p300 harbour both structural and enzymatic roles. To investigate the involvement of acetylation specifically in enhancer function, I isolated the P300 core domain, containing the KAT and BRD domain (Hilton *et al.*, 2015) and fused it to the TetR DNA binding domain, creating a TetR-$P300^{Core}$ fusion protein, and analysed the transcriptional effects at the *TetO* locus. This system has the advantage of eliminating the complexity of transcription from nearby genes, or the potential activities of other proteins bound to the chromatin. I found that the $P300^{Core}$ could activate distal elements, including an increase in chromatin accessibility measured by ATAC-seq and an increase in transcription, measured by polyA- RNA-seq, from these distal elements. This finding corroborates previous findings using the same $P300^{Core}$ domain in a dCas9 construct (Hilton *et al.,* 2015). In this study, they directed the dCas9-$P300^{Core}$ to gene promoters and enhancers of *IL1RN*, *MYOD* and *OCT4,* leading to robust gene activation from the corresponding gene. The strong transcription observed in this study, assessed by qRT-PCR, may have been more pronounced compared with TetR-$P300^{Core}$ due to the direct targeting to endogenous enhancer/promoter pairings that are likely already co-bound by important TFs and co-activators. Interestingly, some of these activated distal elements were shared with MYB^{TA} and some distinct. This suggests that MYB^{TA} and $P300^{Core}$ act partially through distinct mechanisms in their roles in enhancer activity.

Future Work

To further explore the involvement of MYB, CBP/P300 and H3K27ac in enhancer activity, there are several outstanding questions.

*1. What are the properties of chr8-*TetO *that allow it to be activated compared to chrX-*TetO*?*

In chapter 3 and 5, I demonstrated that TetR-MYB^{TA} and TetR-P300^{Core} can activate distal chromatin at chr8-*TetO* but can only activate local chromatin at chrX-*TetO*. To address why this might be, it will be important to investigate the DNA sequence and chromatin profile surrounding (+/- 100kb) both chr8-*TetO* and chrX-*TetO* to provide insight into why distal regions in chr8-*TetO* were activated by TetR-MYB^{TA} and TetR-P300^{Core} but not distal regions in chrX-*TetO*. Motif analysis at the activated regions will identify potential transcription factor binding sites that may explain their activation, which could further be validated by ChIP for these TFs. Additionally, ChIP-seq for H3K4me1 may reveal whether the activated regions in chr8-*TetO* are 'primed' cryptic regulatory sites.

2. Are CBP/P300 necessary for maintaining chromatin structure?

It will be intriguing to investigate whether CBP and P300 are necessary for the maintaining chromatin structure. ATAC-seq and MCC after PROTAC treatment would provide information on the role of CBP and P300 in chromatin accessibility and 3D chromatin structure (in particular in maintaining proximity between the enhancer and promoter) respectively. This includes MCC at endogenous ESC-specific genes including *Sox2*, capturing from the promoter and the enhancer, indicating the involvement of CBP/P300-mediated acetylation in enhancer-promoter interactions. Additionally, MCC capturing from the *TetO* array bound by TetR-MYB^{TA} would provide insight on whether a reduction in acetylation at an artificial locus causes a reduction in 3D chromatin contacts driven by TetR-MYB^{TA}. By using the ultra-high resolution

technique MCC, allowing us to observe even subtle differences in interaction profile, we would be able to identify whether there are precise or more broad regions of DNA that are dependent on CBP/P300 for interactions.

3. What other factors may be involved in MYB^{TA} and P300^{Core} enhancer activity?

In chapter 5, I demonstrated that MYB^{TA} and P300^{Core} have distinct roles in enhancer activity at chr8-*TetO*. Additionally, in chapter 4, I showed that reducing CBP/P300-mediated acetylation does not abrogate enhancer activity at chr8-*TetO*. This suggests that TetR-MYB^{TA} and TetR-P300^{Core} may be interacting with distinct factors in order to activate the different distal elements. Therefore, it would be intriguing to carry out mass spectrometry with the TetR-MYB^{TA} and TetR-P300^{Core} constructs to identify which proteins they interact with and may therefore recruit to the *TetO* array, which may, along with further experiments, provide information on which proteins drive their different activities.

Appendix

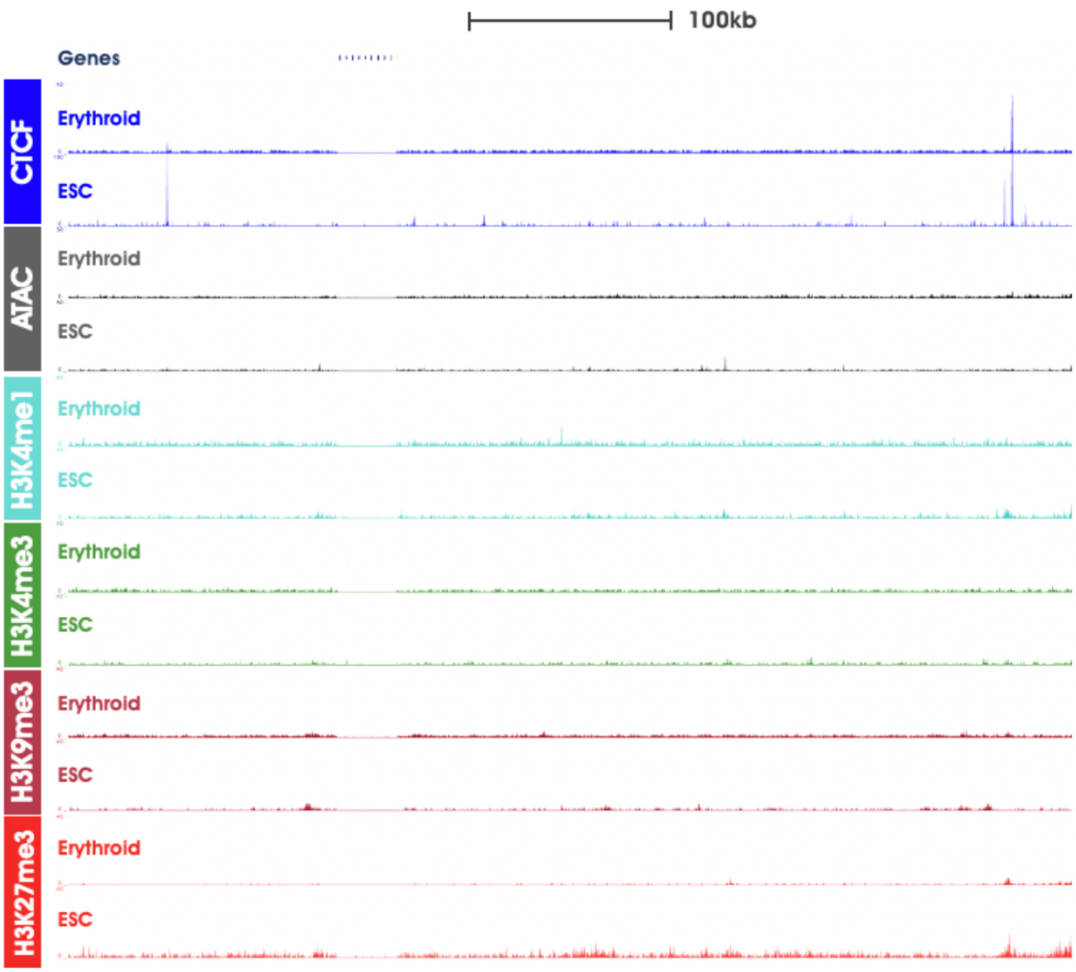


Figure 7.1 Identification of a neutral region of chromatin in mESCs

Figure taken from Dr Emily Georgiades' thesis (Georgiades, E. 2022), data generated by Dr Caz Harrold (Harrold, C. 2019).

chrX:10,935,857-11,046,248 (mm9). A neutral region in this context is defined as a region of chromatin that is depleted of CTCF and the histone marks: H3K4me1, H3K4me3, H3K9me3 and H3K27me3 (by ChIP-seq), open chromatin sites (by ATAC-seq), annotated genes and with no underlying TAD structure. For full details of track availability refer to Dr Caz Harrold's thesis (Harrold, C. 2019).

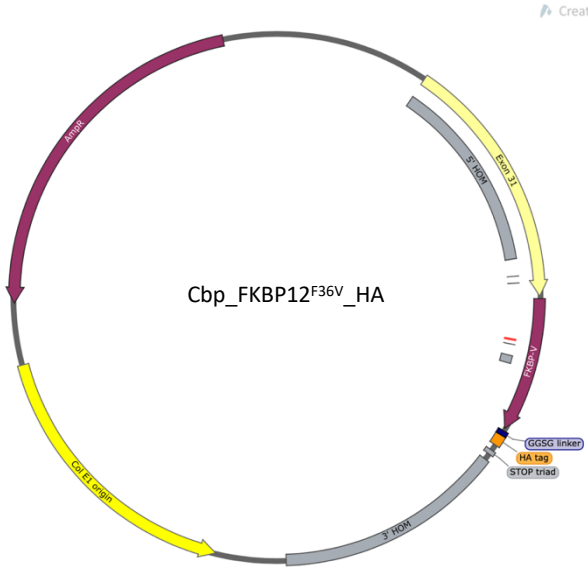


Figure 7.2: Cbp_FKBP12^{F36V}_HA donor vector map

Total vector size is 3901bp containing the FKBP12^{F36V} domain, flanked by ~500bp homology arms targeting the C terminus of Cbp. The vector includes an HA tag joined to FKBP12^{F36V} with an GGSG linker and restriction enzyme sites for easy exchange of homology arms.

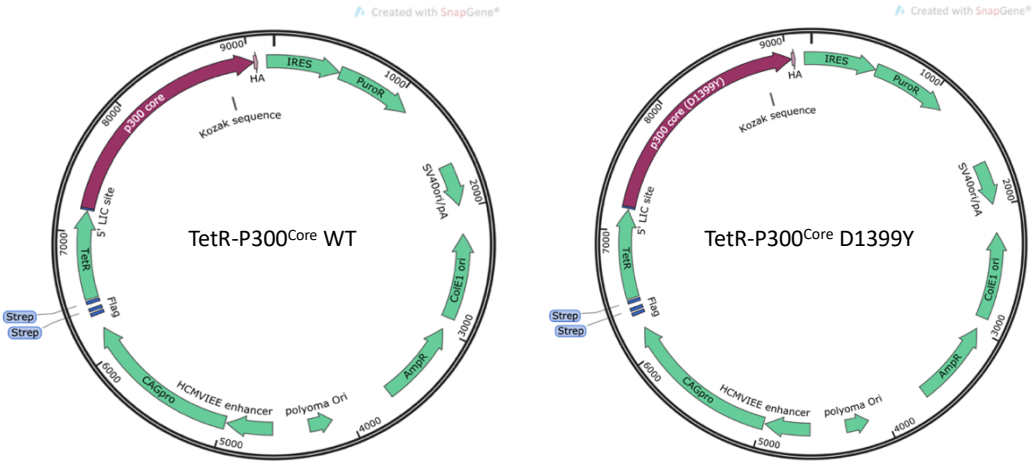


Figure 7.3 TetR-P300^{Core} plasmid maps

The total vector size is 9195bp for both plasmids. The D1399Y vector differs by one amino acid in the p300 core domain, which abrogates HAT activity. The vector also includes restriction enzyme sites for easy exchange of the p300 core domain for future experiments. A flag-strep

tag is at the C-terminus of the TetR protein and an HA tag at the N terminus. An IRES separates the TetR-P300^{Core} from the puromycin resistance marker, used for making stable cell lines.

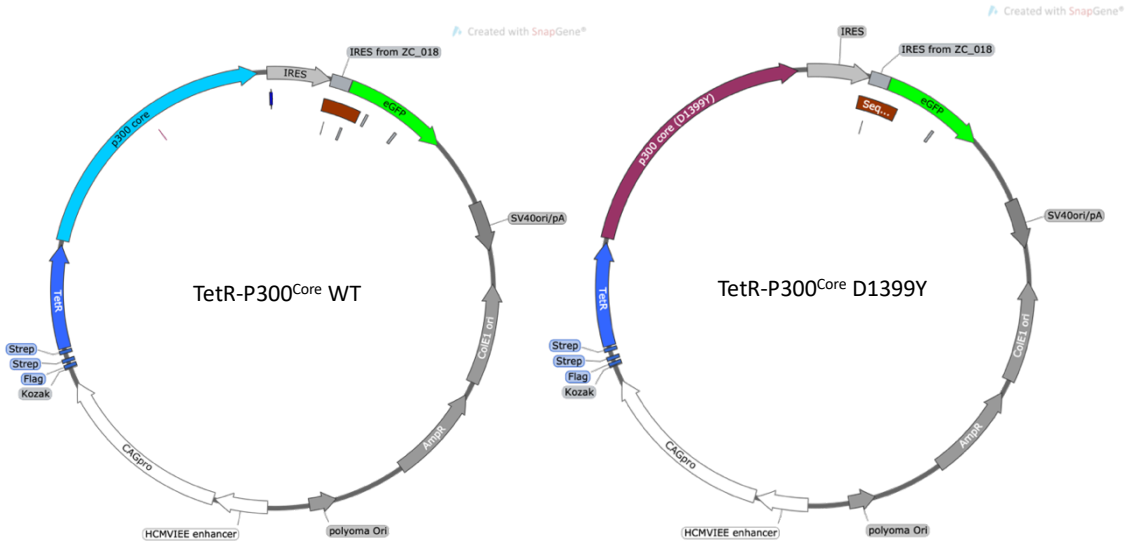


Figure 7.4 TetR-P300^{Core} plasmid maps with GFP marker

The total vector size is 9195bp for both plasmids. The D1399Y vector differs by one amino acid in the p300 core domain, which abrogates HAT activity. The vector also includes restriction enzyme sites for easy exchange of the p300 core domain for future experiments. A flag-strep tag is at the C-terminus of the TetR protein and an HA tag at the N terminus. An IRES separates the TetR-P300^{Core} from the GFP marker, which can be used as a gate for sorting experiments.

Table 7.1: Chromosome 8 *TetO* Primers for ChIP-qPCR and RT-qPCR

Primer Name	Sequence
TetO_-5_F	TTGGACATAGCACTGAGAAAGAAG
TetO_-5_F	TCCAGACAGGGACTTTGAGTG
TetO_-1_F	TTTTCCTCATGATGGTACATGG
TetO_-1_R	AATTTGTTGTAGGTGCCATTCA
TetO_-0.5_F	CTTTAGGAGGGGAGCTGCAT
TetO_-0.5_R	TTGCCAGTTTTAAGAGAATTTGC
TetO_F	TGTGGGCAGGTCACAAGTTA
TetO_R	ATCTCATCCCCAATGCTGTC
TetO_+0.5_F	AAAGGAAAATGAGGGGCCTA
TetO_+0.5_R	CAGGCAGGTAAACACGATAAGTT
TetO_+1_F	CAAATCCGTCCTTGATCTTGA
TetO_+1_R	GCAGTAGGATGGGGAGACAG
TetO_+5_F	TTGAGAGGGGAGAAAAATGG
TetO_+5_R	TGGAATCACAAATTTGAACTGG
MYB_TetO_Chr8_F	TGAGACAACATCTCGCCTTG
MYB_TetO_Chr8_R	AGGCAGGAGAATCACTTGAAC
P300_TetO_Chr8_F	GCCTGTAATCCCAGCACTTT
P300_TetO_Chr8_R	CCACCACACCCAGCTAATTT

Table 7.2: Chromosome X *TetO* Primers for ChIP-qPCR and RT-qPCR

Primer Name	Sequence
TetO_-5_F	TGGTGTCTCTGATTGGTCTTTC
TetO_-5_R	GATCTTCTTCTGGCTTAGAACATAAAC
TetO_-1_F	ATGTGGGTAGTACCAGTGAATAAG
TetO_-1_R	CACATGATGGTGGTCAGGAA
TetO_-0.5_F	CATGGCATTGCTGAATGTATCC
TetO_-0.5_R	CAACACTGACCCAAGGAAGT
TetO_F	TGTGGGCAGGTCACAAGTAT
TetO_R	TGGTGGACTGCTATCTCTATC
TetO_+0.5_F	CCCTGAAGGTAAGAACAGGAATTA
TetO_+0.5_R	AGGGAGAAACAGAGACATAGGA
TetO_+1_F	GCAACTTACTGGTGGAACTCT
TetO_+1_R	CATGATGGCCTTTGCTGTTG

Table 7.3: Endogenous Loci Primers for ChIP-qPCR and RT-qPCR

Primer Name	Sequence
Oct4 Enhancer Fw	GTGGTGTGAGCAAGTAGGTAG
Oct4 Enhancer Rv	CTGGCTTAGAGTGTGAACCTTAT
Sox2 Enhancer 3 Fw	GCTTACTGGAACTTTCCCTAGAA
Sox2 Enhancer 3 Rv	GGAGACCTGCCATCATCAAA
Sox2 Enhancer 1 Fw	CCTTCCAGAGACAACGCTAAT
Sox2 Enhancer 1 Rv	GGCTCGGCCATCTATGATTTA
GAPDH Promoter Fw	AGAGAGGGAGGAGGGGAAATG
GAPDH Promoter Rv	AACAGGGAGGAGCAGAGAGCAC
18205 Fw	TGGCATAGTCCAAGCAGGGATAG
18205 Rv	TCAGATGAGAGACCAAGGCAGAG
GAPDH Intron 2 Fw	CGGAGCAACAGATGTGTGTA
GAPDH Intron 2 Rv	CGCCCTGATCTGAGGTTAAAT
GAPDH Exon 3 Fw	CAGTGGCAAAGTGGAGATTG
GAPDH Exon 3 Rv	AATTTGCCGTGAGTGGAGTC

Table 7.4 Characterisation of Chr8 Primers for Genotyping

Primer Name	Sequence
mouse_A (1)	TGCGCAGTCGCATTGTTTAC
human_TetO_B (3)	GGAGCTTACCATCCAGGATCCATC
human_TetO_A (2)	GGAAGTGACAGGAGTGAGAAAGG
NeoR_B (5)	AGCCGAATAGCCTCTCCACCC
NeoR_B (4)	TTTACGGTATCGCCGCTCCC
mouse_A (6)	ATACACGATCTTGGTTAGCTGGAGG

Table 7.5: Screening Primers for ChrX *TetO* Insertion

Primer Name	Sequence
ChrX_TetO_LH_Fw	GTGCTGCAATGAATGTCAGAAGG
ChrX_TetO_LH_Rv	ACCTGCCCACATCTCTATCAC
ChrX_TetO_RH_Fw	AGACCAAATCTCCTGGGTAAGAG
ChrX_TetO_RH_Rv	GGAAGGGAGAGAGACATGGAGAG

Table 7.6 Screening Primers for Cbp degron cell line

Primer Name	Sequence
Cbp_N_LH_Fw	TTGGAGATGCCCACAGAGTTCAC
Cbp_N_LH_Rv	GCCACTCTCGTCTTCGATGTGG
Cbp_N_RH_Fw	AACTTAAAGGGCTTGTTTCTGTCGC
Cbp_N_RH_Rv	CCCCAACATGACAAACATGAATCC

References

Alexander, J.M. *et al.* (2019) 'Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity', *eLife*. Edited by R.H. Singer, K. Struhl, and Z. Liu, 8, p. e41769. Available at: https://doi.org/10.7554/eLife.41769.

Aljahani, A. *et al.* (2022) 'Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF', *Nature Communications*, 13(1), p. 2139. Available at: https://doi.org/10.1038/s41467-022-29696-5.

Attar, N. and Kurdistani, S.K. (2017) 'Exploitation of EP300 and CREBBP Lysine Acetyltransferases by Cancer', *Cold Spring Harbor Perspectives in Medicine*, 7(3), p. a026534. Available at: https://doi.org/10.1101/cshperspect.a026534.

Baek, H.J., Kang, Y.K. and Roeder, R.G. (2006) 'Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly', *The Journal of Biological Chemistry*, 281(22), pp. 15172–15181. Available at: https://doi.org/10.1074/jbc.M601983200.

Banerji, J., Rusconi, S. and Schaffner, W. (1981) 'Expression of a β-globin gene is enhanced by remote SV40 DNA sequences', *Cell*, 27(2 PART 1), pp. 299–308. Available at: https://doi.org/10.1016/0092-8674(81)90413-X.

Bannister, A.J. and Kouzarides, T. (2011) 'Regulation of chromatin by histone modifications', *Cell Research*, 21(3), pp. 381–395. Available at: https://doi.org/10.1038/cr.2011.22.

Bedford, D.C. *et al.* (2010) 'Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases', *Epigenetics*, 5(1), pp. 9–15. Available at: https://doi.org/10.4161/epi.5.1.10449.

Benoist, C. and Chambon, P. (1981) 'In vivo sequence requirements of the SV40 early promoter region', *Nature*, 290(5804), pp. 304–310. Available at: https://doi.org/10.1038/290304a0.

Bernstein, B.E. *et al.* (2010) 'The NIH Roadmap Epigenomics Mapping Consortium', *Nature Biotechnology*, 28(10), pp. 1045–1048. Available at: https://doi.org/10.1038/nbt1010-1045.

Blackledge, N.P. *et al.* (2014) 'Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation', *Cell*, 157(6), p. 1445. Available at: https://doi.org/10.1016/j.cell.2014.05.004.

Blayney, J.W. *et al.* (2023) 'Super-enhancers include classical enhancers and facilitators to fully activate gene expression', *Cell*, 186(26), pp. 5826-5839.e18. Available at: https://doi.org/10.1016/j.cell.2023.11.030.

Bleckwehl, T. *et al.* (2021) 'Enhancer-associated H3K4 methylation safeguards in vitro germline competence', *Nature Communications*, 12(1), p. 5771. Available at: https://doi.org/10.1038/s41467-021-26065-6.

Blobel, G.A. *et al.* (2021) 'Testing the super-enhancer concept', *Nature Reviews Genetics*, 22(12), pp. 749–755. Available at: https://doi.org/10.1038/s41576-021-00398-w.

Boija, A. *et al.* (2018) 'Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains', *Cell*, 175(7), pp. 1842-1855.e16. Available at: https://doi.org/10.1016/j.cell.2018.10.042.

Bonn, S. *et al.* (2012) 'Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development', *Nature Genetics*, 44(2), pp. 148–156. Available at: https://doi.org/10.1038/ng.1064.

Borrow, J. *et al.* (1996) 'The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB–binding protein', *Nature Genetics*, 14(1), pp. 33–41. Available at: https://doi.org/10.1038/ng0996-33.

Bose, D.A. *et al.* (2017) 'RNA Binding to CBP Stimulates Histone Acetylation and Transcription', *Cell*, 168(1), pp. 135-149.e22. Available at: https://doi.org/10.1016/j.cell.2016.12.020.

Bose, D.A. and Berger, S.L. (2017) 'eRNA binding produces tailored CBP activity profiles to regulate gene expression', *RNA Biology*, 14(12), pp. 1655–1659. Available at: https://doi.org/10.1080/15476286.2017.1353862.

Bowers, E.M. *et al.* (2010) 'Virtual Ligand Screening of the p300/CBP Histone Acetyltransferase: Identification of a Selective Small Molecule Inhibitor', *Chemistry & Biology*, 17(5), pp. 471–482. Available at: https://doi.org/10.1016/j.chembiol.2010.03.006.

Boyes, J. *et al.* (1998) 'Regulation of activity of the transcription factor GATA-1 by acetylation', *Nature*, 396(6711), pp. 594–598. Available at: https://doi.org/10.1038/25166.

Boyson, S.P. *et al.* (2021) 'Functional Roles of Bromodomain Proteins in Cancer', *Cancers*, 13(14), p. 3606. Available at: https://doi.org/10.3390/cancers13143606.

Brooks, N. *et al.* (2021) 'CCS1477, a Novel p300/CBP Bromodomain Inhibitor, Enhances Efficacy of Azacitidine and Venetoclax in Pre-Clinical Models of Acute Myeloid Leukaemia and Lymphoma', *Blood*, 138, p. 3291. Available at: https://doi.org/10.1182/blood-2021-148295.

Buenrostro, J. *et al.* (2015) 'ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide', *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, 109, p. 21.29.1-21.29.9. Available at: https://doi.org/10.1002/0471142727.mb2129s109.

Bulynko, Y.A. and O'Malley, B.W. (2011) 'Nuclear receptor coactivators: structural and functional biochemistry', *Biochemistry*, 50(3), pp. 313–328. Available at: https://doi.org/10.1021/bi101762x.

Chan, H.M. and La Thangue, N.B. (2001) 'p300/CBP proteins: HATs for transcriptional bridges and scaffolds', *Journal of Cell Science*, 114(13), pp. 2363–2373. Available at: https://doi.org/10.1242/jcs.114.13.2363.

Chang, P.-Y. *et al.* (2010) 'Binding of the MLL PHD3 finger to histone H3K4me3 is required for MLL-dependent gene transcription', *Journal of Molecular Biology*, 400(2), pp. 137–144. Available at: https://doi.org/10.1016/j.jmb.2010.05.005.

Chen, R.H. and Lipsick, J.S. (1993) 'Differential transcriptional activation by v-myb and c-myb in animal cells and Saccharomyces cerevisiae.', *Molecular and Cellular Biology*, 13(7), pp. 4423–4431.

Chen, J. *et al.* (2014) 'Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells', *Cell*, 156(6), pp. 1274–1285. Available at: https://doi.org/10.1016/j.cell.2014.01.062.

Chen, X. *et al.* (2008) 'Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells', *Cell*, 133(6), pp. 1106–1117. Available at: https://doi.org/10.1016/j.cell.2008.04.043.

Cho, W.-K. *et al.* (2018) 'Mediator and RNA polymerase II clusters associate in transcription-dependent condensates', *Science (New York, N.Y.)*, 361(6400), pp. 412–415. Available at: https://doi.org/10.1126/science.aar4199.

Cirillo, L.A. *et al.* (2002) *Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4 binding of GAL4 to the GAL1 promoter in yeast precedes recruitment of SAGA (Dudley et al., 1999). It remains to be determined how these and other transcription factors recognize their binding sites within silent chromatin and whether they require other proteins for initial binding*, *Molecular Cell*.

Crawford, T.D. *et al.* (2016) 'Discovery of a Potent and Selective in Vivo Probe (GNE-272) for the Bromodomains of CBP/EP300', *Journal of Medicinal Chemistry*, 59(23), pp. 10549–10563. Available at: https://doi.org/10.1021/acs.jmedchem.6b01022.

Cremer, T. *et al.* (1982) 'Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments', *Human Genetics*, 62(3), pp. 201–209. Available at: https://doi.org/10.1007/BF00333519.

Cremer, T. and Cremer, C. (2001) 'Chromosome territories, nuclear architecture and gene regulation in mammalian cells', *Nature Reviews Genetics*, 2(4), pp. 292–301. Available at: https://doi.org/10.1038/35066075.

Creyghton, M.P. *et al.* (2010) 'Histone H3K27ac separates active from poised enhancers and predicts developmental state', *Proceedings of the National Academy of Sciences*, 107(50), pp. 21931–21936. Available at: https://doi.org/10.1073/pnas.1016071107.

Crump, N.T. *et al.* (2021) 'BET inhibition disrupts transcription but retains enhancer-promoter contact', *Nature Communications*, 12(1), p. 223. Available at: https://doi.org/10.1038/s41467-020-20400-z.

Crump, N.T. *et al.* (2023) 'MLL-AF4 cooperates with PAF1 and FACT to drive high-density enhancer interactions in leukemia', *Nature Communications*, 14(1), p. 5208. Available at: https://doi.org/10.1038/s41467-023-40981-9.

Cui, K. *et al.* (2009) 'Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation', *Cell Stem Cell*, 4(1), pp. 80–93. Available at: https://doi.org/10.1016/j.stem.2008.11.011.

Cutter, A.R. and Hayes, J.J. (2015) 'A brief review of nucleosome structure', *FEBS letters*, 589(20 Pt A), pp. 2914–2922. Available at: https://doi.org/10.1016/j.febslet.2015.05.016.

Dai, P. *et al.* (1996) 'CBP as a transcriptional coactivator of c-Myb.', *Genes & Development*, 10(5), pp. 528–540. Available at: https://doi.org/10.1101/gad.10.5.528.

Dancy, B.M. and Cole, P.A. (2015) 'Protein Lysine Acetylation by p300/CBP', *Chemical Reviews*, 115(6), pp. 2419–2452. Available at: https://doi.org/10.1021/cr500452k.

Davies, J.O.J. *et al.* (2016) 'Multiplexed analysis of chromosome conformation at vastly improved sensitivity', *Nature Methods*, 13(1), pp. 74–80. Available at: https://doi.org/10.1038/nmeth.3664.

Davies, J.O.J. *et al.* (2017) 'How best to identify chromosomal interactions: a comparison of approaches', *Nature Methods*, 14(2), pp. 125–134. Available at: https://doi.org/10.1038/nmeth.4146.

Delvecchio, M. *et al.* (2013) 'Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation', *Nature Structural & Molecular Biology*, 20(9), pp. 1040–1046. Available at: https://doi.org/10.1038/nsmb.2642.

de Wit, E. *et al.* (2015) 'CTCF Binding Polarity Determines Chromatin Looping', *Molecular Cell*, 60(4), pp. 676–684. Available at: https://doi.org/10.1016/j.molcel.2015.09.023.

Dixon, J.R. *et al.* (2012) 'Topological domains in mammalian genomes identified by analysis of chromatin interactions', *Nature*, 485(7398), pp. 376–380. Available at: https://doi.org/10.1038/nature11082.

Dou, Y. *et al.* (2006) 'Regulation of MLL1 H3K4 methyltransferase activity by its core components', *Nature Structural & Molecular Biology*, 13(8), pp. 713–719. Available at: https://doi.org/10.1038/nsmb1128.

Doyon, Y. *et al.* (2006) 'ING Tumor Suppressor Proteins Are Critical Regulators of Chromatin Acetylation Required for Genome Expression and Perpetuation', *Molecular Cell*, 21(1), pp. 51–64. Available at: https://doi.org/10.1016/j.molcel.2005.12.007.

Dunham, I. *et al.* (2012) 'An integrated encyclopedia of DNA elements in the human genome', *Nature*, 489(7414), pp. 57–74. Available at: https://doi.org/10.1038/nature11247.

Dyson, H.J. and Wright, P.E. (2016) 'Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300', *Journal of Biological Chemistry*, 291(13), pp. 6714–6722. Available at: https://doi.org/10.1074/jbc.R115.692020.

Eberharter, A. and Becker, P.B. (2002) 'Histone acetylation: a switch between repressive and permissive chromatin', *EMBO Reports*, 3(3), pp. 224–229. Available at: https://doi.org/10.1093/embo-reports/kvf053.

Ernst, P. *et al.* (2001) 'MLL and CREB Bind Cooperatively to the Nuclear Coactivator CREB-Binding Protein', *Molecular and Cellular Biology*, 21(7), p. 2249. Available at: https://doi.org/10.1128/MCB.21.7.2249-2258.2001.

Esnault, C. *et al.* (2008) 'Mediator-dependent recruitment of TFIIH modules in preinitiation complex', *Molecular Cell*, 31(3), pp. 337–346. Available at: https://doi.org/10.1016/j.molcel.2008.06.021.

Fang, F. *et al.* (2014) 'Coactivators p300 and CBP Maintain the Identity of Mouse Embryonic Stem Cells by Mediating Long-Range Chromatin Structure', *Stem Cells*, 32(7), pp. 1805–1816. Available at: https://doi.org/10.1002/stem.1705.

Fauquier, L. *et al.* (2018) 'CBP and P300 regulate distinct gene networks required for human primary myoblast differentiation and muscle integrity', *Scientific Reports*, 8(1), p. 12629. Available at: https://doi.org/10.1038/s41598-018-31102-4.

Feaver, W. 4 *et al.* (1994) *Relationship of CDK-Activating Kinase and RNA Polymerase II CTD Kinase TFIIH/TFIIK*.

Felzien, L.K. *et al.* (1999) 'Specificity of Cyclin E-Cdk2, TFIIB, and E1A Interactions with a Common Domain of the p300 Coactivator', *Molecular and Cellular Biology*, 19(6), pp. 4241–4246.

Feng, Q. *et al.* (2002) 'Methylation of H3-Lysine 79 Is Mediated by a New Family of HMTases without a SET Domain', *Current Biology*, 12(12), pp. 1052–1058. Available at: https://doi.org/10.1016/S0960-9822(02)00901-6.

Ferrari, K.J. *et al.* (2014) 'Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity', *Molecular Cell*, 53(1), pp. 49–62. Available at: https://doi.org/10.1016/j.molcel.2013.10.030.

Ferrie, J.J. *et al.* (2023) 'p300 Is an Obligate Integrator of Combinatorial Transcription Factor Inputs'. bioRxiv, p. 2023.05.18.541220. Available at: https://doi.org/10.1101/2023.05.18.541220.

Filippakopoulos, P. *et al.* (2012) 'Histone recognition and large-scale structural analysis of the human bromodomain family', *Cell*, 149(1), pp. 214–231. Available at: https://doi.org/10.1016/j.cell.2012.02.013.

Fishburn, J. *et al.* (2015) 'Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation', *Proceedings of the National Academy of Sciences of the United States of America*, 112(13), pp. 3961–3966. Available at: https://doi.org/10.1073/pnas.1417709112.

Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Haberle V, Lassmann T, Kulakovskiy IV, Lizio M, Itoh M, Andersson R, Mungall CJ, Meehan TF, Schmeier S, Bertin N, Jorgensen M, Dimont E, Arner E, Schmidl C, Schaefer U *et al* (2014) A promoter-level mammalian expression atlas. *Nature* 507: 462–470. Available at: https://doi.org/10.1038/nature13182.

Fudenberg, G. *et al.* (2016) 'Formation of Chromosomal Domains by Loop Extrusion', *Cell Reports*, 15(9), pp. 2038–2049. Available at: https://doi.org/10.1016/j.celrep.2016.04.085.

Fuglerud, B.M. *et al.* (2017) 'A c-Myb mutant causes deregulated differentiation due to impaired histone binding and abrogated pioneer factor function', *Nucleic Acids Research*, 45(13), pp. 7681–7696. Available at: https://doi.org/10.1093/nar/gkx364.

Furlong, E.E.M. and Levine, M. (2018) 'Developmental enhancers and chromosome topology', *Science*. American Association for the Advancement of Science, pp. 1341–1345. Available at: https://doi.org/10.1126/science.aau0320.

Gartner, L.P. and Hiatt, J.L. (2006) *Color Textbook of Histology E-Book*. Elsevier Health Sciences.

Georgiades, E. *et al.* (2023) 'Active regulatory elements recruit cohesin to establish cell-specific chromatin domains'. bioRxiv, p. 2023.10.13.562171. Available at: https://doi.org/10.1101/2023.10.13.562171.

Georgiades, E. University of Oxford (2022).

Gilan, O. *et al.* (2016) 'Functional interdependence of BRD4 and DOT1L in MLL leukemia', *Nature Structural and Molecular Biology*, 23(7), pp. 673–681. Available at: https://doi.org/10.1038/nsmb.3249.

Glozak, M.A. *et al.* (2005) 'Acetylation and deacetylation of non-histone proteins', *Gene*, 363, pp. 15–23. Available at: https://doi.org/10.1016/j.gene.2005.09.010.

Godfrey, L. *et al.* (2019) 'DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation', *Nature Communications*, 10(1), p. 2803. Available at: https://doi.org/10.1038/s41467-019-10844-3.

Gong, F., Chiu, L.-Y. and Miller, K.M. (2016) 'Acetylation Reader Proteins: Linking Acetylation Signaling to Genome Maintenance and Cancer', *PLoS Genetics*, 12(9), p. e1006272. Available at: https://doi.org/10.1371/journal.pgen.1006272.

Goto, N.K. *et al.* (2002) 'Cooperativity in Transcription Factor Binding to the Coactivator CREB-binding Protein (CBP): THE MIXED LINEAGE LEUKEMIA PROTEIN (MLL) ACTIVATION DOMAIN BINDS TO AN ALLOSTERIC SITE ON THE KIX DOMAIN *', *Journal of Biological Chemistry*, 277(45), pp. 43168–43174. Available at: https://doi.org/10.1074/jbc.M207660200.

Grant, P.A. *et al.* (1999) 'Expanded Lysine Acetylation Specificity of Gcn5 in Native Complexes *', *Journal of Biological Chemistry*, 274(9), pp. 5895–5900. Available at: https://doi.org/10.1074/jbc.274.9.5895.

Gryder, B.E., Wu, L., *et al.* (2019) 'Chemical genomics reveals histone deacetylases are required for core regulatory transcription', *Nature Communications*, 10, p. 3004. Available at: https://doi.org/10.1038/s41467-019-11046-7.

Gryder, B.E., Pomella, S., *et al.* (2019) 'Histone hyperacetylation disrupts core gene regulatory architecture in rhabdomyosarcoma', *Nature genetics*, 51(12), pp. 1714–1722. Available at: https://doi.org/10.1038/s41588-019-0534-4.

Gu, W. and Roeder, R.G. (1997) 'Activation of p53 Sequence-Specific DNA Binding by Acetylation of the p53 C-Terminal Domain', *Cell*, 90(4), pp. 595–606. Available at: https://doi.org/10.1016/S0092-8674(00)80521-8.

Gurova, K. *et al.* (2018) 'Structure and function of the histone chaperone FACT – resolving FACTual issues', *Biochimica et biophysica acta. Gene regulatory mechanisms*, pp. S1874-9399(18)30159–7. Available at: https://doi.org/10.1016/j.bbagrm.2018.07.008.

Hager, G.L., McNally, J.G. and Misteli, T. (2009) 'Transcription dynamics', *Molecular Cell*, 35(6), pp. 741–753. Available at: https://doi.org/10.1016/j.molcel.2009.09.005.

Halsall, J.A. *et al.* (2015) 'Cells adapt to the epigenomic disruption caused by histone deacetylase inhibitors through a coordinated, chromatin-mediated transcriptional response', *Epigenetics & Chromatin*, 8, p. 29. Available at: https://doi.org/10.1186/s13072-015-0021-9.

Hansen, A.S. *et al.* (2019) 'Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF', *Molecular Cell*, 76(3), pp. 395-411.e13. Available at: https://doi.org/10.1016/j.molcel.2019.07.039.

Hanssen, L.L.P. *et al.* (2017) 'Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo', *Nature Cell Biology*, 19(8), pp. 952–961. Available at: https://doi.org/10.1038/ncb3573.

Harrold, C. University of Oxford (2019).

Hay, D., Hughes, J.R., Babbs, C., Davies, James O. J., *et al.* (2016) 'Genetic dissection of the α-globin super-enhancer in vivo', *Nature Genetics*, 48(8), pp. 895–903. Available at: https://doi.org/10.1038/ng.3605.

Heintzman, N.D. *et al.* (2007) 'Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome', *Nature Genetics*, 39(3), pp. 311–318. Available at: https://doi.org/10.1038/ng1966.

Henry, R.A. *et al.* (2015) 'Changing the Selectivity of p300 by Acetyl-CoA Modulation of Histone Acetylation', *ACS Chemical Biology*, 10(1), pp. 146–156. Available at: https://doi.org/10.1021/cb500726b.

Henry, R.A., Kuo, Y.-M. and Andrews, A.J. (2013) 'Differences in specificity and selectivity between CBP and p300 acetylation of histone H3 and H3/H4', *Biochemistry*, 52(34), pp. 5746–5759. Available at: https://doi.org/10.1021/bi400684q.

Hilton, I.B. *et al.* (2015) 'Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers', *Nature Biotechnology*, 33(5), pp. 510–517. Available at: https://doi.org/10.1038/nbt.3199.

von Hippel, P.H. and Berg, O.G. (1989) 'Facilitated target location in biological systems', *The Journal of Biological Chemistry*, 264(2), pp. 675–678.

Hnisz, D. *et al.* (2013) 'Super-Enhancers in the Control of Cell Identity and Disease', *Cell*, 155(4), pp. 934–947. Available at: https://doi.org/10.1016/j.cell.2013.09.053.

Hogg, S.J. *et al.* (2021) 'Targeting histone acetylation dynamics and oncogenic transcription by catalytic P300/CBP inhibition', *Molecular cell*, 81(10), pp. 2183-2200.e13. Available at: https://doi.org/10.1016/j.molcel.2021.04.015.

Hörnblad, A. *et al.* (2021) 'Dissection of the Fgf8 regulatory landscape by in vivo CRISPR-editing reveals extensive intra- and inter-enhancer redundancy', *Nature Communications*, 12(1), p. 439. Available at: https://doi.org/10.1038/s41467-020-20714-y.

Hua, P. *et al.* (2021) 'Defining genome architecture at base-pair resolution', *Nature*, 595(7865), pp. 125–129. Available at: https://doi.org/10.1038/s41586-021-03639-4.

Hughes, J.R. *et al.* (2013) 'High-resolution analysis of *cis* -acting regulatory networks at the α-globin locus', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1620), p. 20120361. Available at: https://doi.org/10.1098/rstb.2012.0361.

Hung, H.-L. *et al.* (1999) 'CREB-Binding Protein Acetylates Hematopoietic Transcription Factor GATA-1 at Functionally Important Sites', *Molecular and Cellular Biology*, 19(5), p. 3496. Available at: https://doi.org/10.1128/mcb.19.5.3496.

Ida, K. *et al.* (1997) 'Adenoviral E1A-Associated Protein p300 Is Involved in Acute Myeloid Leukemia With t(11; 22)(q23; q13)', *Blood*, 90(12), pp. 4699–4704. Available at: https://doi.org/10.1182/blood.V90.12.4699.

Imhof, A. *et al.* (1997) 'Acetylation of general transcription factors by histone acetyltransferases', *Current Biology*, 7(9), pp. 689–692. Available at: https://doi.org/10.1016/S0960-9822(06)00296-X.

Jacquet, K. *et al.* (2016) 'The TIP60 Complex Regulates Bivalent Chromatin Recognition by 53BP1 through Direct H4K20me Binding and H2AK15 Acetylation', *Molecular Cell*, 62(3), pp. 409–421. Available at: https://doi.org/10.1016/j.molcel.2016.03.031.

Jang, M.K. *et al.* (2005) 'The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription', *Molecular cell*, 19(4), pp. 523–534. Available at: https://doi.org/10.1016/j.molcel.2005.06.027.

Jin, C. *et al.* (2009) 'H3.3/H2A.Z double variant–containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions', *Nature Genetics*, 41(8), pp. 941–945. Available at: https://doi.org/10.1038/ng.409.

Jin, C. and Felsenfeld, G. (2007) 'Nucleosome stability mediated by histone variants H3.3 and H2A.Z', *Genes & Development*, 21(12), pp. 1519–1529. Available at: https://doi.org/10.1101/gad.1547707.

Jin, Q. *et al.* (2011) 'Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation', *The EMBO Journal*, 30(2), pp. 249–262. Available at: https://doi.org/10.1038/emboj.2010.318.

Jishage, M. *et al.* (2012) 'Transcriptional regulation by Pol II(G) involving mediator and competitive interactions of Gdown1 and TFIIF with Pol II', *Molecular Cell*, 45(1), pp. 51–63. Available at: https://doi.org/10.1016/j.molcel.2011.12.014.

John, S. *et al.* (2000) 'The Something About Silencing protein, Sas3, is the catalytic subunit of NuA3, a yTAFII30-containing HAT complex that interacts with the Spt16 subunit of the yeast CP (Cdc68/Pob3)–FACT complex', *Genes & Development*, 14(10), pp. 1196–1208.

Johnson, K.M. *et al.* (2002) 'TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA', *Genes & Development*, 16(14), pp. 1852–1863. Available at: https://doi.org/10.1101/gad.995702.

Josling, G.A. *et al.* (2012) 'The Role of Bromodomain Proteins in Regulating Gene Expression', *Genes*, 3(2), pp. 320–343. Available at: https://doi.org/10.3390/genes3020320.

Kasper, L.H. *et al.* (2002) 'A transcription-factor-binding surface of coactivator p300 is required for haematopoiesis', *Nature*, 419(6908), pp. 738–743. Available at: https://doi.org/10.1038/nature01062.

Kasper, L.H. *et al.* (2010) 'CBP/p300 double null cells reveal effect of coactivator level and diversity on CREB transactivation', *The EMBO Journal*, 29(21), pp. 3660–3672. Available at: https://doi.org/10.1038/emboj.2010.235.

Keegan, L., Gill, G. and Ptashne, M. (1986) 'Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein', *Science*, 231(4739), pp. 699–704. Available at: https://doi.org/10.1126/science.3080805.

Kerry, J. *et al.* (2017) 'MLL-AF4 Spreading Identifies Binding Sites that Are Distinct from Super-Enhancers and that Govern Sensitivity to DOT1L Inhibition in Leukemia', *Cell Reports*, 18(2), pp. 482–495. Available at: https://doi.org/10.1016/j.celrep.2016.12.054.

Kim, J.L., Nikolov, D.B. and Burley, S.K. (1993) 'Co-crystal structure of TBP recognizing the minor groove of a TATA element', *Nature*, 365(6446), pp. 520–527. Available at: https://doi.org/10.1038/365520a0.

Kimura, A. and Horikoshi, M. (1998) 'Tip60 acetylates six lysines of a specific class in core histones in vitro', *Genes to Cells: Devoted to Molecular & Cellular Mechanisms*, 3(12), pp. 789–800. Available at: https://doi.org/10.1046/j.1365-2443.1998.00229.x.

Klose, R.J. *et al.* (2013) 'Chromatin Sampling—An Emerging Perspective on Targeting Polycomb Repressor Proteins', *PLOS Genetics*, 9(8), p. e1003717. Available at: https://doi.org/10.1371/journal.pgen.1003717.

Kornberg, R.D. (1977) 'Structure of Chromatin', *Annual Review of Biochemistry*, 46(1), pp. 931–954. Available at: https://doi.org/10.1146/annurev.bi.46.070177.004435.

Kostrewa, D. *et al.* (2009) 'RNA polymerase II-TFIIB structure and mechanism of transcription initiation', *Nature*, 462(7271), pp. 323–330. Available at: https://doi.org/10.1038/nature08548.

Lai, F. *et al.* (2013) 'Activating RNAs associate with Mediator to enhance chromatin architecture and transcription', *Nature*, 494(7438), pp. 497–501. Available at: https://doi.org/10.1038/nature11884.

Lamonica, J.M. *et al.* (2011) 'Bromodomain protein Brd3 associates with acetylated GATA1 to promote its chromatin occupancy at erythroid target genes', *Proceedings of the National Academy of Sciences of the United States of America*, 108(22), pp. E159-168. Available at: https://doi.org/10.1073/pnas.1102140108.

Lamonica, J.M., Vakoc, C.R. and Blobel, G.A. (2006) 'Acetylation of GATA-1 is required for chromatin occupancy', *Blood*, 108(12), pp. 3736–3738. Available at: https://doi.org/10.1182/blood-2006-07-032847.

Larson, A.G. *et al.* (2017) 'Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin', *Nature*, 547(7662), pp. 236–240. Available at: https://doi.org/10.1038/nature22822.

Lasko, L.M. *et al.* (2017) 'Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours', *Nature*, 550(7674), pp. 128–132. Available at: https://doi.org/10.1038/nature24028.

Lau, I. University of Oxford (2019).
Lee, J.-H. and Skalnik, D.G. (2005) 'CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex', *The Journal of Biological Chemistry*, 280(50), pp. 41725–41731. Available at: https://doi.org/10.1074/jbc.M508312200.

Li, B., Carey, M. and Workman, J.L. (2007) 'The Role of Chromatin during Transcription', *Cell*, 128(4), pp. 707–719. Available at: https://doi.org/10.1016/j.cell.2007.01.015.

Li, W. *et al.* (2013) 'Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation', *Nature*, 498(7455), pp. 516–520. Available at: https://doi.org/10.1038/nature12210.

Li, Y. *et al.* (2014) 'CRISPR Reveals a Distal Super-Enhancer Required for Sox2 Expression in Mouse Embryonic Stem Cells', *PLOS ONE*, 9(12), p. e114485. Available at: https://doi.org/10.1371/journal.pone.0114485.

Li, Z. *et al.* (2012) 'Foxa2 and H2A.Z Mediate Nucleosome Depletion during Embryonic Stem Cell Differentiation', *Cell*, 151(7), pp. 1608–1616. Available at: https://doi.org/10.1016/j.cell.2012.11.018.

Lieberman-Aiden, E., van Berkum, N.L., *et al.* (2009) 'Comprehensive mapping of long range interactions reveals folding principles of the human genome', *Science (New York, N.Y.)*, 326(5950), pp. 289–293. Available at: https://doi.org/10.1126/science.1181369.

Liu, L. *et al.* (1999) 'p53 Sites Acetylated In Vitro by PCAF and p300 Are Acetylated In Vivo in Response to DNA Damage', *Molecular and Cellular Biology*, 19(2), p. 1202. Available at: https://doi.org/10.1128/mcb.19.2.1202.

Liu, X. *et al.* (2008) 'The structural basis of protein acetylation by the p300/CBP transcriptional coactivator', *Nature*, 451(7180), pp. 846–850. Available at: https://doi.org/10.1038/nature06546.

Long, H.K., Prescott, S.L. and Wysocka, J. (2016) 'Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution', *Cell*, 167(5), pp. 1170–1187. Available at: https://doi.org/10.1016/j.cell.2016.09.018.

Luger, K. *et al.* (1997) 'Crystal structure of the nucleosome core particle at 2.8 A resolution', *Nature*, 389(6648), pp. 251–260. Available at: https://doi.org/10.1038/38444.

Luger, K. and Richmond, T.J. (1998) 'The histone tails of the nucleosome', *Current Opinion in Genetics and Development*, 8(2), pp. 140–146. Available at: https://doi.org/10.1016/S0959-437X(98)80134-2.

Malik, A.N. *et al.* (2014) 'Genome-wide identification and characterization of functional neuronal activity-dependent enhancers', *Nature neuroscience*, 17(10), pp. 1330–1339. Available at: https://doi.org/10.1038/nn.3808.

Maniatis, T., Goodbourn, S. and Fischer, J. (1987) 'Regulation of inducible and tissue-specific gene expression', *Science*, 236(4806), pp. 1237–1245. Available at: https://doi.org/10.1126/science.3296191.

Marmorstein, R. and Zhou, M.-M. (2014) 'Writers and Readers of Histone Acetylation: Structure, Mechanism, and Inhibition', *Cold Spring Harbor Perspectives in Biology*, 6(7), p. a018762. Available at: https://doi.org/10.1101/cshperspect.a018762.

Martire, S. *et al.* (2020) 'Differential contribution of p300 and CBP to regulatory element acetylation in mESCs', *BMC Molecular and Cell Biology*, 21(1), p. 55. Available at: https://doi.org/10.1186/s12860-020-00296-9.

Mashtalir, N. *et al.* (2021) 'Chromatin landscape signals differentially dictate the activities of mSWI/SNF family complexes', *Science*, 373(6552), pp. 306–315. Available at: https://doi.org/10.1126/science.abf8705.

McManus, K.J. and Hendzel, M.J. (2003) 'Quantitative Analysis of CBP- and P300-Induced Histone Acetylations In Vivo Using Native Chromatin', *Molecular and Cellular Biology*, 23(21), pp. 7611–7627. Available at: https://doi.org/10.1128/MCB.23.21.7611-7627.2003.

Mifsud, B. *et al.* (2015) 'Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C', *Nature Genetics*, 47(6), pp. 598–606. Available at: https://doi.org/10.1038/ng.3286.

Milne, T.A. *et al.* (2002) 'MLL targets SET domain methyltransferase activity to Hox gene promoters', *Molecular Cell*, 10(5), pp. 1107–1117. Available at: https://doi.org/10.1016/s1097-2765(02)00741-4.

Milne, T.A. *et al.* (2010) 'Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis', *Molecular Cell*, 38(6), pp. 853–863. Available at: https://doi.org/10.1016/j.molcel.2010.05.011

Mishima, Y. *et al.* (2011) 'The Hbo1-Brd1/Brpf2 complex is responsible for global acetylation of H3K14 and required for fetal liver erythropoiesis', *Blood*, 118(9), pp. 2443–2453. Available at: https://doi.org/10.1182/blood-2011-01-331892.

Mousavi, K. *et al.* (2013) 'eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci', *Molecular Cell*, 51(5), pp. 606–617. Available at: https://doi.org/10.1016/j.molcel.2013.07.022.

Näär, A.M., Lemon, B.D. and Tjian, R. (2001) 'Transcriptional coactivator complexes', *Annual Review of Biochemistry*, 70, pp. 475–501. Available at: https://doi.org/10.1146/annurev.biochem.70.1.475.

Nabet, B. *et al.* (2018) 'The dTAG system for immediate and target-specific protein degradation', *Nature chemical biology*, 14(5), pp. 431–441. Available at: https://doi.org/10.1038/s41589-018-0021-8.

Nakajima, T. *et al.* (1997) 'RNA helicase A mediates association of CBP with RNA polymerase II', *Cell*, 90(6), pp. 1107–1112. Available at: https://doi.org/10.1016/s0092-8674(00)80376-1.

Nakamura, T. *et al.* (2002) 'ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation', *Molecular Cell*, 10(5), pp. 1119–1128. Available at: https://doi.org/10.1016/s1097-2765(02)00740-2.

Narita, T. *et al.* (2021) 'Enhancers are activated by p300/CBP activity-dependent PIC assembly, RNAPII recruitment, and pause release', *Molecular Cell*, 81(10), pp. 2166-2182.e6. Available at: https://doi.org/10.1016/j.molcel.2021.03.008.

Narita, T. *et al.* (2023) 'Acetylation of histone H2B marks active enhancers and predicts CBP/p300 target genes', *Nature Genetics*, 55(4), pp. 679–692. Available at: https://doi.org/10.1038/s41588-023-01348-4.

Neumayr, C. *et al.* (2022) 'Differential cofactor dependencies define distinct types of human enhancers', *Nature*, 606(7913), pp. 406–413. Available at: https://doi.org/10.1038/s41586-022-04779-x.

Nicosia, L. *et al.* (2023) 'Therapeutic targeting of EP300/CBP by bromodomain inhibition in hematologic malignancies', *Cancer Cell*, pp. S1535-6108(23)00366–5. Available at: https://doi.org/10.1016/j.ccell.2023.11.001.

Nora, E.P. *et al.* (2012) 'Spatial partitioning of the regulatory landscape of the X-inactivation centre', *Nature*, 485(7398), pp. 381–385. Available at: https://doi.org/10.1038/nature11049.

Ogryzko, V.V. *et al.* (1996) 'The Transcriptional Coactivators p300 and CBP Are Histone Acetyltransferases', *Cell*, 87(5), pp. 953–959. Available at: https://doi.org/10.1016/S0092-8674(00)82001-2.

Orlando, D.A. *et al.* (2014) 'Quantitative ChIP-Seq normalization reveals global modulation of the epigenome', *Cell Reports*, 9(3), pp. 1163–1170. Available at: https://doi.org/10.1016/j.celrep.2014.10.018.

Oliveira, A.M.M. *et al.* (2006) 'Differential role for CBP and p300 CREB-binding domain in motor skill learning', *Behavioral Neuroscience*, 120(3), pp. 724–729. Available at: https://doi.org/10.1037/0735-7044.120.3.724.

Parvin, J.D. and Sharp, P.A. (1993) 'DNA topology and a minimal set of basal factors for transcription by RNA polymerase II', *Cell*, 73(3), pp. 533–540. Available at: https://doi.org/10.1016/0092-8674(93)90140-l.

Pasini, D. *et al.* (2010) 'Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes', *Nucleic Acids Research*, 38(15), pp. 4958–4969. Available at: https://doi.org/10.1093/nar/gkq244.

Pattabiraman, D.R. *et al.* (2009) 'Mutations in multiple domains of c-Myb disrupt interaction with CBP/p300 and abrogate myeloid transforming ability', *Molecular cancer research: MCR*, 7(9), pp. 1477–1486. Available at: https://doi.org/10.1158/1541-7786.MCR-09-0070.

Pattabiraman, D.R. *et al.* (2014) 'Interaction of c-Myb with p300 is required for the induction of acute myeloid leukemia (AML) by human AML oncogenes', *Blood*, 123(17), pp. 2682–2690. Available at: https://doi.org/10.1182/blood-2012-02-413187.

Pelham-Webb, B. *et al.* (2021) 'H3K27ac bookmarking promotes rapid post-mitotic activation of the pluripotent stem cell program without impacting 3D chromatin reorganization', *Molecular Cell*, 81(8), pp. 1732-1748.e8. Available at: https://doi.org/10.1016/j.molcel.2021.02.032.

Peterlin, B.M. and Price, D.H. (2006) 'Controlling the Elongation Phase of Transcription with P-TEFb', *Molecular Cell*. Mol Cell, pp. 297–305. Available at: https://doi.org/10.1016/j.molcel.2006.06.014.

Petrif, F. *et al.* (1995) 'Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP', *Nature*, 376(6538), pp. 348–351. Available at: https://doi.org/10.1038/376348a0.

Plaschka, C. *et al.* (2015) 'Architecture of the RNA polymerase II-Mediator core initiation complex', *Nature*, 518(7539), pp. 376–380. Available at: https://doi.org/10.1038/nature14229.

Plotnikov, A.N. *et al.* (2014) 'Structural Insights into Acetylated-Histone H4 Recognition by the Bromodomain-PHD Finger Module of Human Transcriptional Co-Activator CBP', *Structure (London, England : 1993)*, 22(2), pp. 353–360. Available at: https://doi.org/10.1016/j.str.2013.10.021.

Rada-Iglesias, A. *et al.* (2011) 'A unique chromatin signature uncovers early developmental enhancers in humans', *Nature*, 470(7333), pp. 279–283. Available at: https://doi.org/10.1038/nature09692.

Radzisheuskaya, A. *et al.* (2021) 'Complex-dependent histone acetyltransferase activity of KAT8 determines its role in transcription and cellular homeostasis', *Molecular Cell*, 81(8), pp. 1749-1765.e8. Available at: https://doi.org/10.1016/j.molcel.2021.02.012.

Rahnamoun, H. *et al.* (2018) 'RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation', *Nature Structural & Molecular Biology*, 25(8), pp. 687–697. Available at: https://doi.org/10.1038/s41594-018-0102-0.

Raisner, R. *et al.* (2018) 'Enhancer Activity Requires CBP/P300 Bromodomain-Dependent Histone H3K27 Acetylation', *Cell Reports*, 24(7), pp. 1722–1729. Available at: https://doi.org/10.1016/j.celrep.2018.07.041.

Ran, F.A. *et al.* (2013) 'Genome engineering using the CRISPR-Cas9 system', *Nature Protocols*, 8(11), pp. 2281–2308. Available at: https://doi.org/10.1038/nprot.2013.143.

Rao, S.S.P. *et al.* (2014) 'A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping', *Cell*, 159(7), pp. 1665–1680. Available at: https://doi.org/10.1016/j.cell.2014.11.021.

Rathert, P. *et al.* (2015) 'Transcriptional plasticity promotes primary and acquired resistance to BET inhibition', *Nature*, 525(7570), pp. 543–547. Available at: https://doi.org/10.1038/nature14898.

Romero, F.A. *et al.* (2017) 'GNE-781, A Highly Advanced Potent and Selective Bromodomain Inhibitor of Cyclic Adenosine Monophosphate Response Element Binding Protein, Binding Protein (CBP)', *Journal of Medicinal Chemistry*, 60(22), pp. 9162–9183. Available at: https://doi.org/10.1021/acs.jmedchem.7b00796.

Rougvie, A.E. and Lis, J.T. (1988) *The RNA Polymerase II Molecule at the 5' End of the Uninduced hspTO Gene of D. melanogaster Is Transcriptionally Engaged*, *Cell*.

Rühlmann, F. *et al.* (2019) 'The prognostic capacities of CBP and p300 in locally advanced rectal cancer', *World Journal of Surgical Oncology*, 17, p. 224. Available at: https://doi.org/10.1186/s12957-019-1764-8.

Ruthenburg, A.J. *et al.* (2011) 'Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions', *Cell*, 145(5), pp. 692–706. Available at: https://doi.org/10.1016/j.cell.2011.03.053.

Sabari, B.R. *et al.* (2018) 'Coactivator condensation at super-enhancers links phase separation and gene control', *Science*, 361(6400), p. eaar3958. Available at: https://doi.org/10.1126/science.aar3958.

Sainsbury, S., Bernecky, C. and Cramer, P. (2015) 'Structural basis of transcription initiation by RNA polymerase II', *Nature Reviews Molecular Cell Biology*, 16(3), pp. 129–143. Available at: https://doi.org/10.1038/nrm3952.

Sakaguchi, K. *et al.* (1998) 'DNA damage activates p53 through a phosphorylation–acetylation cascade', *Genes & Development*, 12(18), p. 2831. Available at: https://doi.org/10.1101/gad.12.18.2831.

Sandberg, M.L. *et al.* (2005) 'c-Myb and p300 Regulate Hematopoietic Stem Cell Proliferation and Differentiation', *Developmental Cell*, 8(2), pp. 153–166. Available at: https://doi.org/10.1016/j.devcel.2004.12.015.

Sano, Y. and Ishii, S. (2001) 'Increased Affinity of c-Myb for CREB-binding Protein (CBP) after CBP-induced Acetylation *', *Journal of Biological Chemistry*, 276(5), pp. 3674–3682. Available at: https://doi.org/10.1074/jbc.M006896200.

Sanyal, A. *et al.* (2012) 'The long-range interaction landscape of gene promoters', *Nature*, 489(7414), pp. 109–113. Available at: https://doi.org/10.1038/nature11279.

Sartorelli, V. and Lauberth, S.M. (2020) 'Enhancer RNAs are an important regulatory layer of the epigenome', *Nature Structural & Molecular Biology*, 27(6), pp. 521–528. Available at: https://doi.org/10.1038/s41594-020-0446-0.

Schneider, I. (1972) 'Cell lines derived from late embryonic stages of Drosophila melanogaster', *Development*, 27(2), pp. 353–365. Available at: https://doi.org/10.1242/dev.27.2.353.

Seizl, M. *et al.* (2011) 'Mediator head subcomplex Med11/22 contains a common helix bundle building block with a specific function in transcription initiation complex stabilization', *Nucleic Acids Research*, 39(14), pp. 6291–6304. Available at: https://doi.org/10.1093/nar/gkr229.

Sexton, T. *et al.* (2012) 'Three-dimensional folding and functional organization principles of the Drosophila genome', *Cell*, 148(3), pp. 458–472. Available at: https://doi.org/10.1016/j.cell.2012.01.010.

Shen, X. *et al.* (2008) 'EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency', *Molecular Cell*, 32(4), pp. 491–502. Available at: https://doi.org/10.1016/j.molcel.2008.10.016.

Shin, H.Y. *et al.* (2016) 'Hierarchy within the mammary STAT5-driven Wap super-enhancer', *Nature Genetics*, 48(8), pp. 904–911. Available at: https://doi.org/10.1038/ng.3606.

Shvedunova, M. and Akhtar, A. (2022) 'Modulation of cellular processes by histone and non-histone protein acetylation', *Nature Reviews Molecular Cell Biology*, 23(5), pp. 329–349. Available at: https://doi.org/10.1038/s41580-021-00441-y.

Shvets, A.A., Kochugaeva, M.P. and Kolomeisky, A.B. (2018) 'Mechanisms of Protein Search for Targets on DNA: Theoretical Insights', *Molecules : A Journal of Synthetic Chemistry and Natural Product Chemistry*, 23(9), p. 2106. Available at: https://doi.org/10.3390/molecules23092106.

Simó-Riudalbas, L. *et al.* (2015) 'KAT6B Is a Tumor Suppressor Histone H3 Lysine 23 Acetyltransferase Undergoing Genomic Loss in Small Cell Lung Cancer', *Cancer Research*, 75(18), pp. 3936–3945. Available at: https://doi.org/10.1158/0008-5472.CAN-14-3702.

Staby, L. *et al.* (2017) 'Eukaryotic transcription factors: Paradigms of protein intrinsic disorder', *Biochemical Journal*. Portland Press Ltd, pp. 2509–2532. Available at: https://doi.org/10.1042/BCJ20160631.

Stolper, R.J. *et al.* (2023) 'Loop extrusion by cohesin plays a key role in enhancer-activated gene expression during differentiation'. bioRxiv, p. 2023.09.07.556660. Available at: https://doi.org/10.1101/2023.09.07.556660.

Sun, X.-J. *et al.* (2015) 'The Role of Histone Acetyltransferases in Normal and Malignant Hematopoiesis', *Frontiers in Oncology*, 5, p. 108. Available at: https://doi.org/10.3389/fonc.2015.00108.

Sungalee, S. *et al.* (2021) 'Histone acetylation dynamics modulates chromatin conformation and allele-specific interactions at oncogenic loci', *Nature Genetics*, 53(5), pp. 650–662. Available at: https://doi.org/10.1038/s41588-021-00842-x.

Tafessu, A. and Banaszynski, L.A. (2020) 'Establishment and function of chromatin modification at enhancers', *Open Biology*, 10(10), p. 200255. Available at: https://doi.org/10.1098/rsob.200255.

Takahashi, S. and Yokoyama, A. (2020) 'The molecular functions of common and atypical MLL fusion protein complexes', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1863(7), p. 194548. Available at: https://doi.org/10.1016/j.bbagrm.2020.194548.

Tanaka, Y. *et al.* (2000) 'Extensive brain hemorrhage and embryonic lethality in a mouse null mutant of CREB-binding protein', *Mechanisms of Development*, 95(1), pp. 133–145. Available at: https://doi.org/10.1016/S0925-4773(00)00360-9.

Tang, Z. *et al.* (2013) 'SET1 and p300 Act Synergistically, through Coupled Histone Modifications, in Transcriptional Activation by p53', *Cell*, 154(2), pp. 297–310. Available at: https://doi.org/10.1016/j.cell.2013.06.027.

Thompson, P.R. *et al.* (2004) 'Regulation of the p300 HAT domain via a novel activation loop', *Nature Structural & Molecular Biology*, 11(4), pp. 308–315. Available at: https://doi.org/10.1038/nsmb740.

Tie, F. *et al.* (2009) 'CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing', *Development*, 136(18), pp. 3131–3141. Available at: https://doi.org/10.1242/dev.037127.

Tomita, A. *et al.* (2000) 'c-Myb acetylation at the carboxyl-terminal conserved domain by transcriptional co-activator p300', *Oncogene*, 19(3), pp. 444–451. Available at: https://doi.org/10.1038/sj.onc.1203329.

Tonnemacher, S., Eltsov, M. and Jakob, B. (2020) 'Correlative Light and Electron Microscopy (CLEM) Analysis of Nuclear Reorganization Induced by Clustered DNA Damage Upon Charged Particle Irradiation', *International Journal of Molecular Sciences*, 21(6), p. 1911. Available at: https://doi.org/10.3390/ijms21061911.

Tsai, F.T. and Sigler, P.B. (2000) 'Structural basis of preinitiation complex assembly on human pol II promoters', *The EMBO journal*, 19(1), pp. 25–36. Available at: https://doi.org/10.1093/emboj/19.1.25.

Vannam, R. *et al.* (2021) 'Targeted degradation of the enhancer lysine acetyltransferases CBP and p300', *Cell Chemical Biology*, 28(4), pp. 503-514.e12. Available at: https://doi.org/10.1016/J.CHEMBIOL.2020.12.004.

Vaquerizas, J.M. *et al.* (2009) 'A census of human transcription factors: function, expression and evolution', *Nature Reviews Genetics*, 10(4), pp. 252–263. Available at: https://doi.org/10.1038/nrg2538.

Verrijzer, C.P. *et al.* (1995) 'Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II', *Cell*, 81(7), pp. 1115–1125. Available at: https://doi.org/10.1016/s0092-8674(05)80016-9.

Vo ngoc, L. *et al.* (2020) 'Identification of the human DPR core promoter element using machine learning', *Nature*, 585(7825), pp. 459–463. Available at: https://doi.org/10.1038/s41586-020-2689-7.

Voss, A.K. *et al.* (2012) 'MOZ regulates the Tbx1 locus, and Moz mutation partially phenocopies DiGeorge syndrome', *Developmental Cell*, 23(3), pp. 652–663. Available at: https://doi.org/10.1016/j.devcel.2012.07.010.

Waanders, E. *et al.* (2020) 'Mutational Landscape and Patterns of Clonal Evolution in Relapsed Pediatric Acute Lymphoblastic Leukemia', *Blood Cancer Discovery*, 1(1), pp. 96–111. Available at: https://doi.org/10.1158/0008-5472.BCD-19-0041.

Wada, T. *et al.* (1998) 'DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs', *Genes & Development*, 12(3), pp. 343–356. Available at: https://doi.org/10.1101/gad.12.3.343.

Waddington, C.H. (1942) 'Canalization of Development and the Inheritance of Acquired Characters', *Nature*, 150(3811), pp. 563–565. Available at: https://doi.org/10.1038/150563a0.

Wang, L. *et al.* (2008) 'Structure and chemistry of the p300/CBP and Rtt109 histone acetyltransferases: Implications for histone acetyltransferase evolution and function', *Current opinion in structural biology*, 18(6), pp. 741–747. Available at: https://doi.org/10.1016/j.sbi.2008.09.004.

Wang, Z. *et al.* (2010) 'Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression', *Cell*, 141(7), pp. 1183–1194. Available at: https://doi.org/10.1016/j.cell.2010.05.016.

Wang, S.-A. *et al.* (2014) 'Phosphorylation of p300 increases its protein degradation to enhance the lung cancer progression', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1843(6), pp. 1135–1149. Available at: https://doi.org/10.1016/j.bbamcr.2014.02.001.

Wang, C. *et al.* (2016) 'Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition', *Proceedings of the National Academy of Sciences of the United States of America*, 113(42), pp. 11871–11876. Available at: https://doi.org/10.1073/pnas.1606857113.

Weinert, B.T. *et al.* (2018) 'Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/p300 Acetylome', *Cell*, 174(1), pp. 231-244.e12. Available at: https://doi.org/10.1016/j.cell.2018.04.033.

Whyte, W.A. *et al.* (2013) 'Master transcription factors and mediator establish super-enhancers at key cell identity genes', *Cell*, 153(2), pp. 307–319. Available at: https://doi.org/10.1016/j.cell.2013.03.035.

Won, K.-J. *et al.* (2012) 'Global identification of transcriptional regulators of pluripotency and differentiation in embryonic stem cells', *Nucleic Acids Research*, 40(17), pp. 8199–8209. Available at: https://doi.org/10.1093/nar/gks584.

Xu, Y. *et al.* (2018) 'A TFIID-SAGA Perturbation that Targets MYB and Suppresses Acute Myeloid Leukemia', *Cancer Cell*, 33(1), pp. 13-28.e8. Available at: https://doi.org/10.1016/j.ccell.2017.12.002.

Yamaguchi, Y. *et al.* (1999) 'NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation', *Cell*, 97(1), pp. 41–51. Available at: https://doi.org/10.1016/S0092-8674(00)80713-8.

Yanazume, T. *et al.* (2003) 'Biological role of p300 in cardiac myocytes', *Molecular and Cellular Biochemistry*, 248(1), pp. 115–119. Available at: https://doi.org/10.1023/A:1024132217870.

Yang, X.-J. and Ullah, M. (2007) 'MOZ and MORF, two large MYSTic HATs in normal and cancer stem cells', *Oncogene*, 26(37), pp. 5408–5419. Available at: https://doi.org/10.1038/sj.onc.1210609.

Yang, Z. *et al.* (2005) 'Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4', *Molecular Cell*, 19(4), pp. 535–545. Available at: https://doi.org/10.1016/j.molcel.2005.06.029.

Yao, T.-P. *et al.* (1998) 'Gene Dosage–Dependent Embryonic Development and Proliferation Defects in Mice Lacking the Transcriptional Integrator p300', *Cell*, 93(3), pp. 361–372. Available at: https://doi.org/10.1016/S0092-8674(00)81165-4.

Yokoyama, A. and Cleary, M.L. (2008) 'Menin critically links MLL proteins with LEDGF on cancer-associated target genes', *Cancer Cell*, 14(1), pp. 36–46. Available at: https://doi.org/10.1016/j.ccr.2008.05.003.

Yuh, C.H. *et al.* (1994) ‘Complexity and organization of DNA-protein interactions in the 5′-regulatory region of an endoderm-specific marker gene in the sea urchin embryo’, *Mechanisms of Development*, 47(2), pp. 165–186. Available at: https://doi.org/10.1016/0925-4773(94)90088-4.

Zeng, L. *et al.* (2008) ‘Structural Basis of Site-Specific Histone Recognition by the Bromodomains of Human Coactivators PCAF and CBP/p300’, *Structure*, 16(4), pp. 643–652. Available at: https://doi.org/10.1016/j.str.2008.01.010.

Zeng, L. *et al.* (2010) ‘Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b’, *Nature*, 466(7303), pp. 258–262. Available at: https://doi.org/10.1038/nature09139.

Zentner, G.E., Tesar, P.J. and Scacheri, P.C. (2011) ‘Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions’, *Genome Research*, 21(8), pp. 1273–1283. Available at: https://doi.org/10.1101/gr.122382.111.

Zhang, T. *et al.* (2020) ‘Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells’, *Genome Biology*, 21(1), p. 45. Available at: https://doi.org/10.1186/s13059-020-01957-w.

Zhang, Y. *et al.* (2018) ‘The ZZ domain of p300 mediates specificity of the adjacent HAT domain for histone H3’, *Nature Structural & Molecular Biology*, 25(9), pp. 841–849. Available at: https://doi.org/10.1038/s41594-018-0114-9.

Zhao, D. *et al.* (2017) ‘YEATS Domain-A Histone Acylation Reader in Health and Disease’, *Journal of Molecular Biology*, 429(13), pp. 1994–2002. Available at: https://doi.org/10.1016/j.jmb.2017.03.010.

Zor, T. *et al.* (2004) ‘Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of c-Myb’, *Journal of Molecular Biology*, 337(3), pp. 521–534. Available at: https://doi.org/10.1016/j.jmb.2004.01.038.

Zuber, J. *et al.* (2011) ‘RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia’, *Nature*, 478(7370), pp. 524–528. Available at: https://doi.org/10.1038/nature10334.