University of Sheffield

Too T-ALL for a HAT? Deconstructing genetic regulation at a model enhancer

Nicola Carruthers

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield

Faculty of Science

School of Bioscience

30th June 2023

Acknowlegements

Firstly I would like to thank my supervisor Dan Bose, without whom none of this would have been possible. Thank you Dan for your endless support, encouragement and belief, and for continually pushing me to be a better Scientist and person. I will forever be grateful to have had this opportunity.

Thank you to the Bose lab because you have never failed to cheer me on, offer advice and make me smile. I will never forget the joy, the support and the memories shared with you all. Thank you to Katie Gelder, my lab partner in crime throughout our PhDs. We have been through every step on this journey together and without your wisdom, joy and strength this would have been a completely different experience. Thank you to Laura Harrison, for teaching me to be strong, to believe in myself and for always encouraging me. Thank you to Archna Shah, for your advice, unending kindness and always looking out for me. Thank you to Grace Gilbert, for never failing to encourage me and always keeping me smiling. To Petra Celadova thank you for leading the way. Without your advice and support I would not have become the Scientist I am today. Thank you also to Katy Boswell, who's encouragement and belief in me never failed, and for somehow always knowing what I needed before I did. Many thanks also must go to all of our fourth year students who have shared our lab with us over the years. In particular to Sophie Ball, for her work on the $MYB_{Halotag}$ project. Thank you all for being a pleasure to teach, laugh and commiserate with and for always working so hard.

Thank you to the people with whom we have shared our office and lab spaces who have also shaped my PhD journey. Thank you to Peter Daniels, for being my E21 office buddy and for always being a ray of sunshine. Thank you to Emma Thomson, who's advice and support has always been so appreciated, and I will always be grateful for your keeping an eye on me. Thank you to the Thomson lab, who have always made

Abstract

Enhancers are regions of DNA that act as hubs for protein binding of transcriptional machinery. This machinery can be donated to the promoter of a target gene, upregulating gene expression. Enhancers are bidirectionally transcribed, producing non-coding enhancer RNAs (eRNAs). CREB binding protein (CBP) is a transcriptional coactivator found at most enhancers and upon eRNA binding its histone acetyltransferase (HAT) activity is stimulated (D. A. Bose et al. 2017). In a subset of T-cell acute lymphoblastic leukemia (T-ALL), somatic mutations form a new enhancer that requires the transcription factor (TF) MYB for initiation (Mansour, Abraham, et al. 2014). This enhancer drives oncogenic *TAL1* expression by recruiting CBP and a regulatory complex, consisting of proto-oncogenic TFs, including TAL1 and MYB (Sanda et al. 2012; Mansour, Abraham, et al. 2014).

Increasing evidence is emerging for TFs to have RNA binding properties, and despite the growth in eRNA studies, interactions with eRNAs and protein complexes at enhancers have not been well studied. We sought to understand how CBP and TFs found at this enhancer interact with RNA, to determine a role for regulating RNAs at this enhancer.

We showed that CBP is part of this complex; which we term CBP-TAL1, and individually CBP-TAL1 binds to the *TAL1* eRNAs. To further investigate CBP-TAL1:RNA interactions, we endogenously tagged MYB and CBP in a single cell line using CRISPR/Cas9. Affinity pull-down experiments followed by RNA sequencing, called tandem affinity cross-linking and analysis of cDNAs (TA-CRAC, Granneman, Petfalski, and Tollervey 2011; Thoms et al. 2015) will shed light on the action of RNAs within this protein complex at enhancers, particularly eRNAs. Additionally we developed a system to deplete initiating MYB at this enhancer, allowing investigation of enhancer formation and assembly.

Overall, we have taken complementary approaches to analyse the *TAL1* enhancer, to understand the action of the CBP-TAL1 complex and eRNAs. This will provide deeper insight into eRNA behaviour, regulatory complex action and enhancer function.

Declaration

I, the author, confirm that this Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means. This work has not been previously presented for an award at this, or any other, University.

Contents

List of Figures

14

List of Tables

Nomenclature

3C	Chromosome conformation capture
4C	Chromosome conformation capture on chip
4SU	4-thiouridine
5C	Carbon copy chromosome conformation capture
5mC	Methylated Cytosine on the 5th carbon
AIL	Auto-inhibitory loop
AML	Acute Myeloid Leukemia
ApE®	A plasmid Editor
AR	Androgen receptor
ARID	AT-rich interactive domain
ARID5B	AT-rich interactive domain 5B
ARIEL	ARID5B-inducing enhancer associated long noncoding RNA
ATAC-seq	Assay for transposase accessible chromatin with high-throughput sequencing
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid

bHLH	basic Helix Loop Helix
BRD4	Bromodomain containing protein 4
BSA	Bovine Serum Albumin
CBP	CREB Binding Protein
CBP-FTH	CBP-Flag-TEV-6xHis
CBP-FTHH	CBP-Flag-TEV-6xHis-Halotag
ChIA-Drop	Chromatin interaction analysis via droplet-based and barcode-linked sequencing
ChIP	Chromatin Immunoprecipitation
ChIP-seq	Chromatin Immunoprecipitation followed by sequencing
CIP	Calf Intestinal Phosphatase
CLIP	Crosslinking and immunoprecipitation
CML	Chronic Myeloid Leukemia
Co-IP	Co-Immunoprecipitation
CpG	Cytosine phosphate Guanine
CRISPR	Clustered Regularly Interspaced short Palindromic Repeats
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
Ct	Cycle threshold
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase

DSB	Double Strand Break
dsDNA BR	Double stranded DNA Broad Range
DTT	Dithiothreitol
EDTA	Ethylenediaminetetra-acetic acid
eRNAs	enhancer RNAs
ESCs	Embryonic stem cells
FACS	Fluorescence Activated Cell Sorting
FBS	Fetal Bovine Serum
FISH	Fluorescence in situ hybridisation
gDNA	genomic DNA
GFP	Green Fluorescent Protein
gRNA	guide RNA
GRO-seq	Global run on sequencing
H3K122ac	Histone 3 lysine 122 acetylation
H3K18ac	Histone 3 lysine 18 acetylation
H3K27ac	Histone 3 lysine 27 acetylation
H3K27me	Histone 3 lysine 27 methylation
H3K27me3	Histone 3 lysine 27 trimethylation
H3K4me	Histone 3 lysine 4 methylation
H3K4me1	Histone 3 lysine 4 monomethylation
H3K4me2	Histone 3 lysine 4 dimethylation
H3K4me3	Histone 3 lysine 4 trimethylation

H3K9ac	Histone 3 lysine 9 acetylation
H4K16ac	Histone 4 lysine 16 acetylation
HaloPROTACs	Halotag proteolysis targeting chimeras
HAT	Histone AcetylTransferase
HDR	Homology Directed Repair
HTHAS	6xHis-TEV-HA/Strep
HyPro	Hybridisation proximity labelling
ICE	Inference of CRISPR Edits
IDR	Intrinsically disordered region
indel	Insertion and deletion mutations
IP	Immunoprecipitation
kb	kilobase pairs
KMTs	Lysine methyltransferase
KRAB	Kruppel associated box
LB	Luria Broth
LINE	Long interspersed nuclear element
LN2	Liquid nitrogen
mb	megabase
miRNA	microRNA
MNase	Micrococcal nuclease
MPRA	Massively parallel reporter assay
mRNA	messenger RNA

MYB-HTHAS	MYB-6xHis-TEV-HA/Strep
MYB-HTHAS:CBP-FTH	MYB-6xHis-TEV-HA/Strep:CBP-Flag-TEV-6xHis
ncRNA	non-coding RNA
NHEJ	Non Homologous End Joining
nt	nucleotide
PAM	Protospacer Adjacent Motif
PAR-CLIP	Photoactivatable Ribonucleoside-enhanced Crosslinking and Immunoprecipitation
PBS	Phosphate Buffered Saline
PcG	Polycomb group
PCR	Polymerase Chain Reaction
Pen-strep	Penecillin-streptomycin
PNK	Polynucleotide kinase
PRC2	Polycomb repressive complex 2
PRMTs	Protein arginine methyltransferase
PRO-seq	Precision run-on sequencing
PROTACs	Proteolysis targeting chimeras
PTM	Post-translational modifications
RBR-ID	RNA binding region identification
rcf	relative centrifugal field
RIP	RNA Immunoprecipitation
RISC	RNA induced silencing complex

RNA	Ribonucleic acid
RNA Pol II	RNA Polymerase II
RNA-seq	RNA Sequencing
RNAi	RNA interference
RNP	Ribonucleoprotein
rpm	rotations per minute
RPMI	Roswell Park Memorial Institute
RT-qPCR	Real Time quantitative PCR
RUNX1	Runt related transcription factor 1
SCL	Stem cell leukemia protein
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
sgRNA	single guide RNA
SHAPE-Map	Selective 2'-hydroxyl acylation analysed by primer extension coupled with mutational profiling
Shh	Sonic hedghehog
shRNAs	short hairpin RNAs
siRNA	small interfering RNA
SNPs	Single nucleotide polymorphisms
ssODN	Single stranded oligo-deoxyribonucleotide
STARR-seq	self-transcribing active regulatory region sequencing
T-ALL	T cell acute lymphoblastic leukemia

T2A	Thosea asigna virus 2A protein
T7E1	T7 Endonuclease 1
TAD	Topologically associated domain
TAL1	T cell acute lymphoblastic leukemia protein 1
TBS-T	Tris Buffered Saline with added Tween
TF	Transcription Factor
TIDE	Tracking of Indels by Decomposition
TIDER	Tracking of Indels by Decomposition for HDR
TT-seq	Transient transcriptome sequencing
UCSC	University of California, Santa Cruz
WB	Western blot
YY1	Yin-Yang 1

Chapter 1

Introduction

1.1 Differential gene expression

1.1.1 The coding genome

The central dogma of molecular biology underpins genomic and cell based research (Crick 1970). Our genome is captured within molecules of Deoxyribonucleic acid (DNA), which are then transcribed into Ribonucleic acid (RNA, Jacob and Monod 1961). In turn, these RNA molecules are translated into proteins (Figure 1.1, Brenner, Jacob, and Meselson 1961).
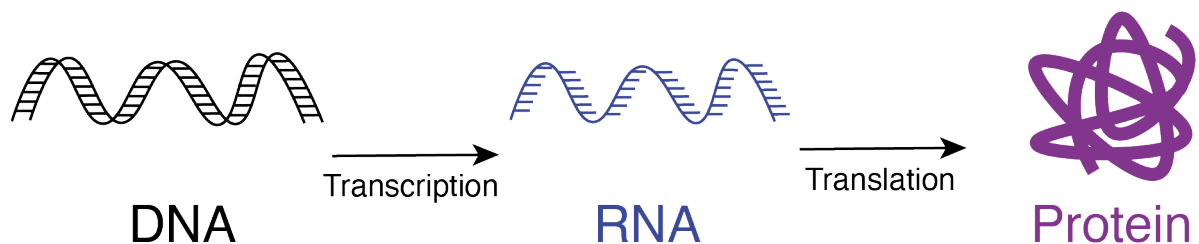
Figure 1.1: The central dogma of molecular biology
Coined by Francis Crick, the central dogma of molecular biology states that DNA is transcribed to make RNA, which in turn is translated to protein (Crick 1970).

1.1.2 Differential gene expression

With the human genome consisting of over 20,000 protein-coding genes, this allows scope for vast diversity in the structure and function of each protein gene product (Harrow et al. 2012). However, in reality, 20,000 genes is not that many to make up an entire human being, with particular regard to the diversity in cell type and function exhibited throughout the body. This raises the question of how our cells are able to have such drastic differences in terms of function and morphology, when each cell contains the same set of genes.

1.1.2.1 Selectivity of messenger RNA

There are many mechanisms by which gene expression is controlled and manipulated to achieve increased diversity in protein expression. When DNA sequences are transcribed into messenger RNA (mRNA), these can be selectively processed and spliced. For example during the removal of non-coding intronic regions differential patterns of exonic regions can form mRNAs to achieve distinct protein products originating from the same encoding gene (reviewed by Marasco and Kornblihtt 2023).

1.1.2.2 Regulation of translation

After being transcribed to mRNA, a further level of selection can take place at the step of translation. In mammalian cells, apart from particular groups of proteins, such as proteins that are secreted by the cell, levels of mRNA and for the final protein product do not correlate well, implying translation of mRNA is highly regulated (Anderson and Seilhamer 1997; Mignone et al. 2002). Untranslated regions (UTRs) of mRNAs can exist either or both at the 5' and the 3' end and can function as an additional step to regulate translation, as specific motifs or secondary structures in the mRNA UTR can be recognised by proteins or other non-coding RNAs for translation or degradation (reviewed by Mignone et al. 2002). mRNA UTR secondary structure can also affect recruitment of the ribosome, and dissolution of the 5' cap by the initiation complex is required prior to translation (reviewed by Livingstone et al. 2010). Altogether this shows that there are many stages of mRNA translation that offer levels of selectivity to control production of proteins.

1.1.2.3 Post-translational modification of proteins can increase protein diversity

Translated proteins can then be differentially post-translationally modified to achieve alternate functions in the cell and further increase the variety of proteins available to the cell (reviewed by Y.-C. Wang, Peterson, and Loring 2014). Some examples include acetylation, methylation, phosphorylation and ubiquitination, which can change the localisation and activity and control interactions with other proteins (reviewed by Mann and Jensen 2003)

1.1.2.4 The non-coding genome

The first investigation into assigning function to the genome as a whole implied that only 1.2% of the genome directly codes for proteins (ENCODE Project Consortium 2012). This highlighted that the vast majority of our DNA does not contain exonic protein coding sequence and therefore there must be large quantities of non-coding sequences and DNA elements that function to aid gene expression, yet do not directly code for protein sequences. In this research project we will focus particularly on how gene expression can be regulated through non-coding regions within the genome.

1.1.3 Chromatin

Within the human genome, DNA wraps around a core of histone proteins to form nucleosomes, that protect the DNA and cluster together to form chromatin (Figure 1.2, reviewed by Wolffe and Guschin 2000). Chromatin is mainly organised into closed chromatin regions, termed heterochromatin, and more open regions known as euchromatin (Figure 1.2, reviewed by Morrison and Thakur 2021). Open euchromatic regions occupy genomic and transcriptionally active regions, as it is necessary for DNA to be accessible to the transcriptional machinery and therefore unwound from the protection of the nucleosome (Figure 1.2, reviewed by Venkatesh and Workman 2015). Hence, heterochromatic regions occupy transcriptionally silent regions, where transcriptional machinery is not required to access DNA sequences, so DNA is wound around histones forming nucleosome dense regions (reviewed by Morrison and Thakur 2021).
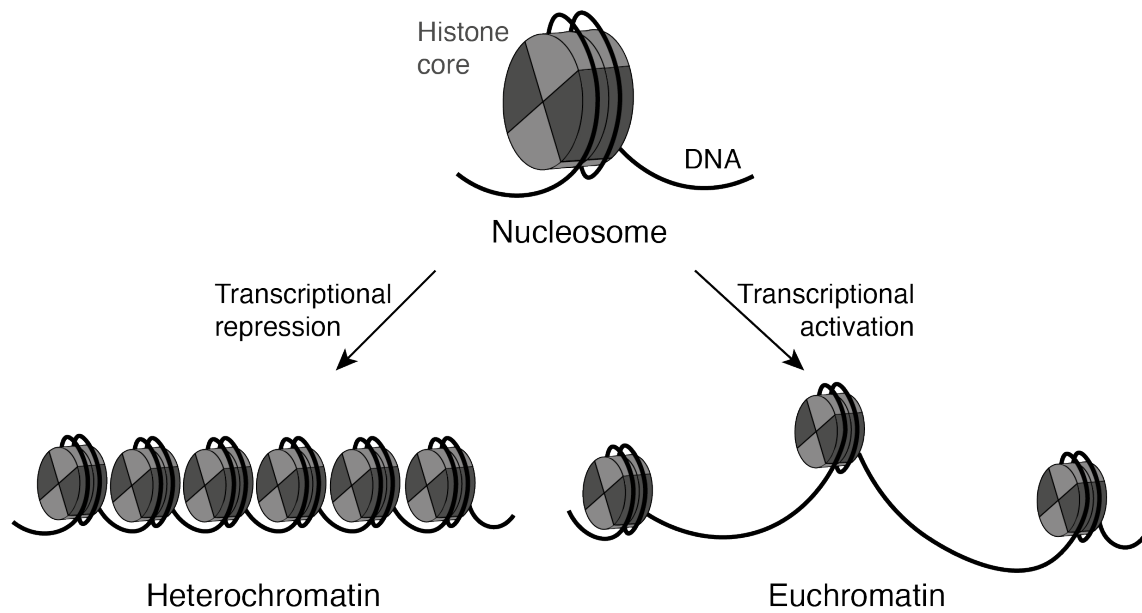
Figure 1.2: Nucleosomes are organised into heterochromatin and euchromatin DNA wraps around histone cores to form nucleosomes. In transcriptionally repressive regions, nucleosomes cluster densely together, forming regions of heterochromatin. In transcriptionally active regions of the genome, nucleosomes are less densely clustered, and this forms regions of euchromatin.

This opening and closing of chromatin can be dynamic, and post-translational modifications of DNA and the histone proteins can be used to either drive or maintain chromatin status (reviewed by Karolin Luger, Dechassa, and Tremethick 2012 and Lai and Pugh 2017). In addition, variants of the core histone proteins found within nucleosomes can be varied to favour either activating or repressive chromatin states (reviewed by Martire and Banaszynski 2020).

The discovery of these modifications have helped to form the field of epigenetic research: investigating how changes to chromatin can enable changes in gene expression, without manipulating the DNA nucleotide sequence itself.

1.1.4   DNA methylation

The addition of methyl groups to DNA, particularly to cytosine nucleotides, on the fifth carbon (5mC), are often found preceding a guanine (CpG) (reviewed by Greenberg and Bourc'his 2019). Methylation of DNA can be associated with repression of

transcription, for example the mechanism for inactivation of the X chromosome utilises DNA methylation (Mohandas, Sparkes, and Shapiro 1981). Methylation of promoters and the first exon of a gene is associated with repression of gene expression, however methylation found downstream, within the body of the gene, is associated with an increase in transcription in actively dividing cells (reviewed by Moore, Le, and Fan 2013).

DNA methyltransferase (DNMT) enzymes are one example of a group of enzymes that deposit methyl groups onto cytosine residues (reviewed by Greenberg and Bourc'his 2019). At methylated cytosine nucleotides, DNMTs have been shown to recruit histone modifying enzymes and can induce both active or repressive modifications onto the histones depending upon the transcriptional environment (reviewed by Moore, Le, and Fan 2013). Therefore, DNA methylation consists of a complex interplay within the local gene environment, the cell type and function, and is utilised to aid in differential gene expression.

1.1.5 Histone modifications

The histone core within a nucleosome is an octamer, consisting of two H2A/H2B heterodimers alongside a tetramer containing two H3 and H4 proteins (Figure 1.3A, Luger et al. 1997). The first evidence for post-translational modifications (PTMs) of histones was identified almost 60 years ago (Allfrey, Faulkner, and Mirsky 1964), and a variety of modifications have since been described upon different amino acids. Histones can be subjected to PTMs, both within the histone core and upon the N and C-terminal histone tails (reviewed by Tessarz and Tony Kouzarides 2014). The N-terminal tail of histone proteins within the octamer core can protrude from the nucleosome and can vary from between 15-37 residues in length across the four histone proteins (Luger et al. 1997). N-terminal histone tails are highly enriched for amino acids that can be modified: particularly lysines and arginines, but also serines, threonines, glutamines and tyrosines (reviewed by Millán-Zambrano et al. 2022). There are many types of PTMs deposited onto histones, but these include acetylation, methylation, phosphorylation, ubiquitylation and sumoylation (reviewed by Andrew J Bannister and Tony Kouzarides 2011).

These PTMs on histones are deposited, recognised and removed by particular

groups of enzymes. *Writers* are proteins that function to add PTMs to specific residues on histones, which in turn are interpreted by *Readers* (Figure 1.3B and C respectively). To complete the toolkit, *Eraser* enzymes are able to get rid of histone marks, enabling different histone marks to be deposited (Figure 1.3D, reviewed by Millán-Zambrano et al. 2022).

A

H3 H4 tetramer

Histone tails

H2A/H2B heterodimer

Nucleosome

B

ac/me

HATs and HMTs

Writers

C

Recruitment of binding partners and downstream effects

Readers

D

HDACs and HDMs

Erasers
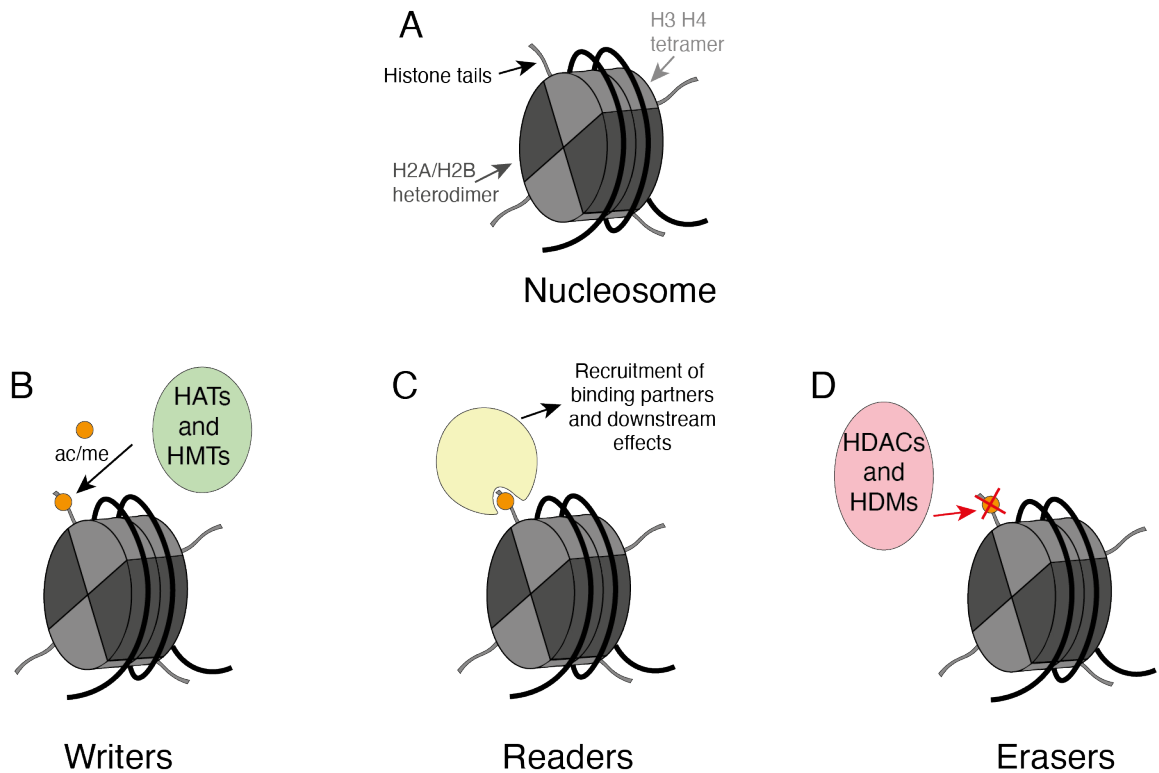
Figure 1.3: Histones are subject to post-translational modifications
(A) The nucleosome comprises of two H2A/H2B heterodimers (dark grey) and a H3/H4 tetramer (light grey). Histone tails protrude from the nucleosome and histone core and can be subject to post-translational modifications (PTMs). (B) Writer enzymes (green) can deposit histone PTMs, such as acetylation (ac) or methylation (me) onto the histone tails. (C) Reader enzymes (yellow) can recognise the histone modifications and can cause downstream events, such as recruitment of binding proteins. (D) Eraser enzymes (red) remove histone PTMs.

As well as PTMs, the core histone proteins can be exchanged for modified histone proteins, that can aid in manipulating the local chromatin environment. These histone variants can have small or large changes, from just a few amino acids to addition of

domains (reviewed by Venkatesh and Workman 2015).

This highlights the dynamic nature of chromatin, and how PTMs and histone variants are used by the local chromatin environment to fine tune and to precisely turn gene expression on and off. This adds further layers to how the cell is able to manipulate its own differential gene expression patterns.

1.1.6   Histone methylation

In addition to methyl groups being deposited upon DNA, particular amino acid residues can be methylated within histones post-translationally. Methylation is mainly found upon arginine and lysine residues on histones, where it is possible that more than one methyl group can be deposited upon these amino acids (reviewed by Millán-Zambrano et al. 2022). Differential methylation upon alternative residues of the histone tails, or with regard to the number of methyl groups upon the same residue can affect the local chromatin and transcriptional environment. Histone methylation modifications have been shown to both activate and repress transcription, depending on the residue and the local environment (reviewed by Hyun et al. 2017; Millán-Zambrano et al. 2022).

Methylation upon lysine residues is deposited by classes of lysine methyltransferases (KMTs, reviewed by Hyun et al. 2017) and protein arginine methyltransferases (PRMTs) are responsible for methylating arginines (reviewed by J. Zhang et al. 2019).

1.1.6.1   H3K4me1/2/3

One example of histone methylation is histone 3 lysine 4 methylation (H3K4me), which is generally associated with transcriptional activation (reviewed by Black, Van Rechem, and Whetstine 2012). The number of methyl groups present on this particular lysine residue can mark specific regulatory elements and can have different functions through intergenic and intragenic regions.

Histone 3 lysine 4 monomethylation (H3K4me1) appears to be a more general characteristic of open and active chromatin regions, which can be found at promoters and other *cis*-regulatory elements (Rada-Iglesias et al. 2011), for example at enhancers (described later in detail).

Two methyl groups present at histone 3 lysine 4 (H3K4me2) have been observed at different regions of the gene and regulatory regions, depending upon the gene being analysed (Pekowska et al. 2010). H3K4me2 can be enriched at promoters, but can also be found at high levels within the gene body in genes associated with tissue specificity (Pekowska et al. 2010). A role has been suggested for H3K4me2 to retain epigenetic memory of important genes during final differentiation steps from progenitor cells, for example in the differentiation of adult smooth muscle cells (M. Liu et al. 2021)

Trimethylation of lysine 4 on histone 3 (H3K4me3) is localised to transcriptional start sites (TSS) and promoter regions, and is therefore often used as a characteristic of such features to define promoters (reviewed by Millán-Zambrano et al. 2022). H3K4me3 has been shown to aid in the recruitment of RNA Polymerase II (RNA Pol II), therefore initiating transcription (Vermeulen et al. 2007).

1.1.6.2 H3K27me3

On the other hand, histone methylation PTMs can be associated with transcriptional repression. Histone 3 lysine 27 methylation (H3K27me), particularly trimethylation of H3K27 (H3K27me3), is a marker of heterochromatin and is important in the repression of gene transcription (reviewed by Millán-Zambrano et al. 2022). After deposition of H3K27me3 modifications, these can be spread along nucleosomes, helping to further silence the local chromatin environment (reviewed by Hyun et al. 2017).

1.1.6.3 Histone methylation affects the local chromatin environment

These examples of H3K4me1/2/3 and H3K27me3 highlight that methylation can correlate with both repressive or activating transcriptional regions. Further to this, the number of methyl groups deposited on a particular residue can change depending upon the local chromatin environment, and can be used as markers for regulatory elements within the gene environment (reviewed by Hyun et al. 2017).

1.1.7 Histone acetylation

An alternative histone PTM is the deposition of acetyl groups, where acetylation of histones is associated with transcriptionally active regions of chromatin (reviewed by

Millán-Zambrano et al. 2022). As lysine residues are positive, the addition of an acetyl group neutralises this charge and as a result causes the nucleosomes to distance from each other, opening the chromatin and therefore making DNA more accessible for transcriptional machinery (reviewed by Millán-Zambrano et al. 2022).

1.1.7.1 Lysine acetylation in histone tails

Within the N-terminal tails of histones, amino acids can be acetylated; particularly lysine residues. Due to the high number of lysines that are found within histone tails, this can give rise to many different patterns of acetylated residues (reviewed by Ghoneim, Fuchs, and Musselman 2021). Acetylation modifications generally correlate with active gene expression and regions of transcriptional activation, however the pattern of the amino acids that are acetylated can vary depending on the genomic environment (Zhibin Wang et al. 2008).

1.1.7.2 Histone acetyltransferases

Acetylation PTMs can also vary in that they can be deposited onto histones by different histone acetyltransferases (HATs). There are a number of different HATs within the genome, that can acetylate histones to result in slightly different functions. There are four families of HATs that vary slightly in function and substrate preferences, that comprise of GCN5/PCAF, CBP/p300, the MYST family and the nuclear receptor family (Bedford et al. 2010).

The paralogues p300 and cAMP Response Element (CREB) binding protein (CBP) are both HATs that act to increase transcriptional activation (Ogryzko et al. 1996 and A J Bannister and Kouzarides 1996, respectively). CBP and p300 have been observed to have distinct substrates and functions within the genome, compared to other acetyltransferase families, and aid in increasing the transcriptional activation of the genomic environment (Q. Jin et al. 2011). CBP and p300 are also transcriptional coactivators, whereby they have the ability to acetylate transcription factors to help upregulate gene transcription, for example p300 and CBP were implicated in acetylation of the tumour suppressor protein p53 (Gu and R G Roeder 1997).

1.1.7.3 H3K9ac

Histone 3 lysine 9 acetylation (H3K9ac) has been shown to mainly localise to the transcriptional start site (TSS) of active genes (Zhibin Wang et al. 2008). H3K9ac is mainly deposited by GCN5 and knockout of GCN5 has been shown to specifically reduce H3K9ac modifications at promoters (Q. Jin et al. 2011). However, this reduction in H3K9ac did not affect gene expression or recruitment of RNA Polymerase II (RNA Pol II) to the promoter, indicating that H3K9ac is not directly required for gene expression, and may be correlative with active gene expression (Q. Jin et al. 2011).

1.1.7.4 H3K18ac

Histone 3 lysine 18 acetylation (H3K18ac) is mostly deposited by CBP and p300, as deletion of these proteins saw drastic reductions in H3K18ac (Q. Jin et al. 2011). H3K18ac modifications are also often found at the TSS of genes, however H3K18ac modifications appear to be deposited at an earlier stage in transcriptional activation, prior to the recruitment of RNA Pol II (Q. Jin et al. 2011).

1.1.7.5 H3K27ac

Acetylation of lysine 27 on histone 3 (H3K27ac) is deposited by both CBP and p300 and is a marker of active transcriptional regions (Q. Jin et al. 2011). H3K27ac is classically used as a marker of *cis*-regulatory elements known as enhancers (described later in more detail), and can be used to distinguish between different developmental and transcriptional states at such regions (Creyghton et al. 2010).

1.1.7.6 H3K122ac

Histone 3 lysine 122 acetylation (H3K122ac) is an example of a PTM that is deposited within the histone core, rather than the histone tails, and is found at a region of histone 3 that is very important for histones to interact with DNA to form the nucleosome (Luger et al. 1997; Hall et al. 2009). Therefore, it is likely that acetylation of H3K122 disrupts the DNA-histone interactions, causing DNA to be less tightly wound around the nucleosomes leading to a chromatin environment where nucleosomes are more open and less densely clustered (Tropberger et al. 2013).

H3K122ac is deposited by CBP and p300, and has also been shown to be a characteristic marker of enhancer regions (described later in more detail, Tropberger et al. 2013; Pradeepa et al. 2016). When H3K122ac is found within chromatin *in vitro*, this has been shown to be sufficient for stimulation of *in vitro* transcription, compared to non-acetylated H3K122 where transcription was much reduced (Tropberger et al. 2013).

1.1.7.7   H4K16ac

Acetylation of lysine 16 on histone 4 (H4K16ac) is one of four lysine residues that are able to be acetylated on the H4 N-terminal tail (Dion et al. 2005). However, studies in yeast have shown that without acetylation of K16 in particular, gene expression was downregulated, suggesting H4K16ac has the largest contribution to transcriptional activation compared to the other lysine residues in the H4 tail (Dion et al. 2005). H4K16ac has also been shown to prevent complete chromatin condensation *in vitro* (Shogren-Knaak et al. 2006), which has been explained due to the loss of the positive charge on K16 no longer being able to interact with a negatively charged region within the H2A/H2B heterodimer, therefore preventing interactions between nucleosomes (Luger et al. 1997; Robinson et al. 2008).

H4K16ac is deposited in mammalian cells by the HAT KAT8 of the MYST family (Taylor et al. 2013) and has been mainly associated with transcriptionally active enhancer regions (Pal et al. 2023, described later in detail).

1.1.7.8   Histone crotonylation

There are also examples of other acyl groups, similar to acetyl groups that can be deposited onto histones and influence gene expression. Crotonyl groups are larger than acetyl groups and can also be deposited onto histones by CBP and p300 (Sabari, Tang, et al. 2015). H3K18cr has been implied to activate gene transcription through localising to transcriptional start sites (Sabari, Tang, et al. 2015).

1.1.7.9   "The histone code" is not quite as simple

Traditionally the complex interplay of the range of modifications has been termed as "the histone code", implying that specific groups of histone modifications dictate

the environment that each modification is found within (Strahl and Allis 2000). To allow a better understanding of the role histone PTMs play within chromatin regions, in recent years the techniques available to study histone modifications, coupled with their genomic environment and the transcriptional activity of the region has vastly increased. For example, mass spectrometry has been to developed to include cross-linking (O'Reilly and Rappsilber 2018), alongside the development of genome-wide approaches such as Chromatin-Immunoprecipitation (ChIP, T. H. Kim and Ren 2006; Johnson et al. 2007), Cleavage Under Targets & Release Under Nuclease (CUT&RUN, Skene and Henikoff 2017) and Cleave Under Targets & Tagmentation (CUT&Tag, Henikoff et al. 2020). In one study, due to the correlation between transcription and histone modifications, a pipeline was established to use nascent RNA sequencing data, representing transcription, to generate predictive patterns of many different histone modifications correctly at many different chromatin environments (Zhong Wang et al. 2022). However, this still raises the question of whether histone modifications and the transcriptional environment are causative or correlative.

Further studies continue to add evidence that histone modifications reflect a complex dialogue with the local chromatin environment, where in some cases histone PTMs can arise as a result of the changes to the transcriptional environment itself, and in other cases they can be correlative. Therefore, putting histone modifications in boxes to truly define chromatin regions can be misleading, whereas using these modifications as individual pieces of a puzzle to help visualise a whole picture can be a lot more informative.

1.2 Enhancers

Enhancers are part of the non-coding genome, and are *cis*-acting regulatory DNA elements, meaning they act to increase transcriptional activity on targets found within the same strand of DNA (reviewed by Heinz et al. 2015; Long, Prescott, and Wysocka 2016). Each enhancer element is able to regulate transcription of its target promoter, irrespective of being located up or downstream of the promoter, with these interactions being able to occur over vast distances (Long, Prescott, and Wysocka 2016).

The first enhancer was discovered as a DNA region distal from the β-*Globin* gene, that although at the time was not identified as being transcribed itself, its presence was necessary for high levels of β-Globin transcription (Banerji, Rusconi, and Schaffner 1981). Since that first discovery, enhancers have been shown to act on target promoters that can lie between a hundred bases away to over multiple megabases away (reviewed by Heinz et al. 2015). The first long range enhancer was of the Sonic hedgehog (*Shh*) gene, where the distal enhancer was located 1 megabase (mb) away from the TSS (Lettice et al. 2003).

In order to help define these regulatory elements and to enable identification of putative regulatory elements and mapping of enhancers genome wide, there are specific features that have classically been assigned to enhancers. These include transcription factor and coactivator binding, such as CBP, patterns of histone modifications, DNase I hypersensitivity and the transcription of non-coding enhancer RNAs (Figure 1.6, described later in detail).

There are however limitations to the model of having a strict set of characteristics that each enhancer must display in order to be classed as an active enhancer, as it has been well documented that enhancers are likely to exhibit most, but not all of these features. Therefore, using a combination of these characteristics can be helpful for defining enhancer regions, however the lack of a single feature should not automatically prevent a putative element from being classed as an enhancer. This does present challenges in the search for putative enhancers, however it does provide more flexibility in the definition of these regulatory elements.

1.2.1 Enhancers are hubs for protein binding

Enhancer elements contain high concentrations of protein binding sites within a short stretch of sequence; classically between 100 to 1000 bp (Figure 1.4A, Long, Prescott, and Wysocka 2016). Transcription factors (TFs) are proteins that bind to short 6-12 bp DNA motifs and influence transcriptional activity (reviewed by Lambert et al. 2018) and enhancers comprise many binding sites for such TFs. TF binding at enhancers can recruit transcriptional coactivators, which are large proteins or whole complexes that have distinct roles in activating transcription (Figure 1.4B and C, reviewed by Lambert et al. 2018), for example recruitment of the coactivating Mediator complex recruits RNA Pol II (Kagey et al. 2010).
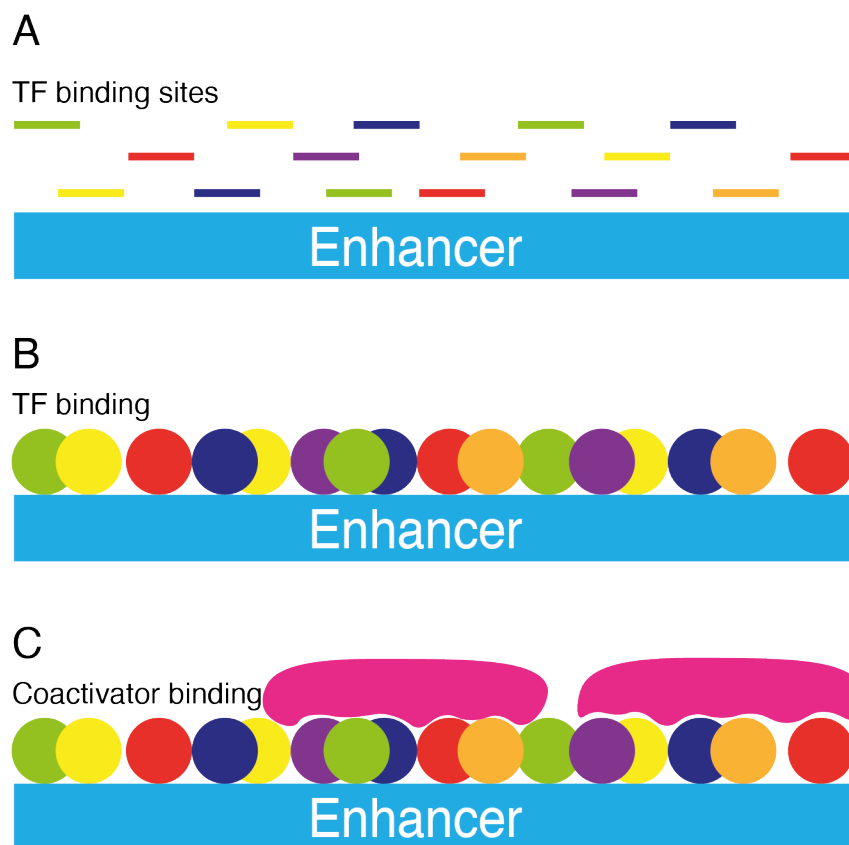
Figure 1.4: Clusters of protein binding motifs are located within enhancers
(A) Many binding sites for different transcription factors (TFs) are found within enhancer sequences. (B) TFs are able to bind to their binding site and cluster together to form active enhancers. (C) Bound TFs are then able to recruit coactivators (pink).

1.2.1.1 Enhancers donate transcriptional machinery to the target promoter

After recruiting the transcriptional machinery, through interactions between bound TFs and transcriptional coactivators, this collective of transcriptional machinery is then able to be donated to the target promoter, aiding activation of transcription of the target gene (Figure 1.5, reviewed by Furlong and Levine 2018).
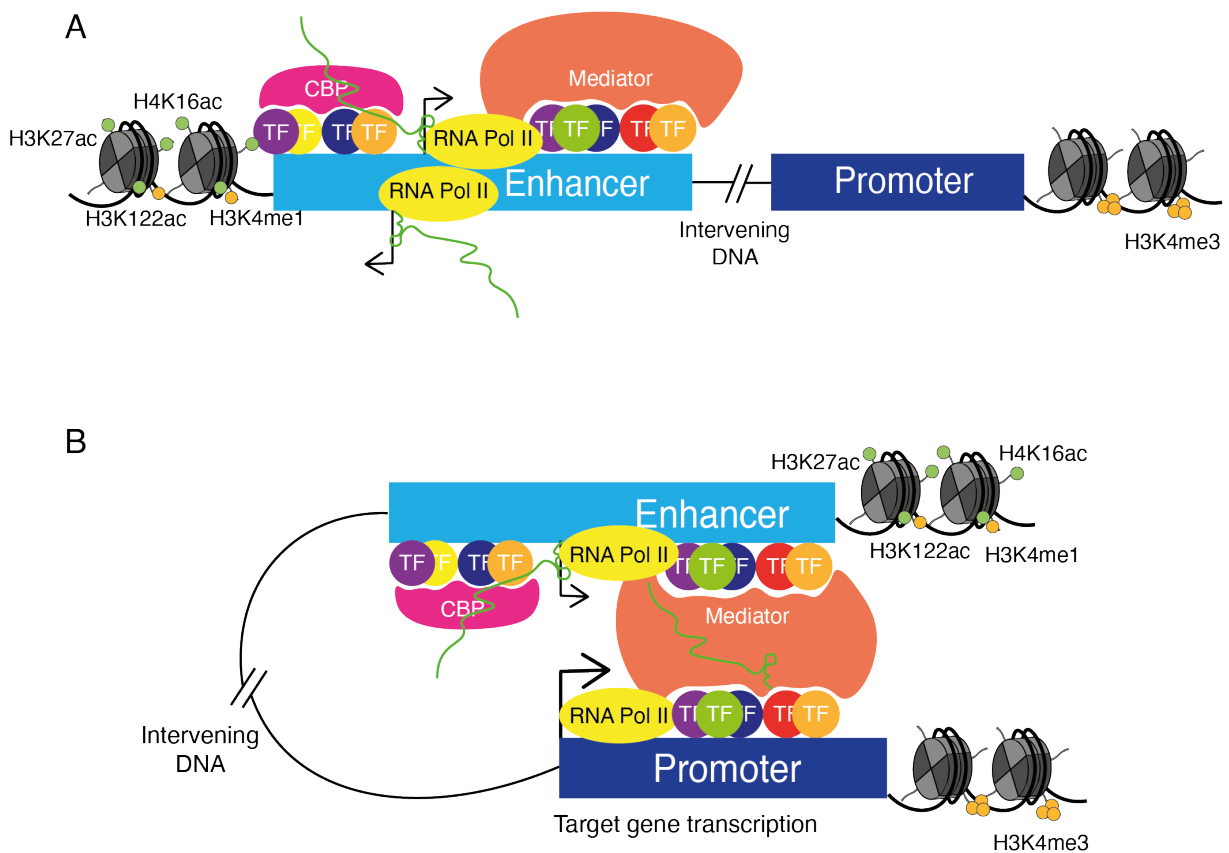
Figure 1.5: Enhancers donate transcriptional machinery to target promoters (A) Transcriptional machinery including transcription factors (TFs), coactivators such as CBP and Mediator and RNA Pol II assemble at enhancers. enhancer RNAs (eRNAs) (green) are transcribed due to RNA Pol II binding. (B) Enhancer-promoter interactions enable target machinery to reach the target promoter, activating target gene transcription.

1.2.2 Chromatin accessibility

Enhancers are DNase I hypersensitive as these regulatory regions are depleted of nucleosomes in order to allow the binding of transcription factors and transcriptional coactivators. Therefore, DNA is more exposed and is not protected through being tightly bound to histone proteins, allowing DNase I to be particularly effective at such sites (Figure 1.6, Malin, Aniba, and Hannenhalli 2013). Through the determination of areas that are highly susceptible to degradation by this enzyme, it can allow the identification of putative regulatory elements (Gross and Garrard 1988).

Accessible chromatin regions can be mapped using assay for transposase accessible chromatin with high-throughput sequencing (ATAC-seq, Buenrostro et al. 2015). Use of a Tn5 transposase allows sequencing adapters to be inserted directly into accessible DNA sequences within chromatin, reducing the number of steps within the protocol compared to other techniques such as treating with DNase or Micrococcal nuclease (MNase) prior to sequencing (Buenrostro et al. 2015).

1.2.3 Enhancers can be defined by histone modifications

A series of histone modifications are found at enhancers, which have classically been used to identify and classify the type of regulatory element and its state of activation. There are limitations to defining enhancers as requiring a full complement of histone modifications, or by only examining histone modifications and not taking into account other characteristics. Further examples continue to be elucidated whereby different types of enhancers exhibit different patterns of modifications, for example active compared to poised enhancers (described later in detail), and differential chromatin environments may correlate with or cause different patterning of enhancers.
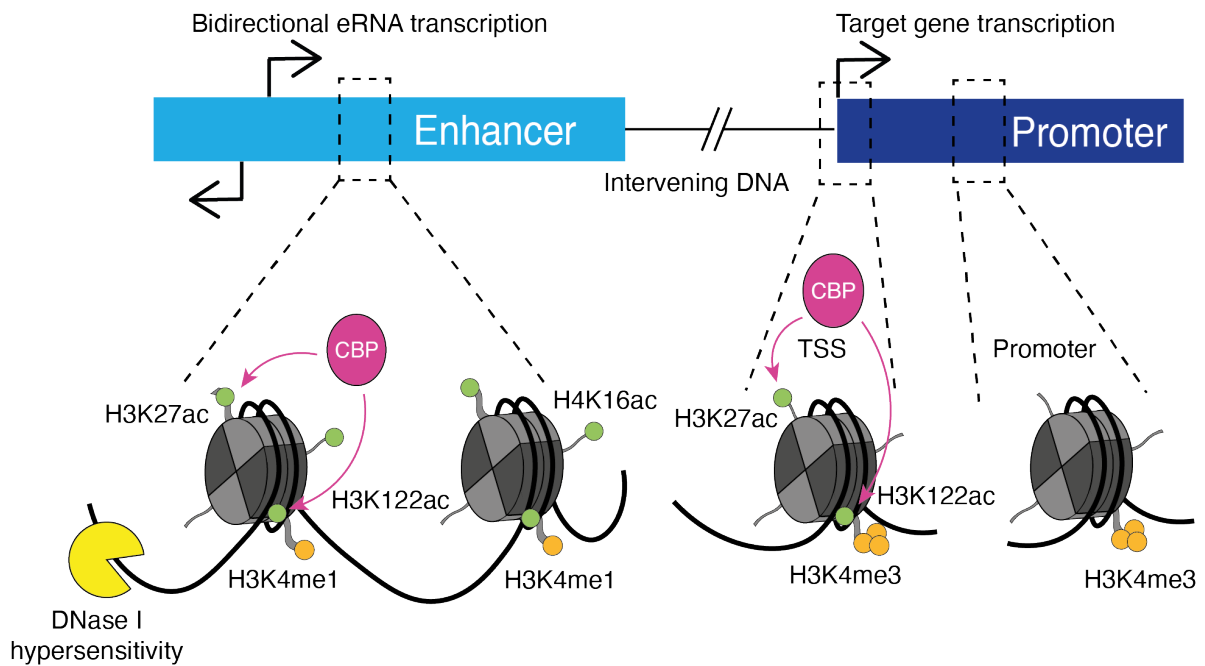
Figure 1.6: Enhancers have specific signatures

DNA within nucleosomes is hypersensitive to DNase I digestion, as the DNA is more freely accessible. Histones within enhancers have H3K27ac and H3K122ac modifications deposited by CBP, H4K16ac and H3K4me1 modifications. The transcriptional start site (TSS) has H3K27ac and H3K122ac deposited by CBP and H3K4me3. The promoters mainly exhibit H3K4me3 PTMs. The H3K27ac at enhancers is deposited by CBP, and CBP binding is a further marker of enhancer activity. Enhancers are bidirectionally transcribed into non-coding enhancer RNAs (eRNAs), where target genes are transcribed to mRNA.

1.2.3.1 H3K4me1

High levels of monomethylation of lysine 4 on histone 3 (H3K4me1) when found alongside low levels of trimethylation of the same residue (H3K4me3) indicate an active enhancer, whereas high H3K4me3 levels often mark a promoter (Figure 1.6, Heintzman et al. 2007). H3K4me1 is not only found at enhancers, so areas of the genome with these PTMs may only be cautiously labelled as true enhancers. But, in combination with other defining factors it can allow the identification of enhancers across the genome (Calo and Wysocka 2013).

1.2.3.2 H3K27ac

A further histone modification used in the identification of enhancers is H3K27ac, which has been implicated as a key marker of active enhancer activity (Figure 1.6, Creyghton et al. 2010). H3K27ac is indicative of open chromatin regions, allowing space for the transcriptional machinery to bind and influence transcriptional activity (reviewed by Verdin and Ott 2015). H3K27ac at enhancers aids in the donation of the transcriptional machinery to the target promoter, as loss of H3K27ac has been shown to reduce interactions between enhancers and promoters (Sungalee et al. 2021).

H3K27ac PTMs are deposited at enhancers by CBP and p300, which are also used as markers of active enhancer activity (A J Bannister and Kouzarides 1996; Ogryzko et al. 1996; Creyghton et al. 2010).

1.2.3.3 H3K122ac

H3K122ac has been shown to generally correlate with H3K27ac and CBP and p300 binding, but does not correlate as well with H3K4me1 (Tropberger et al. 2013). Although both H3K27ac and H3K122ac have been shown to be markers of active enhancer regions, they do not always co-localise and it has been shown that differences between the two histone PTMs correlate with differential transcription factor binding sites (Tropberger et al. 2013). For example, GATA3 binding sites were more likely to be modified with H3K27ac and reduced H3K122ac, whereas enhancers with binding sites for c-Myc were more likely to exhibit H3K122ac (Tropberger et al. 2013).

1.2.3.4 H4K16ac

H4K16ac is also increasingly being used as a marker of transcriptionally active enhancers (Taylor et al. 2013). At a majority of enhancers investigated in mouse embryonic stem cells, H3K27ac and H4K16ac levels correlated, however there were group of active enhancers that exhibited either H3K27ac or H4K16ac modifications, highlighting that enhancers can be active yet only exhibit one of these modifications (Taylor et al. 2013).

In human embryonic stem cells, H4K16ac was shown to be deposited at the L1 type of DNA sequences termed as long interspersed nuclear element (LINE, Pal et al. 2023). These are transposable elements, where through being transcribed into an

mRNA and joining with a group of proteins, these sequences can move and insert themselves at alternative regions within the genome (Ewing and Kazazian 2011). H4K16ac has been shown to be enriched at the 5' UTRs of these L1 elements, and this was shown in multiple cell types (Pal et al. 2023). Upon further investigation into these 5' UTRs, enhancer characteristics and functions were exhibited and where H4K16ac was deposited at such L1 enhancer elements, increased enhancer-promoter interactions were observed (Pal et al. 2023). Therefore when H4K16ac is deposited at such L1 elements, it correlates with enhancer activity of such regions, aiding increased gene expression (Pal et al. 2023).

1.2.4 Binding of CBP is a marker of enhancers

1.2.4.1 CBP and p300

CBP and its paralogue p300 have both been shown to be responsible for deposition of H3K27ac and H3K122ac PTMs at transcriptionally active enhancers (Tie et al. 2009; Creyghton et al. 2010; Q. Jin et al. 2011). As CBP and p300 are paralogues and have many overlapping functions, often in the literature CBP and p300 are used interchangeably, particularly when describing markers of enhancer activity. However, both CBP and p300 have been shown to be embryonic lethal, meaning when each protein was knocked out during mouse development, the resulting embryos did not survive to birth (Yao et al. 1998; Kung et al. 2000). This highlighted that both proteins must have different functions that are necessary for development, and one cannot completely compensate for the other meaning caution must be taken when terming these proteins interchangeably (reviewed by Kalkhoven 2004).

1.2.4.2 CBP at enhancers

CBP contains seven structured domains, with four domains being key for protein-protein interactions: KIX, TAZ1, TAZ2 and ZZ (Zeng, Q. Zhang, et al. 2008). This enables the CBP interactome to contain over 400 proteins (Bedford et al. 2010). Upon binding of TFs to binding motifs clustered at enhancers, it is therefore very likely that at least one of these bound TFs will bind to CBP, resulting in recruitment of CBP to the enhancer (Figure 1.4). Consequently, CBP is found at almost all enhancers

throughout the genome and is used as a marker of enhancer regions (Figure 1.6, Holmqvist and Mannervik 2013).

CBP in addition to being an acetyltransferase, has the separate ability to scaffold proteins. This means that CBP can help group and hold together specific proteins, and this can be utilised at enhancers to scaffold transcriptional machinery in one place, to help drive transcriptional activation (reviewed by Chan and La Thangue 2001).

1.2.5 Enhancers are transcribed into non-coding enhancer RNAs

A further key feature of active enhancers, is that they themselves are bidirectionally transcribed by RNA Pol II, producing enhancer RNAs (eRNAs, De Santa et al. 2010; T.-K. Kim et al. 2010, Figure 1.5). These non-coding transcripts are generally classed as being of 200-2,000 bp in length, non-polyadenylated and short-lived within the cell (Sartorelli and Lauberth 2020; Harrison and D. Bose 2022).

The full extent of the role of eRNAs at enhancers and with transcription is yet to be elucidated, however many functions for eRNAs have been proposed.

1.2.5.1 Nascent RNA-sequencing can identify eRNAs

When the first β-*Globin* enhancer was identified, a key feature of this element used to define it as an enhancer was that it was not itself transcribed into mRNA (Banerji, Rusconi, and Schaffner 1981). As eRNAs are prone to degradation, they are difficult to observe, extract and sequence, therefore when this first enhancer was identified, there was not the technology available to define the transcription of enhancer RNAs (Banerji, Rusconi, and Schaffner 1981).

The development of nascent RNA sequencing has allowed the capture of RNA populations that are unstable and newly transcribed, which has been used particularly in the dicovery of non-coding RNA species. The first nascent RNA sequencing, termed global run-on sequencing (GRO-seq) captured newly transcribed transcripts (Core, Waterfall, and Lis 2008) and high throughput RNA sequencing was able to show some of the first evidence for the bidirectional transcription of eRNAs (T.-K. Kim et al. 2010).

Since then, GRO-seq has evolved, with precision run-on sequencing (PRO-seq) offering increased resolution (Mahat et al. 2016), and the development of transient

transcriptome sequencing (TT-seq) uses 4-thiouridine (4SU) to label newly transcribed RNAs for sequencing (Schwalb et al. 2016).

Both a sense and antisense eRNA have now been reported to be transcribed from enhancers of the β-Globin gene, and this has been shown through RT-qPCR and fluorescence in situ hybridisation (FISH), which showed co-localisation of the β-Globin gene and the β-Globin eRNAs (Gurumurthy et al. 2021).

1.2.5.2   The act of enhancer transcription may play a regulatory role

There is evidence that the act of transcription itself at enhancers may be important for transcriptional activation, as well as the output of the eRNA transcripts themselves (Kaikkonen et al. 2013). This study linked eRNA transcription to H3K4 methylation at *de novo* enhancers and suggested that the transcription of eRNAs by RNA Pol II aids in the recruitment of HMTs, thus further activating the local chromatin environment (Kaikkonen et al. 2013).

1.2.5.3   eRNAs can regulate H3K27ac activity of CBP

The HAT domain, key to the HAT activity of CBP contains an activation loop, sometimes termed the auto-inhibitory loop (AIL) that is necessary for acetylation (Thompson et al. 2004; D. A. Bose et al. 2017). The activation loop can be found within two conformations: inactive where the AIL binds in the active site and there is no ability to acetylate, or active where the AIL is displaced from the active site, allowing substrates to bind and acetylation PTMs can be deposited onto histones and further substrates (Thompson et al. 2004; D. A. Bose et al. 2017). eRNAs can bind to the activation loop, displacing it from the active site and stimulating catalytic acetyltransferase activity (D. A. Bose et al. 2017). Therefore, eRNAs play an important role in the deposition of activating histone modifications and increasing the transcriptional environment at enhancers (Figure 1.7).

44

CBP   **eRNA**

RNA polymerase II

ac   ac

Histone acetylation
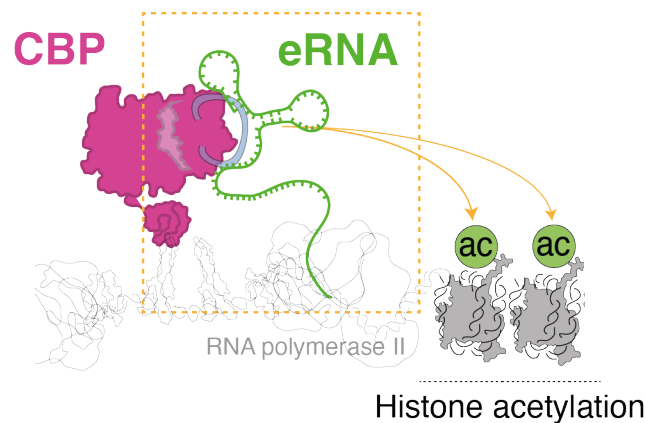
Figure 1.7: eRNA binding stimulates CBP acetyltransferase activity
eRNAs transcribed from enhancers by RNA Polymerase II, bind to the activation loop
of CBP. This induces a conformational change of CBP, causing the activation loop to
swing out, enabling histone acetyltransferase activity by CBP. Figure by Dan Bose.

1.2.5.4   eRNAs bind to BRD4

Bromodomain containing protein 4 (BRD4) is a histone reader protein, and has been
shown to bind to acetylated histones, including H3K27ac (Zeng and M. M. Zhou 2002).
Upon BRD4 binding to acetylated lysines, such as H3K27ac modifications, eRNAs
can also bind, strengthening these interactions between BRD4 and the histone tails
(Rahnamoun et al. 2018). Binding of eRNAs help to retain RNA Pol II at enhancers
and thus increases the activation of the local transcriptional environment (Rahnamoun
et al. 2018).

1.2.5.5   eRNAs aid in enhancer-promoter interactions

Transcripts produced from enhancers have been shown to have important functions
at target promoters (Mousavi et al. 2013). When examining the *MYOD1* promoter,
knockdown of an eRNA transcribed from a regulatory enhancer reduced RNA Pol II
localisation to the *MYOD1* TSS (Mousavi et al. 2013).

In an additional context, an eRNA has also been implicated in enhancer-promoter
interactions in the androgen-receptor (AR) pathway (Hsieh et al. 2014). The eRNA
KLK3e is required for the formation of a complex with Mediator and AR, and the role
of this complex is to activate transcription at target promoters (Hsieh et al. 2014).

These examples imply that eRNAs may play a role in assembling or directing transcriptional machinery from the enhancer to the target promoter (Figure 1.5).

1.2.5.6 Transcription factors interact with eRNAs at enhancers

Additionally eRNAs may have a role in aiding binding of transcription factors to enhancer regions. For example, the transcription factor Yin-Yang 1 (YY1) has been shown to bind to both enhancers and the transcribed eRNAs from that enhancer (Sigova et al. 2015). Reduction of eRNA transcription reduced YY1 binding to that enhancer, suggesting eRNAs may be acting to tether transcription factors to enhancers. This role for eRNAs has been termed "transcription factor trapping" (Sigova et al. 2015).

1.2.6 Assessment of enhancer activity

Using the characteristics often attributed to enhancers, combinations of ATAC-seq, ChIP-seq for CBP and activating histone modifications alongside nascent RNA sequencing can identify enhancer candidate regions of interest. These characteristics are used to define the extent of the activity of the enhancer, for example an increased region of H3K27ac can indicate a super-enhancer element (Hnisz, Abraham, et al. 2013, described later in more detail). However, these characteristics cannot imply the extent to which target gene expression is controlled and manipulated by a particular enhancer of interest. To study the effect of enhancer regions on target gene expression, reporter assays are often used.

1.2.6.1 Reporter assays to study single enhancers

Since the identification of the first β-Globin enhancer, reporter assays have become a key tool for assessing the ability of an enhancer to influence target gene function (Banerji, Rusconi, and Schaffner 1981). The first reporter assays were developed to study single or small numbers of candidate enhancers, where the enhancers were often cloned into plasmids upstream of a reporter gene, for example GFP or Luciferase, under the control of a promoter which cannot highly upregulate transcription by itself (Hannah, Jennens-Clough, and K. V. Wood 1998; Soboleski, Oaks, and Halford

2005). These plasmids can be transfected into cells, and reporter gene expression can be measured.

Reporter assays have been further developed to take advantage of *in vivo* systems. As there are many key aspects that additively could be or are important for enhancer function, for example transcription factors, coactivators and histone modifications, many researchers have turned to using reporter assays in whole organisms, to ensure that the candidate enhancers are able to access all of the enhancer machinery required. For example, reporter assays have been developed in Zebrafish models, where the effect of disease associated single nucleotide polymorphisms (SNPs) in enhancer regions can be compared against the WT enhancer region in the same embryo, through reporter fluorescence of two different colours (Bhatia et al. 2021).

1.2.6.2 High-throughput reporter assays

Although reporter assays can be useful when investigating a small number of enhancer regions, it is not easy to carry these out in a high-throughput manner. The development of massively parallel reporter assays (MPRA), such as self-transcribing active regulatory region sequencing (STARR-seq, Arnold et al. 2013) has enabled many thousands of candidate enhancers to be screened within a single experiment. In STARR-seq, reporter libraries are generated, where candidate enhancer regions of interest are cloned into plasmids upstream of target genes containing barcodes. After transfection of the library into cells and transcription of the target genes, the transcripts can be sequenced and the barcodes utilised to identify the enhancers increasing the number of transcripts present in the population (Arnold et al. 2013).

MPRA are useful in their strategy for assessing enhancer strength in a high-throughput manner, however even though many thousands are being screened simultaneously, these techniques are still used to screen single enhancers in parallel. Evidence continues to grow that enhancers may not function solely on their own, but can cluster with other regulatory elements to have an additive effect on gene expression (Hnisz, Abraham, et al. 2013; Thomas et al. 2021). For example, in mouse embryonic stem cells, a cluster of enhancers are found at the *Fgf5* locus (Thomas et al. 2021). Some of these enhancers did not previously exhibit any enhancer activity when analysed during STARR-seq analysis, however when coupled with other regulatory regions at

this locus, were found to have an additive effect that did contribute to target gene expression (Thomas et al. 2021). This therefore suggested that despite the advantages offered by MPRA, enhancers identified as being low strength that could have been previously dismissed, may actually be required for target gene expression when in combination with other regulatory elements at the target locus (Thomas et al. 2021).

1.2.7 Enhancers through development

To proceed through development, a complex network is required in order to turn on and turn off gene expression of required genes at the correct time (reviewed by Furlong and Levine 2018; Robson, Ringel, and Mundlos 2019). A single gene can interact with multiple enhancer sequences, with each one being activated by a different combination of signals and transcription factors. Therefore, at different stages of development, gene expression can easily be manipulated to finely control the genes that are actively being translated and the genes that are silent across the genome, (reviewed by Furlong and Levine 2018; Robson, Ringel, and Mundlos 2019).

Particularly during development, intermediate enhancer states found between inactive and active have been documented (Rada-Iglesias et al. 2011). These stages require the binding and activity of many different transcription factors and cofactors. This is likely due to the cell generating a way to fine-tune gene expression, and ensure that only the exact pattern of required genes are being expressed at any one time. The transition from inactive through to active enhancers has been investigated from embryonic stem cell through differentiation, in many different cell differentiation pathways, as a model for this switching of enhancer behaviour (Rada-Iglesias et al. 2011; Buecker and Wysocka 2012; Spitz and Furlong 2012).

To maintain gene expression within a required cell state, expression of genes is tightly controlled (reviewed by Buecker and Wysocka 2012). One such way is that the gene body and the corresponding enhancers of genes not being expressed are switched off and found within nucleosome dense regions, where the transcriptional machinery is not able to access the DNA sequences (reviewed by Heinz et al. 2015). However, when a switch in gene expression pattern occurs, for example during cell differentiation, it is necessary to be able to turn on expression of the genes that are tied up in repressive chromatin (reviewed by Heinz et al. 2015).
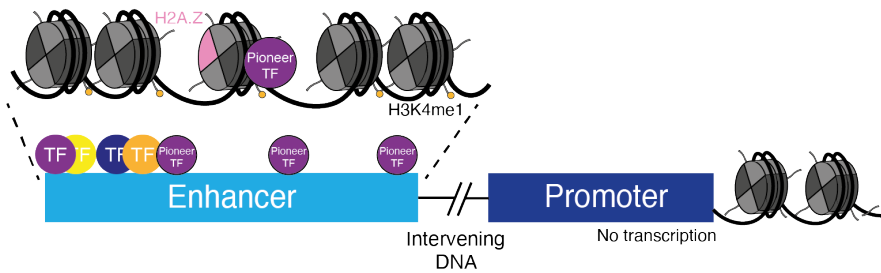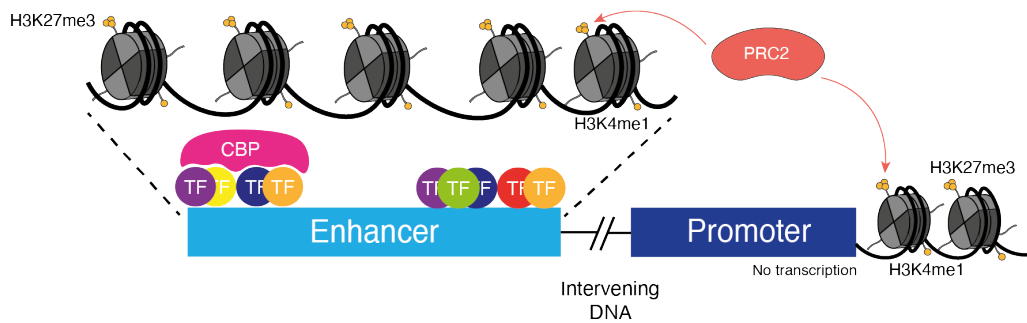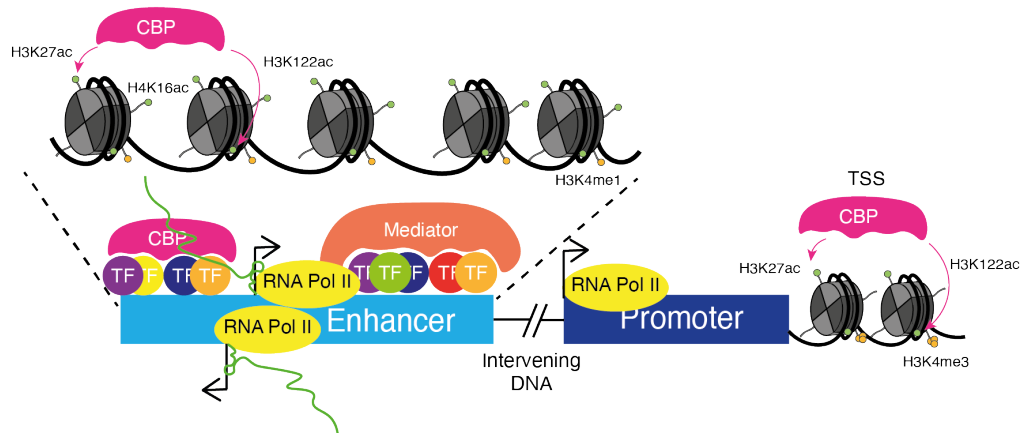
Figure 1.8: Enhancers can be found at different stages of activation in development (A) Primed enhancers are opened by pioneer factors and marked by H3K4me1. Nucleosomes are remodelled to contain histone variants that promote open chromatin, such as H2A.Z. (B) Poised enhancers interact with the target promoter through PRC2, but there is no target gene expression. CBP is bound but inactive, so H3K27me3 and H3K4me1 are the predominant histone PTMs. (C) Active enhancers have H3K27ac, active CBP, eRNA transcription and transcription at the target promoter. The transcriptional start site (TSS) at the promoter has H3K27ac and H3K122ac deposited by CBP and H3K4me3.

1.2.7.1 Pioneer factors bind to heterochromatin

Pioneer factors are a group of TFs that are able to bind within heterochromatic regions, and begin the opening up of the nucleosomes (Figure 1.8A, reviewed by Balsalobre and Drouin 2022). Pioneer factors are able to bind to DNA sequences when they are wrapped around the histone octamer found within nucleosomes, which many TFs are not able to do (Zaret and Mango 2016). Once bound, the pioneer factors are then able to recruit cofactors that begin to activate the local chromatin environment.

1.2.7.2 Histones are remodelled to promote opening of chromatin

Histone or nucleosome remodelling proteins can be recruited, which can change the variants of histone proteins found within the nucleosomes, to promote more active chromatin (Figure 1.8A, Venkatesh and Workman 2015). For example, the incorporation of histone variant H2A.Z rather than H2A causes the nucleosome to be less stable (C. Jin and Felsenfeld 2007). This is due to the H2A.Z nucleosomes being less able to interact with one another, forcing them further apart, causing the DNA to be less tightly wound around the nucleosome and the DNase hypersensitive regions between the nucleosomes to increase (C. Jin and Felsenfeld 2007). Thus, transcriptional activation is favoured, as the transcriptional machinery is more able to access the DNA sequences (reviewed by Giaimo et al. 2019). Nucleosome remodelling proteins can also be recruited, which act to deplete enhancer regions of nucleosomes, making the DNA more accessible to TFs and transcriptional machinery (reviewed by García-González et al. 2016).

1.2.7.3 Primed enhancers

The binding of pioneer transcription factors and the initial remodelling of histones can form what has been termed a "primed enhancer" (Rada-Iglesias et al. 2011). Enhancer priming is the first step from completely inactive enhancer sequences on the way to a truly active enhancer (reviewed by Calo and Wysocka 2013). Primed enhancers are classified as having bound pioneer TFs, with the DNA sequences containing reduced levels of DNA methylation (Figure 1.8A). Histones are marked with H3K4me1, but there is no H3K27ac or further active enhancer histone PTMs. Histone variants such as H2A.Z and H3.3 have also been thought to play a role within enhancer priming

50

(Figure 1.8A). However, the key marker of primed enhancers is that despite these initially activating markers at the enhancer, there is no transcriptional activity at the target promoter (Figure 1.8A, Rada-Iglesias et al. 2011).

1.2.7.4 Poised enhancers

After a primed enhancer has recruited further coactivators and increased the activating environment for transcription, the enhancer element can be classed as a "poised enhancer" (Figure 1.8B). Poised enhancers seem to be held by the cell as an intermediate prior to complete enhancer activation, where enhancer-promoter interactions have been established but no transcription of the target gene has been initiated (Sanyal et al. 2012; Calo and Wysocka 2013).

Poised enhancers exhibit an alternative pattern of histone PTMs. Histones are marked with H3K4me1, as they were previously as a primed enhancer, however H3K27 at this stage is trimethylated (Figure 1.8B, Zentner, Tesar, and Scacheri 2011). The deposition of repressive H3K27me3 modifications at these poised enhancers (Rada-Iglesias et al. 2011; Zentner, Tesar, and Scacheri 2011) is done by the Polycomb repressive complex 2 (PRC2, reviewed by Di Croce and Helin 2013). PRC2 has been shown to aid interactions between poised enhancers and promoters, allowing them to interact, however both the poised enhancer and target promoter have repressive H3K27me3 (Figure 1.8B, Cruz-Molina et al. 2017).

Despite H3K27 being trimethylated, not acetylated, poised enhancers do exhibit binding of CBP (Figure 1.8B, Creyghton et al. 2010). It is likely that during enhancer priming, the pioneer TFs have recruited coactivators such as CBP and PRC2, aiding the further recruitment of other TFs without pioneer activity, helping to open up chromatin. Once the regulatory proteins have bound and have established enhancer-promoter contacts, yet no transcription is active, this is likely when the enhancer can be classed as poised rather than primed (Figure 1.8B).

1.2.7.5 Active enhancers

Upon transition from a poised to an active enhancer, Polycomb group complexes (PcG) are evicted from the enhancer and the repressive H3K27me3 modifications induced by PRC2 are demethylated through the recruitment of the histone demethylase JMJD3

(Figure 1.8C, Vernimmen et al. 2011).

The Mediator complex is a large multisubunit complex, that aids in enhancer-promoter interactions (reviewed by Malik and Robert G Roeder 2010). Coactivator proteins, such as CBP, recruit the Mediator complex to enhancers, which can then go on to recruit RNA Pol II (Figure 1.8C, reviewed by Malik and Robert G Roeder 2010). This initiates bidirectional transcription of enhancers, producing eRNAs (Figure 1.8C, reviewed by W. Li, Notani, and Rosenfeld 2016).

The histone modification most frequently associated with enhancer activity is H3K27ac, which is a marker of active enhancers (Creyghton et al. 2010). Active enhancers also exhibit H3K122ac within the histone core, and some active enhancers have been shown to exhibit H4K16ac (Tropberger et al. 2013; Taylor et al. 2013; Pradeepa et al. 2016). H3K27ac, H3K122ac and H4K16ac are not found at regulatory elements that are poised or primed, implicating them as markers for active enhancer status (Creyghton et al. 2010; Bonn et al. 2012; Tropberger et al. 2013; Taylor et al. 2013; Pradeepa et al. 2016).

RNA Pol II transcription at active enhancers can produce eRNA transcripts which are able to bind to the inactive conformation of the AIL within CBP (D. A. Bose et al. 2017). This binding displaces the AIL to an active conformation, where CBP is then able to deposit activating acetyl PTMs onto histones, such as H3K27ac and H3K122ac, further activating the transcriptional environment (D. A. Bose et al. 2017).

CBP is then able to deposit H3K27ac marks onto the target promoter, which is in close proximity to the enhancer, and these H3K27ac marks are required for RNA Pol II to bind at the promoter (Q. Jin et al. 2011). Consequently, the remaining transcriptional machinery found bound to the enhancer, such as TFs and coactivators, can be donated to the promoter, further aiding target gene expression (Figure 1.8C, Calo and Wysocka 2013).

At this stage of active transcription of both the enhancer element itself and the target gene, the enhancer is termed as active (Figure 1.8C, Calo and Wysocka 2013).

1.2.8   3D organisation of enhancers

Enhancers associate with their target promoter, but only when they are found within the same topologically associated domain (TAD) (Dixon et al. 2012). These domains

are regions of chromatin that are constrained by CTCF binding at boundary sites which recruits rings of cohesin. Cohesin is able to form rings around the chromatin, containing all regulatory elements and transcriptional machinery within, forming a TAD (Dixon et al. 2012). Within TADs, cohesin rings can form smaller loops, constrained by binding of Mediator, known as insulated neighborhoods (Phillips-Cremins et al. 2013; Dowen et al. 2014).
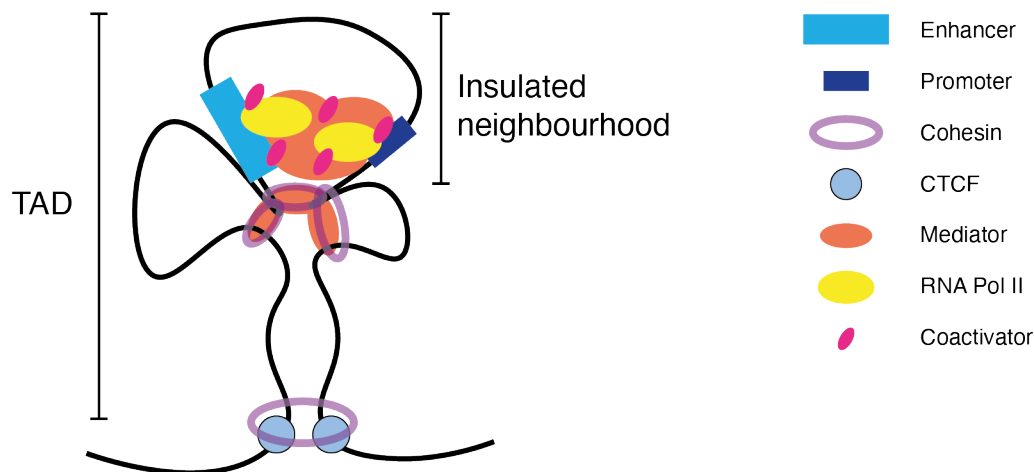
Figure 1.9: 3D organisation of the genome finds enhancers within TADs
Organisation of topologically associated domains (TADs) and insulated neighbourhoods found within TADs, both constrained by cohesin rings. TADs are bound by CTCF, whereas insulated neighbourhoods are bound by Mediator. The domains help to aid enhancer-promoter interactions.

These levels of domains are utilised by the cell to control gene expression and to ensure that only the correct regulatory elements are coming into contact with one another. However, this can be manipulated in many diseases, for example in cancer development this is often done to allow oncogenic expression to be regulated by more powerful enhancers (Hnisz, Weintraub, et al. 2016).

1.2.9 Super-enhancers

A single large enhancer or a group of enhancers clustered together that confer cell fate or determine lineage development can be termed as a super-enhancer (Figure 1.10, Lovén et al. 2013; Whyte et al. 2013).
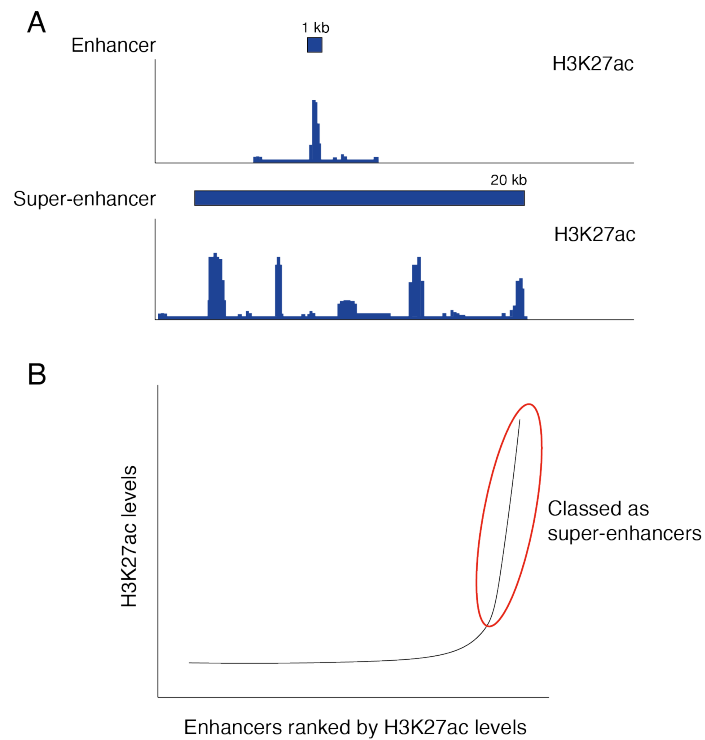
Figure 1.10: Super-enhancers are larger and have higher H3K27ac levels than general enhancers
(A) Representation of genome tracks for H3K27ac from ChIP-seq, identifying an enhancer and a super-enhancer. (B) Representation of plot ranking enhancers by H3K27ac levels, demonstrating cut-offs between super-enhancers and enhancers.

Single enhancers generally occur at the same level of activation across many cell types, whereas super-enhancers are usually found only in a single cell type, due to their role in cell fate-determination (Hnisz, Abraham, et al. 2013). If the super-enhancer is formed from multiple enhancers, as the individual enhancers within the cluster are all influencing the same outcome they are termed as being one large regulatory element (Whyte et al. 2013). Super-enhancers are also typically associated with increased levels of target gene transcription, compared to single enhancers (Whyte et al. 2013).
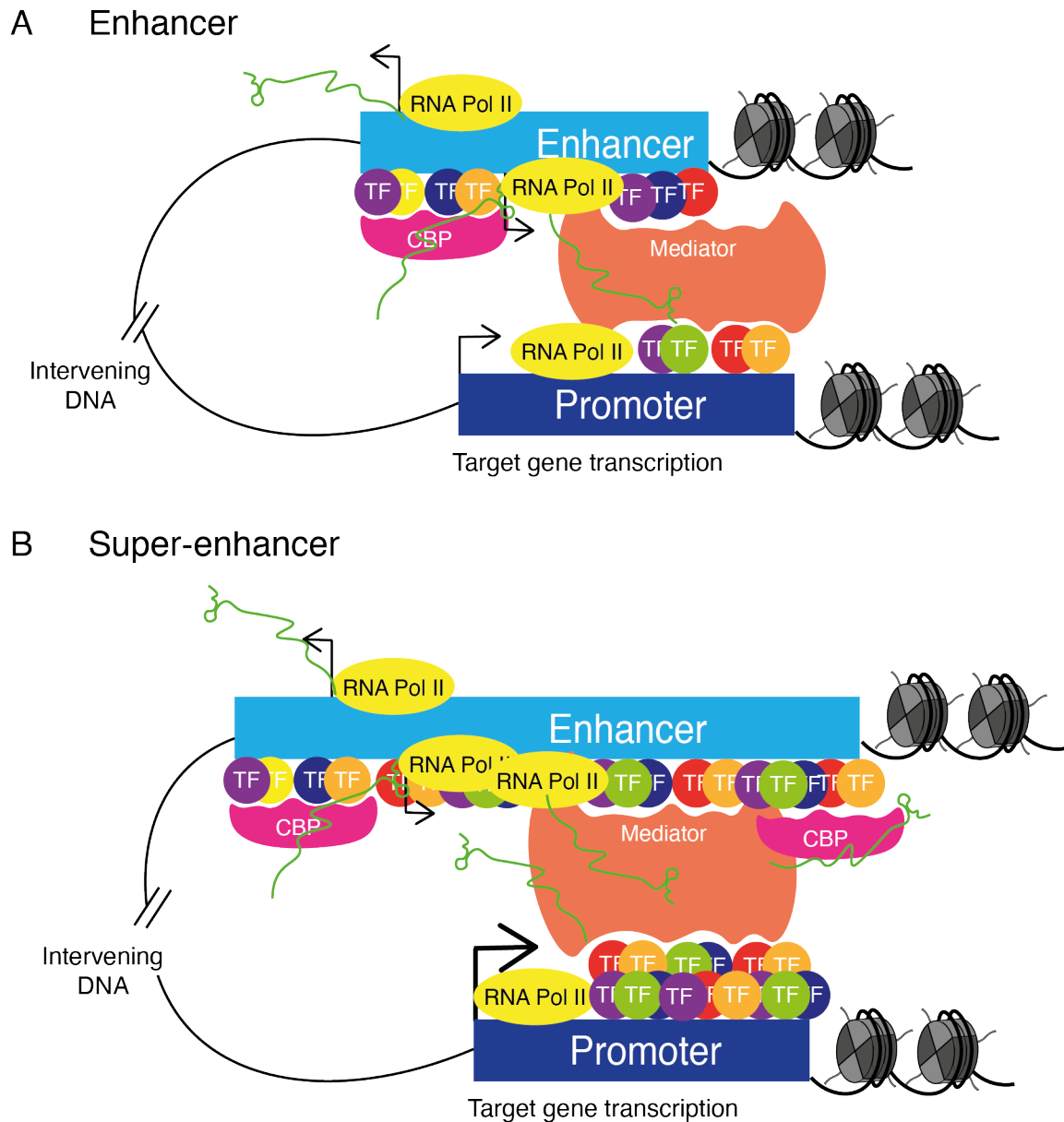
Figure 1.11: Super-enhancers are larger than enhancers and have increased transcriptional machinery

(A) At a general enhancer transcriptional machinery including transcription factors (TFs), coactivators such as CBP and Mediator and RNA Pol II assemble. eRNAs (green) are transcribed due to RNA Pol II binding and the target gene is transcribed. (B) Super-enhancers exhibit the same features of an enhancer, just at greater levels. Increased TF binding causes increased coactivator binding, resulting in higher levels of RNA Pol II. This results in higher levels of eRNA transcripts. As a consequence of increased activity at the super-enhancer, target gene transcription is greatly increased.

1.2.9.1   Super-enhancers are larger than enhancers

Super-enhancers can be a single large enhancer, or a group of multiple enhancers clustered together (Lovén et al. 2013). In one study examining enhancers in B lymphocyte cells, general enhancers were found to be of median 1.3 kb, whereas what was classed as a super-enhancer was of a median of 19.4 kb (Lovén et al. 2013). Super-enhancers are measured in size by a combination of H3K27ac, RNA Pol II and Med1 binding across a locus, with large stretches of enrichment indicating a super-enhancer (Figure 1.10, Whyte et al. 2013; Hnisz, Abraham, et al. 2013).

1.2.9.2   TF binding at super-enhancers

Super-enhancers are typically larger regions of TF binding sites, therefore, higher levels of TF binding are found at super-enhancers as compared to single enhancers (Figure 1.11, Whyte et al. 2013). Super-enhancers were first identified as regulatory elements driving cell fate changes in embryonic stem cells (ESCs), and as such an initial group of TFs key to transcription in ESCs were identified as being key for super-enhancer function (Whyte et al. 2013; Hnisz, Abraham, et al. 2013). This group of TFs included TFs well established in ESC function, including Oct4, Sox2 and Nanog (Whyte et al. 2013).

What sets apart this group of key TFs is that not only is transcription of each TF controlled by its own super-enhancer, each TF has been shown individually to bind to its own super-enhancer (Hnisz, Abraham, et al. 2013). This provided evidence for the formation of positive feedback autoregulatory loops between groups of TFs, which as they all bind to their own and each other's enhancer elements, can rapidly increase transcription of all TFs within the group, to aid transcription across many further super-enhancers throughout the cell (Hnisz, Abraham, et al. 2013; Whyte et al. 2013).

In addition, the identified super-enhancers were further enriched in binding motifs for this group of TFs compared to more general enhancers (Hnisz, Abraham, et al. 2013). This showed that more TF molecules were able to bind to each super-enhancer at any one time, further activating the transcriptional environment (Hnisz, Abraham, et al. 2013).

1.2.9.3 Coactivator binding at super-enhancers

Studies into binding of coactivators have revealed significant increases of such proteins are found at super-enhancers compared to enhancers (Figure 1.11, Lovén et al. 2013). For example, enrichment of the Mediator complex subunit Med1 was 18 times higher in super-enhancers compared to enhancers (Lovén et al. 2013).

Binding of coactivators, such as BRD4 has also been shown to be enriched at super-enhancers compared to general enhancers, with levels at super-enhancers shown to be 16 times higher in B lymphocyte cells (Lovén et al. 2013).

1.2.9.4 Super-enhancers have higher levels of H3K27ac

H3K27ac has been identified as a key determinant for the identification of super-enhancers compared to a more general enhancer (Hnisz, Abraham, et al. 2013). This has enabled use of H3K27ac as a marker for super-enhancer activity. H3K27ac has been used to compare super-enhancer activity in cancer cells. H3K27ac levels were ranked in order of increasing H3K27ac signals, and the H3K27ac signals when plotted produced an exponential curve (Figure 1.10B). Generating a cut-off where the curve becomes exponential has also been thought to be another method of selection between general enhancers and super-enhancers (Figure 1.10, Hnisz, Abraham, et al. 2013).

1.2.9.5 eRNAs are transcribed from super-enhancers

RNA Pol II has been shown to localise more at elements classed at super-enhancers than enhancers (Figure 1.11, Hnisz, Abraham, et al. 2013). Super-enhancers have been shown to have increased levels of Mediator and this could explain why higher levels of RNA Pol II are recruited to super-enhancers (Malik and Robert G Roeder 2010; Lovén et al. 2013).

Due to this increase in RNA Pol II, it was therefore no surprise that super-enhancers have been shown to also transcribe eRNAs (Hnisz, Abraham, et al. 2013); consistently, these elements may produce higher levels of eRNAs than general enhancers, as indicated by RNA-seq experiments (Figure 1.11, Hnisz, Abraham, et al. 2013).

1.2.9.6 Super-enhancers in cancer

Not only are super-enhancers utilised in cell differentiation and development, but due to their nature to be able to highly drive cell regulation and proliferation, they are often manipulated in the development of cancer (Lovén et al. 2013). Analysis of H3K27ac, Med1 and BRD4 levels identified many super-enhancers implicated in upregulating aberrant transcription in multiple myeloma cells (Lovén et al. 2013).

Super-enhancers can be acquired or can form *de novo*, upregulating transcription of specific transcription factors involved in proliferation and growth (Sur and Taipale 2016). Different super-enhancers can form around the same gene in multiple different cancer types (Sur and Taipale 2016). For example, surrounding the *MYC* locus, different super-enhancers have been shown to increase *MYC* expression in leukemia, pancreatic cancer and colorectal cancer (Lovén et al. 2013; Sur and Taipale 2016).

1.2.10 Enhancers and phase separated condensates

Phase-separated condensates act as membraneless organelles, where proteins and nucleic acids cluster closely together at high concentrations through electrostatic interactions into a separated phase, compared to the dilute environment (Figure 1.12A, Brangwynne et al. 2009). Phase separation has rapidly been implicated as a mechanism used by cells in the control of transcription and also at enhancers, to be able to collect the transcriptional machinery and aid donation to a target promoter (Figure 1.12B, Hnisz, Shrinivas, et al. 2017).
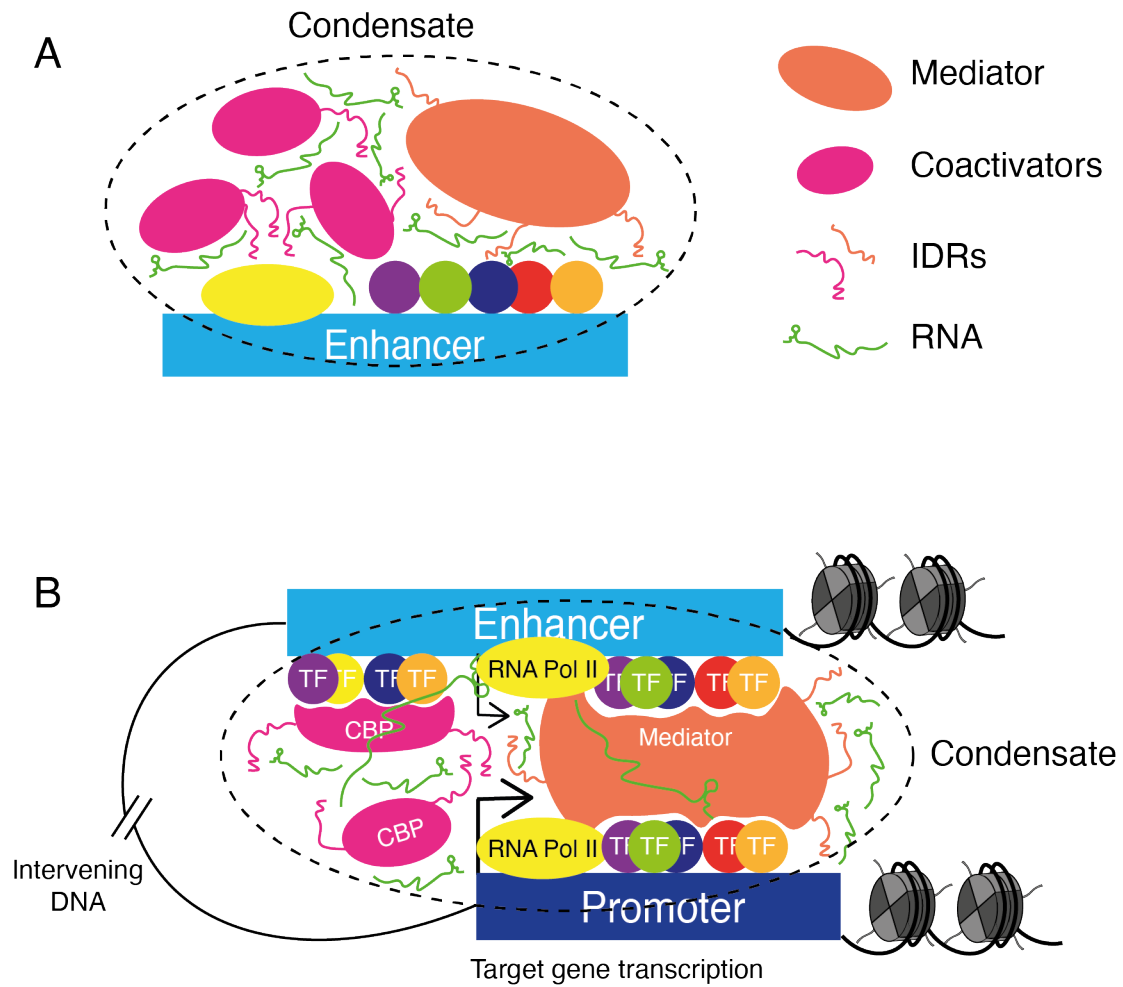
Figure 1.12: Phase separated condensates form at enhancers

(A) Condensates form at enhancers. Coactivators and proteins such as Med1 have intrinsically disordered regions (IDRs) which form electrostatic interactions with RNA, in particular eRNAs, at enhancers. This results in the formation of a condensate, which separates out into a distinct concentrated phase, compared to a dilute environment. (B) The condensate formed at the enhancer can be used to donate the transcriptional machinery to the target gene promoter.

1.2.10.1 Proteins with regions of intrinsic disorder promote phase separation

Specific proteins have been implicated as promoting phase separation, and many of these are known to play roles within transcription and at enhancers. Many of these

proteins contain regions of intrinsic disorder (IDR), which do not have any defined secondary structure (reviewed by Martin and Holehouse 2020). Therefore they can cluster together with nucleic acids and further proteins using electrostatic interactions, forming large regions that separate out as a dense phase compared to the environment (Figure 1.12, reviewed by Martin and Holehouse 2020). In addition, to form phase separated condensates at enhancers, transcription factors have been thought to be required to bind at enhancer TF binding sites, to help recruit coactivators and allow the formation of a surrounding condensate (Shrinivas et al. 2019). Without the DNA, a condensate of proteins only is unable to form (Shrinivas et al. 2019).

IDRs have been identified and shown to promote phase separation in BRD4, Med1 and p300, all of which play important roles in facilitating transcriptional activation at enhancers (Sabari, Dall'Agnese, et al. 2018; Ma et al. 2021). In addition, IDRs have been found to have RNA binding properties, which can further provide evidence for formation of phase separation, as nucleic acids are key to condensate formation (Castello et al. 2016; He et al. 2016).

1.2.10.2 eRNAs may support phase separated condensates at enhancers

Nucleic acids, particularly RNA species, aid in the composition of phase separated condensates due to their negative charge (Figure 1.12). RNA has been shown to act as a scaffold to hold the condensate in place for the remainder of the condensate to form around (Banani et al. 2016). As previously discussed, enhancer elements are bidirectionally transcribed into enhancer RNAs (eRNAs) (De Santa et al. 2010; T.-K. Kim et al. 2010), subsequently eRNAs have been implicated in promoting the formation of phase separated condensates at enhancer regions (Nair et al. 2019).

Phase separation at enhancers may play a role with the transcription of the eRNAs themselves, as condensate disruption has been shown to decrease eRNA transcript levels at highly active enhancers (Nair et al. 2019).

It appears likely that upon initial transcription of eRNAs, the increase in short RNA transcripts which are negatively charged cause a change in the electrostatic forces surrounding the enhancer element where the RNA is being transcribed from and result in the formation of a condensate (Henninger et al. 2021). This condensate is formed surrounding the transcriptional machinery, for example RNA Polymerase II (RNA

Pol II) and Mediator, helping to hold together and scaffold transcriptional machinery (Figure 1.12). Through interactions with the target promoter these condensates cause this transcriptional machinery to access the TSS and aid in transcription of the target gene (Henninger et al. 2021).

However, investigations into the dependency of eRNAs within phase separated condensates have shown that condensates were still able to form upon eRNA degradation using immunofluorescence, highlighting that although eRNAs may be important, phase separation at enhancers can still occur in their absense (Decker et al. 2022).

eRNAs have been shown to interact with protein complexes found at enhancers, forming enhancer RNA ribonucleoproteins (eRNPs), and the presence of eRNAs is required for the eRNP complex as a whole to phase separate (Nair et al. 2019). Upon transcription of the target gene where longer and larger quantities of RNA transcripts are being produced, this shifts the charge balance of the condensate, overwhelming the electrostatic forces and causes the dissolution of the condesate, allowing gene transcription to occur (Henninger et al. 2021).

1.2.10.3 Super-enhancers have increased condensate size

At larger enhancer elements or at super-enhancers, condensate size and strength has been shown to drastically increase (Hnisz, Shrinivas, et al. 2017). This is logical, as super-enhancers have been shown to be larger in size and have increased transcription factor and coactivator binding levels and higher levels of eRNA transcription, therefore there is more machinery present at these super-enhancers that can promote the formation of condensates (Lovén et al. 2013). Disruption of condensates at super-enhancers has been shown to drastically reduce transcriptional activity and target gene expression (Sabari, Dall'Agnese, et al. 2018).

1.3 *TAL1* enhancer

1.3.1 T cell acute lymphoblastic leukemia

T-cell acute lymphoblastic leukemia (T-ALL) is a cancer of T cell progenitors, in which hematopoietic stem cells develop aberrantly into tumour cells, in place of differentiating into mature T cells (Karrman and Johansson 2017). T-ALL accounts for "10-15% of paediatric and 25% of adult ALL cases", with survival rates for children and young adults being estimated between 60 and 75% (Silverman et al. 2001; Hoelzer et al. 2002; Van Vlierberghe and Ferrando 2012).

There are multiple molecular pathways in which tumorigenesis can occur, mainly through the manipulation of oncogenic gene expression to drive cancer development (as reviewed by Van Vlierberghe and Ferrando 2012). One recurring feature of different T-ALLs is the use of transcription factors as oncogenes and oncoproteins, expressed aberrantly in the cell, to maintain cell survival. One common cause of this aberrant expression are chromosomal translocations, which can place oncogenes in different genetic environments. This can activate and highly upregulate transcription of oncogenes and cancer drivers, where in their native transcriptional environment they should be silenced (Van Vlierberghe and Ferrando 2012).

1.3.2 A regulatory complex controls gene expression within Jurkat cells

In Jurkat cells, a T-ALL cell line, three transcription factors were first identified as being critical for gene regulation through the cell: TAL1, GATA3 and RUNX1 (Sanda et al. 2012). Knockdown of each TF within this group was shown to consequently reduce expression of the other two, therefore it was suggested that TAL1, GATA3 and RUNX1 together form an autoregulatory loop to control gene expression in this cell line (Figure 1.13).

Analysis of TF binding motifs found that TAL1, GATA3 and RUNX1 bind at enhancers of each gene, including their own, indicating this autoregulatory loop as a feed-forward mechanism. *MYB* was identified as a downstream target of TAL1, GATA3 and RUNX1, as knockdown of each of these three proteins reduced *MYB* expression and MYB target gene expression (Figure 1.13, Sanda et al. 2012). This

regulatory complex was later expanded to include TAL1 binding partners E2A and HEB and was termed the TAL1 complex (Sanda et al. 2012; Mansour, Abraham, et al. 2014).
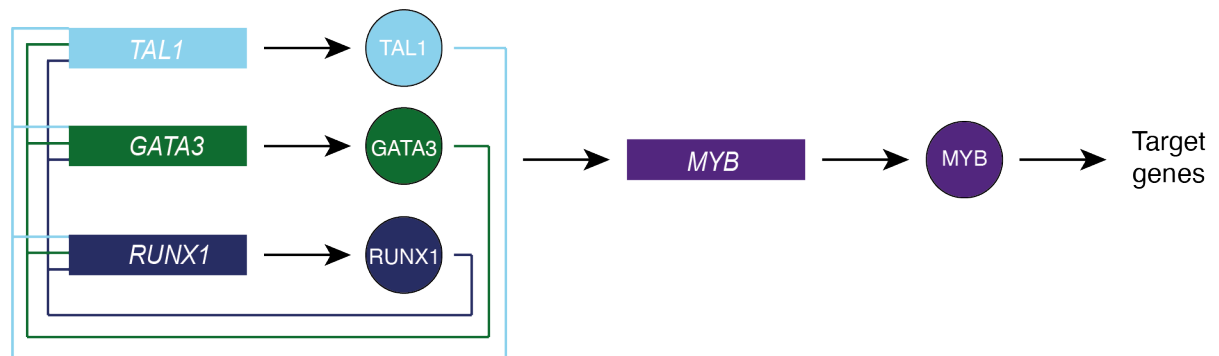
Figure 1.13: An autoregulatory loop between transcription factors forms in Jurkat cells to control gene expression
TAL1, *GATA3* and *RUNX1* form an interconnected loop, whereby they bind at each other's enhancers and genes to positively upregulate transcription. These three proteins upregulate *MYB* expression and increased levels of MYB enable MYB target gene expression to increase. Based upon (Sanda et al. 2012).

1.3.2.1 TAL1

In Jurkat cells, T-Cell Acute Lymphoblastic Leukemia protein 1 (*TAL1*), originally named Stem Cell Leukemia protein (*SCL*), is the key oncogene that drives cancer development. TAL1 is a basic helix-loop-helix (bHLH) transcription factor (Begley et al. 1989) containing a DNA binding domain as well as two alpha-helices used for homo- and heterodimeric interactions (Bernard et al. 1990; Q. Chen et al. 1990). *TAL1* expression is critical for hematopoiesis as it has been shown that homozygous *TAL1* knockout mice are inviable, as they do not survive past embryonic day 10.5 due to lack of red blood cells (Aifantis, Raetz, and Buonamici 2008).

Upregulation of *TAL1* oncogene expression is a key driver of some T-ALL subtypes. Studies in mice have demonstrated the potency of *TAL1* as an oncogene, with one study showing that all mice with ectopic *TAL1* expression died of T-ALL within a period of 70 weeks (Condorelli et al. 1996).

1.3.2.2   TAL1 functions as a heterodimer with E2A or HEB

The TAL1 protein is able to form heterodimers with E2A or HEB (Figure 1.14A, Hsu, Cheng, et al. 1991; Hsu, Wadman, and Baer 1994). Both HEB and E2A are bHLH transcription factors that function in healthy cells through the formation of homodimers (Voronova and F. Lee 1994). The E2A dimers are made up of alternative splicing forms of the *E2A* gene; E12 and E47 (Voronova and F. Lee 1994). These homodimers are key for activating and maintaining transcription at many enhancers throughout the genome. E2A and HEB are both E-box binding proteins that bind to the E-box motif, CANNTG, which is often found at enhancers (Figure 1.14B, Hsu, Cheng, et al. 1991).
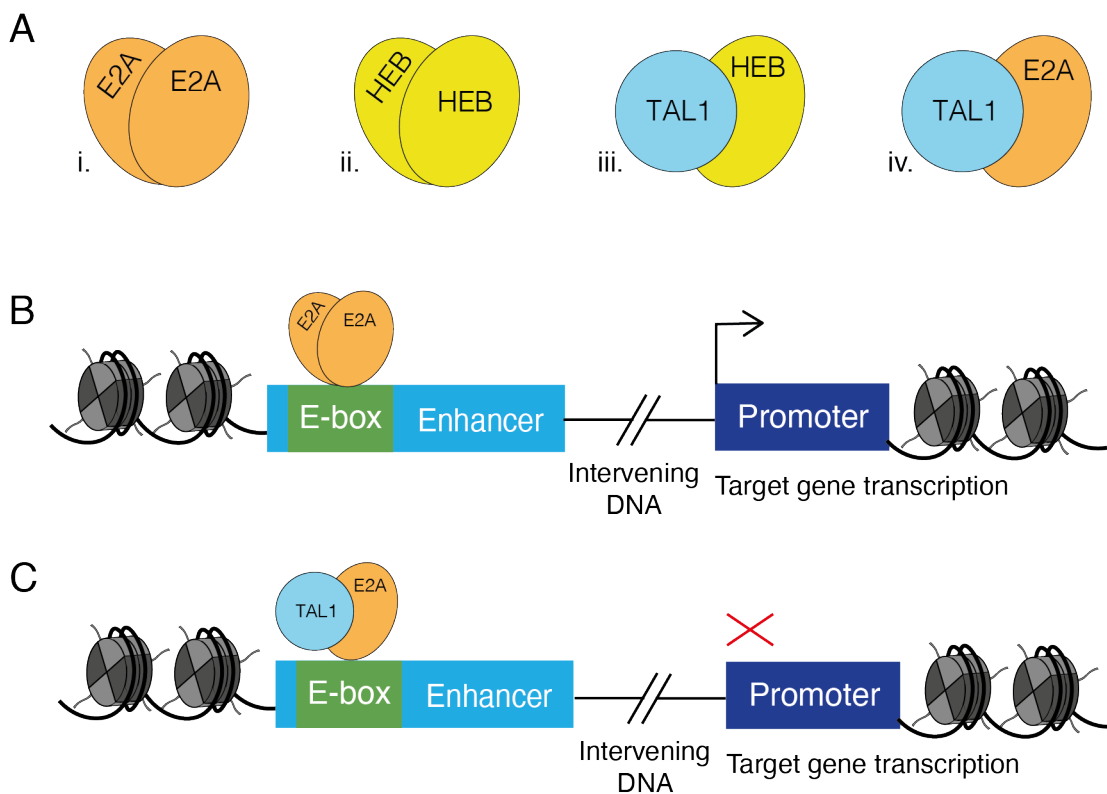
Figure 1.14: TAL1 binds to HEB and E2A to regulate gene expression
(A) E2A (i) and HEB (ii) can form homodimers or heterodimers with TAL1 (iii and iv). (B) Homodimers of E2A and HEB can bind to E-box motifs within enhancers, which activates target gene transcription. (C) Binding of TAL1-HEB or TAL1-E2A heterodimers at E-box motifs in enhancers downregulates target gene transcription.

64

When *TAL1* is more highly expressed, TAL1-E2A or TAL1-HEB heterodimers can form in place of E2A or HEB homodimers (Figure 1.14A, Hsu, Cheng, et al. 1991; Engel et al. 2001). The heterodimers are able to bind to target E-box sequences within enhancers, meaning that these sites are blocked from being bound by the activating E2A or HEB homodimers. Consequently, transcription is downregulated across many genes within the genome (Figure 1.14, Aifantis, Raetz, and Buonamici 2008). The genes predominantly bound by E2A and HEB are those that promote T cell differentiation (Engel et al. 2001). Therefore, through *TAL1* overexpression, transcription of key developmental genes are blocked and TAL1 is able to promote a transcriptional state more similar to hematopoietic stem cell, which is advantageous in tumorigenesis (Engel et al. 2001).

1.3.2.3   GATA3

GATA3 is found within the GATA protein family, these transcription factors bind to GATA DNA motifs using two conserved zinc fingers (reviewed by Katsumura, Bresnick, and GATA Factor Mechanisms Group 2017). GATA3 has been shown to be embryonic lethal, where mice that were homozygous null for GATA3 did not survive past embryonic development day 12 (Pandolfi et al. 1995). GATA3 has been implicated as a transcription factor that is important for hematopoiesis and in T-cell development (Ko et al. 1991).

1.3.2.4   RUNX1

Runt-related transcription factor 1 (RUNX1) is a transcription factor that is often utlised in leukemia development. Chromosome translocations affecting RUNX1 are common in ALL and other leukemias such as acute myeloid leukemia (AML) (Golub et al. 1995; De Braekeleer et al. 2011). Once bound with a cofactor, RUNX1 has been shown to be phosphorylated and subsequently phosphorylate CBP (Wee et al. 2008).

1.3.2.5   c-MYB

The cellular MYB (c-MYB) gene is found within the human genome, but this gene was first reported upon the identification of the homologue viral MYB (v-MYB),

where certain viruses had copied a version of the MYB gene into their own genome (Klempnauer and Bishop 1984).

MYB is a transcription factor that binds to DNA (Biedenkapp et al. 1988), but cannot upregulate transcription without being bound by its coactivator CBP (Dai et al. 1996). MYB also exhibits pioneer factor properties, where it has been shown to bind to DNA wrapped around histone proteins (Fuglerud, Lemma, et al. 2017). For MYB to exhibit its pioneer activity, it recruits CBP as a coactivator, once recruited to histones CBP is then able to donate acetylation modifications onto histones, which causes release of MYB as a pioneer factor (Fuglerud, Ledsaak, et al. 2018). Then once chromatin has further opened and DNA is more accessible, MYB is able to bind to free DNA as a transcription factor rather than a pioneer, aiding in increasing the transcriptional activating environment (Fuglerud, Ledsaak, et al. 2018).

1.3.3 The *TAL1* enhancer is initiated by a MYB binding site

Through the somatic acquisition of a 12 bp insertion, generating two binding sites for the transcription factor MYB, a *de novo* super-enhancer has been shown to form in Jurkat cells (Figure 1.15, Mansour, Abraham, et al. 2014).
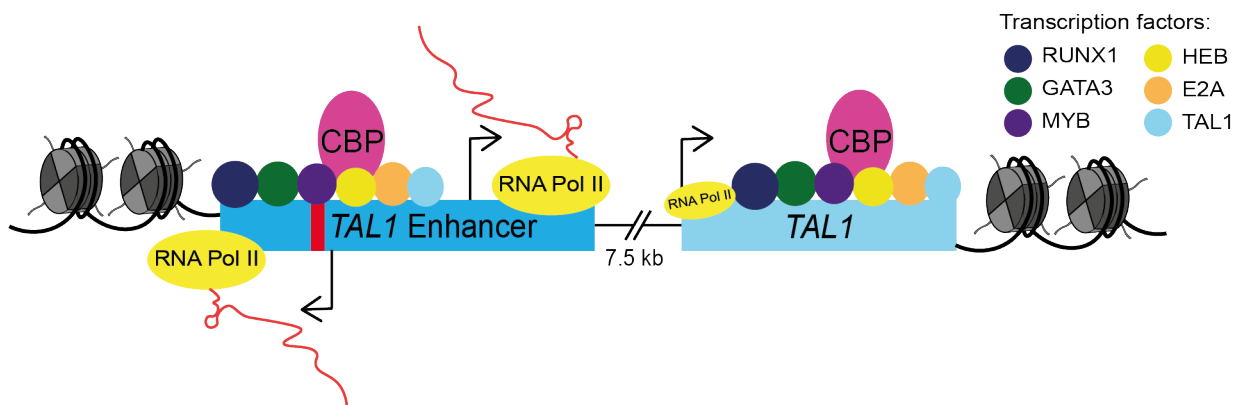
Figure 1.15: The *TAL1* enhancer regulates *TAL1* gene expression
The *TAL1* enhancer contains a 12 bp *de novo* insertion (red stripe) generating binding sites for the transcription factor MYB. CBP and the TAL1 complex are recruited and RNA Pol II bidirectionally transcribes a sense and antisense eRNA. CBP and the TAL1 complex also bind to the *TAL1* target promoter, and *TAL1* target gene transcription is high.

This enhancer has been shown to activate and drive transcription of the *TAL1* oncogene, and this *TAL1* overexpression is the key driver of T-ALL development in these cells (Mansour, Abraham, et al. 2014). Deletion of this locus has been shown to abrogate this enhancer function entirely and dramatically reduce *TAL1* expression (Mansour, Abraham, et al. 2014).

1.3.3.1 Identification of the *TAL1* enhancer

The *TAL1* enhancer element was identified through Chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq). ChIP-seq utilises the cross-linking of DNA to proteins followed by chromatin fragmentation. Fragments of chromatin can then be incubated with a chosen antibody and DNA-protein fragments bound to the antibody of choice can be isolated. Large scale sequencing of these DNA fragments followed by bioinformatic analyses produces tracks visible in a genome browser format, allowing genome-wide mapping of a protein binding or an epigenetic mark, depending upon the antibody chosen (Robertson et al. 2007).

ChIP-seq data analysing the *TAL1* locus in Jurkat cells revealed elevated H3K27ac levels at a site 7.5 kb upstream of the *TAL1* TSS (Mansour, Abraham, et al. 2014). Coupled with high RNA Pol II and Med1 levels, as well as the presence of CBP, this allowed the identification of this site as an enhancer. Due to the extent of the acetylation across the site, it was termed a super-enhancer (Mansour, Abraham, et al. 2014).

ChIP-seq for the factors within the TAL1 complex: TAL1, GATA3, RUNX1, MYB, HEB and E2A revealed that this complex was binding at this locus, with a very distinct peak visible for each protein (Mansour, Abraham, et al. 2014). Analysis of the DNA sequence at this peak revealed a monoallelic 12 bp *de novo* insertion, acquired by this cell line that is not present on the WT allele, nor in the human reference genome (Mansour, Abraham, et al. 2014). Motif analysis of this 12 bp sequence revealed generation of two binding sites for the transcription factor MYB (Mansour, Abraham, et al. 2014). Analysis of this enhancer element within a reporter construct, whereby enhancer activity can be measured by the activation of firefly luciferase, showed that upon *MYB* knockdown, activity of this enhancer element was dramatically reduced (Mansour, Abraham, et al. 2014).

1.3.3.2 The TAL1 complex is key for gene regulation throughout Jurkat cells

Investigations into the action of the TAL1 complex were carried out, showing that MYB and TAL1 directly interact through reciprocal Co-immunoprecipitation (Co-IP) experiments (Mansour, Abraham, et al. 2014). In addition, genome wide MYB binding sites were analysed for co-occupancy by the TAL1 complex and by CBP, through ChIP-seq fragments being displayed as heatmaps. Where fragments containing binding sites for MYB and either TAL1, RUNX1, GATA3 and CBP overlapped, this created a more dense pattern within the heatmap. When examining MYB and CBP co-occupancy, there was a dense cluster observable of CBP peaks at 0 kb away from the MYB binding sites, indicating that on a genome-wide scale, CBP and MYB binding sites co-localise. However, quantitative data with regards to the percentage of overlapping sites was not described (Mansour, Abraham, et al. 2014).

Despite the co-localisation of CBP at many sites as the TAL1 complex, including the *TAL1* enhancer, direct binding of CBP to the other complex components has yet to be shown.

1.3.4 3D organisation of the *TAL1* enhancer

1.3.4.1 Chromosome conformation capture and its derivatives

Chromosome conformation capture (3C) involves cross-linking protein to DNA (Dekker et al. 2002). Where two DNA sequences are held close together in space, mediated by protein-DNA interactions, these can be extracted, producing a map of interactions of differential DNA sequences within the genome. 3C identifies one DNA sequence with one interacting sequence and this was developed into chromosome conformation capture on chip (4C), which can identify interactions between one DNA sequence with many interacting sequences (Zhao et al. 2006). Carbon copy chromosome conformation capture (5C) took this a step further where many DNA sequences can be compared genome wide (Dostie et al. 2006). The development of Hi-C allowed identifications between chromatin regions genome wide (Belton et al. 2012). Hi-C has since been developed further: Capture Hi-C adds a target step to capture all Hi-C fragments containing regions of interest, for example promoter regions (Promoter Hi-

C), to increase sequencing depth and therefore achieve higher resolution of sequencing data upon analysis (Mifsud et al. 2015)

Chromatin interaction analysis by paired-end sequencing (ChIA-PET) produces genome-wide maps of interactions between the genome and a chosen protein (G. Li et al. 2010). Chromatin interaction analysis via droplet-based and barcode-linked sequencing (ChIA-Drop), allows cross-linked and fragmented chromatin to be separated into individual droplets, prepared with unique barcodes, which are ligated upon the library preparation process (Zheng et al. 2019). After high-throughput sequencing and bioinformatic analyses, chromatin regions found in close proximity to one another can be identified through containing the same barcode from individual droplets (Zheng et al. 2019).

1.3.4.2 Identification of the *TAL1* enhancer environment

Through 3C, 4C and Hi-C experiments, the 3D organisation of the *TAL1* locus and its regulatory elements in Jurkat cells have begun to be investigated.

A regulatory hub was identified that surrounded the *TAL1* locus, across many cell types (Y. Zhou et al. 2013). This hub indicated formation of a TAD encompassing nearly 90 kb including the *TAL1* gene. This appeared to be a locus that is often used by leukemia cell lines to drive leukemogenesis, as reprogramming of this hub was often found in different leukemic lines. For example, interactions between the *TAL1* promoter and the neighbouring *STIL* locus were identified in cell lines derived from T-ALL patients (Y. Zhou et al. 2013).

Using 3C in Jurkat cells, high levels of interactions between the *TAL1* promoter and the *TAL1* enhancer were identified, that were not identified in other cell types (Y. Zhou et al. 2013).

Subsequent experiments using ChIA-PET examined genome wide binding of cohesin in Jurkat cells, as cohesin mediates CTCF binding to constrain and mark the boundaries of TADs and insulated neighborhoods (Hnisz, Weintraub, et al. 2016). This provided a deeper insight into the 3D landscape of the Jurkat genome. This confirmed that the *TAL1* gene and the *TAL1* enhancer are located within the same insulated neighborhood. In addition, comparison of cohesin ChIA-PET data in a contrasting cell line of HEK293T cells showed that the insulated neighborhood sur-

rounding the *TAL1* locus is the same as in Jurkats. Therefore it is the addition of this *de novo* enhancer that is highly upregulating *TAL1* expression, and it is not as a result of a change to the 3D regulatory landscape (Hnisz, Weintraub, et al. 2016).

1.3.5 eRNAs are transcribed from the *TAL1* enhancer

eRNAs are relatively short-lived within the cell and are often non-polyadenylated, meaning they can be difficult to capture with traditional RNA-seq methods. Global run-on sequencing (GRO-seq) was developed to capture nascent RNA (Core, Waterfall, and Lis 2008). GRO-seq was carried out in Jurkat cells, and the data was analysed to identify eRNAs within this cell line. The GRO-seq dataset highlighted that eRNAs are transcribed from the *TAL1* enhancer locus, and the bidirectional transcription of the eRNAs centres surrounding the 12 bp *de novo* MYB binding site (Figure 1.16, Danko et al. 2015).

Bidirectional transcription of eRNAs from this locus has since been corroborated through TT-seq (Lidschreiber et al. 2021).
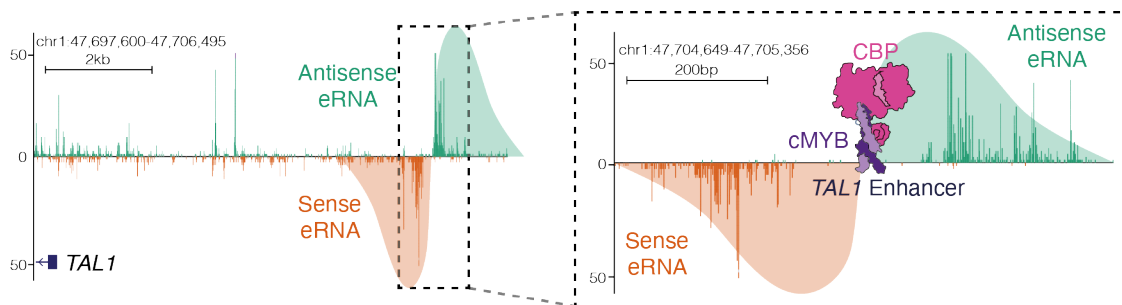
Figure 1.16: The *TAL1* enhancer is bidirectionally transcribed
The locus of the *TAL1* enhancer is bidirectionally transcribed, producing a sense eRNA (orange) and an antisense eRNA (green). The 12 bp MYB binding site is shown with MYB binding (purple) and the recruitment of bound CBP (pink), where the eRNAs are transcribed either sife of this locus. Data from (Danko et al. 2015), Figure by Dan Bose.

1.3.6 Downstream targets of the TAL1 complex

Since the initial discovery of the TAL1 complex being a master regulatory protein complex within Jurkat cells (Sanda et al. 2012; Mansour, Abraham, et al. 2014), in-

vestigations have begun to elucidate further downstream targets of the TAL1 complex.

1.3.6.1 ARID5B

The AT-rich Interactive Domain (ARID) protein family comprise a group of DNA binding proteins containing an ARID domain that binds directly to target DNA (Wilsker et al. 2002). ARID protein 5B is used in the differentiation to B lymphocytes from progenitor cells (Whitson, T. Huang, and Itakura 1999) and has been implicated as a binding partner of PHF2, a H3K9me2 demethylase (Baba et al. 2011). Mass spectrometry revealed protein-protein interactions between ARID5B and the histone deacetylase HDAC1, which implied a role for ARID5B in aiding the removal of further histone modifications (Joshi et al. 2013). The *ARID5B* locus is on chromosome 10 and is regulated by both an upstream enhancer and a super-enhancer (Leong et al. 2017).

In Jurkat cells, knockdown of the TAL1 complex proteins identified *ARID5B* as a downstream target of the TAL1 compex (Leong et al. 2017). ChIP-seq analysis at the *ARID5B* locus showed localisation of the TAL1 complex, CBP and H3K27ac at a super-enhancer 135 kb upstream of the *ARID5B* TSS (Sanda et al. 2012; Leong et al. 2017). Comparative analysis of *ARID5B* expression levels in T cells, hematopoietic stem cells and Jurkat cells showed that *ARID5B* was aberrantly upregulated in Jurkat cells (Leong et al. 2017).

One downstream target of ARID5B has been shown to be the *MYC* oncogene in Jurkat cells, as ARID5B bound to the *MYC* enhancer (Leong et al. 2017). This highlights the complex gene regulatory network in place in Jurkat cells, driven by the TAL1 complex.

1.3.6.2 *ARIEL*

The *ARID5B* super-enhancer is transcribed, with the sense eRNA named ARID5B-inducing enhancer associated long noncoding RNA (*ARIEL*, Ngoc et al. 2018; Tan et al. 2019). This eRNA is only transcribed upon activation by the TAL1 complex (Ngoc et al. 2018; Tan et al. 2019).

The *ARIEL* eRNA has been shown to exhibit functions previously proposed for eR-NAs. For example, knockdown of *ARIEL* reduced *ARID5B* gene expression, implying

evidence for this eRNA to have a role in transcriptional activity (Tan et al. 2019). In addition, loss of *ARIEL* showed decreased recruitment of Mediator to the enhancer, indicating a further role for this eRNA to aid in Mediator recruitment (Tan et al. 2019). Finally *ARIEL* has been shown to associate with both the *ARID5B* enhancer and promoter, adding evidence to eRNAs aiding enhancer-promoter interactions (Tan et al. 2019).

1.3.6.3   miR-223

In Jurkat cells, the TAL1 complex has also been shown to bind to a region 4 kb upstream of the TSS of a microRNA (miRNA) named *miR-223* and subsequently control its expression (Mansour, Sanda, et al. 2013). miR-223 targets the mRNA of the tumour suppressor gene *FBXW7*, an E3 ubiquitin ligase, which in healthy cells marks oncoproteins such as MYB and MYC for degradation (Mansour, Sanda, et al. 2013). Through the silencing of mRNA transcripts of this tumour suppressor gene, this results in upregulation of other oncogenes, and drives development of leukemia (Mansour, Sanda, et al. 2013).

Therefore, through controlling miRNA expression, the TAL1 complex can manipulate the expression of many other oncoproteins, demonstrating the vast transcriptional network this complex dictates in the development of T-ALL (Mansour, Sanda, et al. 2013).

1.4 Aims of this research project

The enhancer found upstream of the *TAL1* locus in Jurkat cells offers a unique opportunity to utilise as a model system for enhancer research.

Despite this enhancer being well studied, there are still questions to be asked as to the mode of action at this regulatory element. Apart from the identification of the eRNAs being transcribed from this enhancer, no work as yet has investigated the role that these eRNAs may play, in particular if there are interactions with the TAL1 complex. In addition, at every stage of TAL1 complex binding and functioning, CBP has also been shown to be present and necessary for function, however CBP has yet to be confirmed as part of this protein complex. Furthermore, due to this enhancer being MYB dependent, there is a possibility to to be able to deconstruct and manipulate this enhancer, to test models for enhancer function.

This research project will seek to use the *TAL1* enhancer and the regulatory protein complex bound at this regulatory element as a model to further uncover enhancer, eRNA and regulatory complex function. The aims of this project are as follows:

1. Determine the protein binding partners found acting at this regulatory element and begin to understand the action of RNA with this regulatory protein complex

2. Generate an endogenously tagged CBP Jurkat cell line, to be used for examining RNA binding with CBP

3. Develop a tandem tagged MYB and CBP Jurkat cell line, to be used in reciprocal pull-down experiments to investigate RNA binding within the CBP-TAL1 complex

4. Generate a model system utilising loss of MYB to dissect the action of the *TAL1* enhancer

Through utilising the *TAL1* enhancer as a model for enhancer function, we seek to decode the action of the localised regulatory protein complex and its interactions with RNA, in particular eRNAs. This will shed further light as to the action of eRNAs, the role of CBP at enhancers and delve deeper into the formation and function of enhancers.

Chapter 2

Materials and Methods

2.1 Antibodies, Plasmids, DNA and RNA oligonu-cleotides

2.1.1 Antibodies

Table 2.1: Table of antibodies
Usage: Western Blot (WB), Co-Immunoprecipitation (Co-IP), RNA
Immunoprecipitation (RIP).

Target	Source	Host species	Concen-tration	Usage
CBP	Cell Signalling (7389S)	Rabbit	0.53 mg/ml	WB, Co-IP, RIP
c-MYB	Abcam (ab45150)	Rabbit	2.2 mg/ml	WB, Co-IP, RIP
E2A	Cell Signalling (12258S)	Rabbit	0.53 mg/ml	WB, Co-IP, RIP
GATA3	Proteintech (10417-1-AP)	Rabbit	0.31 mg/ml	WB, Co-IP, RIP
RUNX1	Proteintech (19555-1-AP)	Rabbit	0.37 mg/ml	WB, Co-IP, RIP
TAL1	Santa Cruz (sc-393287)	Mouse	0.2 mg/ml	WB

Table 2.2: Table of antibodies 2

Target	Source	Host species	Concen-tration	Usage
TAL1	Proteintech (55317-1-AP)	Rabbit	0.27 mg/ml	WB, Co-IP, RIP
TCF12/HEB	Proteintech (14419-1-AP)	Rabbit	0.18 mg/ml	WB, Co-IP, RIP
H3	Abcam (ab1791)	Rabbit	1 mg/ml	RIP
K27ac	Abcam (ab4729)	Rabbit	1 mg/ml	RIP
FLAG tag	Proteintech (66008-3-Ig)	Mouse	0.5 mg/ml	WB, Co-IP
FLAG tag (M2)	Sigma (F3165)	Mouse	3.8 mg/ml	WB, Co-IP, RIP
HA tag	Proteintech (51064-2-AP)	Rabbit	0.8 mg/ml	WB, RIP
Halotag (polyclonal)	Promega (G928A)	Mouse	1 mg/ml	WB
Halotag (monoclonal)	Promega (G9211)	Mouse	2 mg/ml	WB
Rabbit IgG	Proteintech (30000-0-AP)	Rabbit	1.15 mg/ml	WB, Co-IP, RIP
Mouse IgG	Proteintech (10283-1-AP)	Rabbit	1 mg/ml	WB, Co-IP, RIP
GAPDH	Proteintech (60004-1-Ig)	Mouse	1 mg/ml	WB
HRP conjugated Anti-Mouse IgG (Heavy and Light chain)	Proteintech (SA00001-1)	Goat	-	WB Secondary antibody
HRP conjugated Anti-Rabbit IgG (Light chain only)	Stratech (211-032-171)	Mouse	-	WB Secondary antibody

2.1.2 Plasmids

Table 2.3: Table of plasmids

Plasmid	Bacterial resistance	Mammalian resistance	Source
pSpCas9(BB)-2A-GFP (pX458)	Ampicillin	-	Addgene (#48138)
pX458-CBP	Ampicillin	-	Based upon Addgene #48138, cloned in house (NC)
pX458-MYB	Ampicillin	-	Based upon Addgene #48138, cloned in house (NC)
CBP-FTHH-RT	Kanamycin	-	ThermoFisher
CBP-HTHAS-RT	Kanamycin	-	Based upon CBP-FTHH-RT, cloned in house (NC)
pUC19	Ampicillin	-	Addgene (#50005)
MYB-RT	Ampicillin	-	Based upon Addgene #50005, cloned in house (SB)
MYB-Halotag-RT	Ampicillin	-	Based MYB-RT, cloned in house (SB)
MYB-HTHAS-RT	Ampicillin	-	Based upon MYB-RT, cloned in house (NC)
HsCBPv1-GFP	Kanamycin	Neomycin	Based upon pAcGFP-N1, cloned in house (KG)
HsCBPv1-Halotag	Kanamycin	Neomycin	Based upon pAcGFP-N1, cloned in house (KG)

2.1.3 DNA oligonucleotides

Table 2.4: Table of DNA oligonucleotides

Name	Purpose	Sequence	Target
CBP_C-T_F1	CBP C-terminal amplification by PCR	ACCCCAACATGGCGAGTATG	CBP C-terminus
CBP_C-T_R1	CBP UTR amplification by PCR	CGGAAGTCGCAGTTCCATCT	CBP UTR
CBP-C-T_Fwd2	CBP C-terminal amplification by PCR	CAGATCGCCACGTCCCTTAG	CBP C-terminus
MYB_C-T_F1	MYB C-terminal amplification by PCR	CAATTCCAGTGCGTGCATGA	MYB C-terminus
MYB_C-T_R1	MYB C-terminal amplification by PCR	TTGGTGCTGCTCTCAACTGT	MYB C-terminus
U6_promoter-fwd	Sequencing	GAGGGCCTATTTCCCATGATT	U6 promoter
pUC19_pre MYB_fwd	Sequencing	CCATGATTACGCCAAGCTTG	pUC19
MYB_pre HT_fwd	Sequencing	GCATTCTCAGCCCGGACGCT	MYB-RT
codon-optimised_FTH-Fwd	Within tag amplification by PCR to validate clonal lines	GGCAGCGACTATAAGGATGA	Codon optimised Flag-TEV-6xHis tag
codon-optimised_FTH-Rev	Within tag amplification by PCR to validate clonal lines	TCATCCTTATAGTCGCTGCC	Codon optimised Flag-TEV-6xHis tag

2.1.4   RNA sequences

Table 2.5: Table of gRNA sequences

Name of sequence	Purpose	Sequence	Target	Format
CBP C-terminus gRNA	gRNA for CRISPR tagging	CACGCTAGAGAAGTTT GTGG	CBP C-terminus	RNA oligo and plasmid
MYB C-terminus gRNA	gRNA for CRISPR tagging	TGCATTCTCAGCCCGG ACGC	MYB C-terminus	Plasmid
Scramble gRNA	Control gRNA	GCTGATCTATCGCGGT CGTC	Non-targeting control	Plasmid
MYB KO gRNA 1	gRNA for CRISPR KO of MYB	CAGCATATATAGCAGT GACG	MYB exon 2	Plasmid
MYB KO gRNA 2	gRNA for CRISPR KO of MYB	GAAGCAGCCCATCATA GTCA	MYB exon 2	Plasmid
MYB KO gRNA 3	gRNA for CRISPR KO of MYB	AGTCTGGAAAGCGTCA CTTG	MYB exon 2	Plasmid

2.2   DNA and analysis techniques

2.2.1   Buffers, reagents and media

50X TAE: 2 M Tris base, 1 M glacial acetic acid, 50 mM Ethylenediaminetetra-acetic acid (EDTA)

Luria Broth (LB) media: 25 g of LB granules (Melford) dissolved in 1 L Milli-Q H_2O, then sterilised by autoclave at 121ºC for 20 minutes.

LB agar: 32 g of Luria Broth (LB) granules (Melford) dissolved in 1 L Milli-Q H_2O, then sterilised by autoclave at 121ºC for 20 minutes.

SOC media: 2% (wt/vol) Tryptone (Melford), 0.5% (wt/vol) Yeast Extract (Melford, 0.05% (wt/vol) NaCl (Melford), 2.5 mM KCl (Melford), 10 mM MgCl$_2$(Melford). Stocks prepared and sterilised by autoclave at 121°C for 20 minutes prior to addition of 20 mM glucose (Melford), which had been filter sterilised.

2.2.2 Molecular biology kits

Plasmid purification (small scale): Monarch Plasmid Miniprep kit (NEB).

Plasmid purification (medium scale): Plasmid Plus Midi kit (QIAGEN)

PCR product purification: Monarch PCR DNA Cleanup Kit (NEB)

Gel purification: QIAquick Gel Extraction Kit (QIAGEN)

2.2.3 Polymerase Chain Reaction

The Polymerase Chain Reaction (PCR) was carried out using two different polymerase enzymes and therefore PCR conditions. PfuUltra II Fusion High-Fidelity Polymerase (Agilent) was used for higher fidelity reactions and to produce higher quantities of DNA. Taq DNA polymerase (NEB) was used for screening of clones and for reactions not requiring larger quantities of DNA.

Higher fidelity PCR was done in a 50 μl reaction volume. For clone screening PCR using Taq was done in 25 μl reactions to allow scaling and other reactions were done in 50 μl reactions:

Table 2.6: Composition of 50 μl PCR using PfuUltra II Fusion High-Fidelity Polymerase

Component	50 μl reaction
H_2O	To 50 μl
10X PfuUltra II reaction buffer (Agilent)	5 μl
DMSO (Sigma)	2.5 μl
100 μM forward primer (IDT, as detailed in 2.4)	0.25 μl
100 μM reverse primer (IDT, as detailed in 2.4)	0.25 μl
10 mM dNTP mix (Roche)	5 μl
PfuUltra II Fusion High-Fidelity Polymerase (Agilent)	0.5 μl
DNA template	1 μl

Table 2.7: Composition of 25 μl and 50 μl PCR using Taq DNA polymerase

Component	25 μl reaction	50 μl reaction
H_2O	To 25 μl	To 50 μl
10X Thermopol buffer (NEB)	2.5 μl	5 μl
DMSO (Sigma)	1.25 μl	2.5 μl
10 μM forward primer (IDT, as detailed in 2.4)	0.5 μl	1 μl
10 μM reverse primer (IDT, as detailed in 2.4)	0.5 μl	1 μl
10 mM dNTP mix (Roche)	0.5 μl	1 μl
Taq DNA polymerase (NEB)	0.125 μl	0.25 μl
DNA template	0.5 μl	1 μl

A ProFlex PCR thermocycler (Applied Biosystems) was used to run the reaction:

Table 2.8: PCR cycle conditions for Pfu PCR

Cycle step	Temperature	Time	Cycles
Initial denaturation	95°C	5 mins	1
Denaturation	95°C	45 secs	
Annealing	Primer dependent	45 secs	30
Extension	72°C	30 sec/kb	
Final extension	72°C	15 min	1

Table 2.9: PCR cycle conditions for Taq PCR

Cycle step	Temperature	Time	Cycles
Initial denaturation	95°C	5 mins	1
Denaturation	95°C	45 secs	
Annealing	Primer dependent	45 secs	30
Extension	72°C	1 min/kb	
Final extension	72°C	15 min	1

2.2.4   DNA agarose gel electrophoresis

UltraPure Agarose (ThermoFisher) was combined with 1X TAE to a final concentration of either 1% or 1.5% (wt/vol), heated, then supplemented with SYBR Safe DNA Gel Stain (ThermoFisher) to a 1:10,000 dilution. After pouring and setting, the gel was placed in a gel tank of appropriate size, which was filled with 1X TAE. Samples

were loaded into wells with 6X Gel Loading Dye (NEB), with appropriate markers for reference (QuickLoad 100 bp, 1 kb or 1 kb plus DNA ladder, NEB) and run at 80-120 V for 30-60 minutes. Gels were visualised using blue light and imaged using a G:BOX Chemi-XRQ (Syngene).

2.2.5   Bacterial cell transformation

Chemically competent Stbl3 *E. coli* (ThermoFisher) cells were thawed on ice. 1 pg-100 ng plasmid was added to 50 μl Stbl3 cells and incubated on ice for 30 minutes, heat shocked at 42℃ for 30 seconds and returned to the ice for 2 minutes. 950 μl SOC media was added and cells were incubated at 37℃ with shaking at 300 rpm in a table top shaker for 1 hour. Cells were spun down at 400 rcf for 5 minutes, with 900 μl of the media being discarded and the cell pellet resuspended in the remaining volume. This was then plated onto LB agar plates containing antibiotic corresponding to the resistance marker in the plasmid being transformed. Either 100 μg/ml ampicillin or 50 μg/ml kanamycin was used. Plates were incubated overnight at 37℃ and examined for colony growth the next morning.

2.2.6   Plasmid preparation and cloning

2.2.6.1   Small scale plasmid purification

To generate small amounts of a desired plasmid, often to screen colonies after cloning, colonies were selected and grown up into 5 ml LB cultures containing antibiotic corresponding to the resistance marker in the plasmid being transformed. Either 100 μg/ml ampicillin or kanamycin was used. These cultures were grown up overnight at 37℃ with shaking at 180 rotations per minute (rpm) on a shaking platform. The next morning cultures were harvested and plasmids were purified using the Monarch Plasmid Miniprep kit (NEB) according to the manufacturer's instructions.

2.2.6.2   Medium scale plasmid purification

To generate larger amounts of plasmid, often for use in larger scale experiments such as transfections, larger cultures were required. Colonies were selected from plates and grown up into 100 ml LB cultures containing antibiotic corresponding to the resistance

marker in the plasmid being transformed. Either 100 μg/ml ampicillin or 50 μg/ml kanamycin was used. These cultures were grown up overnight at 37^oC with shaking at 180 rpm on a shaking platform. The next morning cultures were harvested and plasmids were purified using the Plasmid Plus Midi Kit (QIAGEN) according to the manufacturer's instructions.

2.2.6.3   Cloning of pX458 based gRNA plasmids

Complementary forward and reverse sequences for the gRNAs were ordered as short single stranded DNA oligonucleotides (IDT). The forward sequence contained a 5' CACC prior to the forward sequence, with the reverse sequence containing a 5' AAAC, thus generating overhangs for BbsI digestion. gRNA sequences detailed in Table 2.5.

gRNA oligos were annealed to form a double stranded sequence, with 1 μl of each 100 μM oligo added to 5 units of T4 polynucleotide kinase (PNK) (NEB) with 10X T4 ligation buffer (NEB) containing Adenosine Triphosphate (ATP) which was essential for the reaction. For annealing, the reactions were placed in a thermocycler and heated to 37^oC for 30 minutes, then heated at 95^oC for a further 5 minutes and cooled to 25^oC at a ramp rate of -5^oC per minute. Annealed gRNAs were then diluted 1:100 in sterile H_2O in preparation for the ligation reaction.

pX458 plasmid was digested with 1 unit of BbsI-HF (NEB) per μg DNA, with 10X NEB Cutsmart buffer, at 37^oC for 15 minutes and then heat inactivated at 65^oC for 20 minutes. Digested plasmid was run on a 1% agarose gel (as described in section 2.2.4), and the desired product was extracted from the gel using the QIAGEN gel extraction kit according to the manufacturer's instructions. This now linearised vector was dephosphorylated to prevent recircularisation of the plasmid using Calf Intestinal Phosphatase (CIP, NEB). 5 units of CIP were used per μg DNA initially digested by BbsI-HF for 30 minutes at 37^oC, followed by purification using the Monarch PCR purification kit, according to the manufacturer's instructions (NEB). DNA concentration was measured using the Qubit broad range dsDNA kit (Thermofisher).

For the ligation reaction, 50 ng vector was combined with 1 μl diluted oligo, alongside 400 units T4 ligase and 10X T4 ligase buffer (NEB). Optimal ligation conditions were 30 minutes at room temperature as opposed to overnight, and then ligation reactions were heat inactived at 65 ^oC for 10 minutes.

2 μl of each ligation reaction was transformed into *E. coli* (as described in section 2.2.5) and plated onto LB agar plates with ampicillin resistance. Colonies were picked the following day and overnight cultures were grown up for minipreps (as described in section 2.2.6.1). Plasmids from the minipreps were then sent for Sanger sequencing to confirm correct insertion of the gRNA sequence. For pX458 gRNA cloning, U6_promoter-fwd primer (Table 2.4) was used for Sanger sequencing.

2.2.6.4 Cloning of MYB RT plasmids

Cloning of MYB homology arms into pUC19 and successive cloning of the Halotag into this donor repair template was done by Sophie Ball.

Different strategies were employed to generate the required homology directed repair (HDR) templates for CRISPR/Cas9 based tagging. For the generation of a HDR template to tag MYB, a donor plasmid was first generated containing the homology arm region with a mutated protospacer adjacent motif (PAM) site only. This provided a master plasmid for any further tag to be cloned into.

A DNA fragment containing MYB homology arms and the mutated PAM sequence (IDT) was cloned into pUC19 (Table 2.3) using the NEBuilder HiFi DNA Assembly kit (NEB). Primers were designed using the NEBuilder (NEB) tool, and are described in Table 2.4, to amplify and linearise the vector at the correct site for the insert to be cloned in. The MYB homology arms fragment was amplified by primers with the addition of sequences that overlap with the vector integration sites, allowing ligation into the plasmid.

The vector and insert sequences were amplified by PCR using the designed primers (as described in section 2.2.3) and all PCR product was run on a 1% agarose gel (as described in section 2.2.4). The desired products were extracted from the gel using the QIAGEN gel extraction kit according to the manufacturer's instructions. DNA concentration was measured using the Qubit broad range dsDNA kit (Thermofisher) according to the manufacturer's instructions.

For the NEBuilder HiFi DNA Assembly reaction 50 ng vector was mixed with insert at a 2:1 ratio (calculations done using NEBioCalculator tool, NEB). The NEBuilder HiFi DNA assembly kit (NEB) was used according to the manufacturer's instructions.

2 μl of each ligation reaction was transformed into *E. coli* (as described in section

83

2.2.5) and plated onto LB agar plates with ampicillin resistance. Colonies were picked the following day and overnight cultures were grown up for minipreps (as described in 2.2.6.1). Plasmids produced from the minipreps were sent for Sanger sequencing to determine if the correct plasmid had been generated successfully. For MYB homology arms cloning, pUC19_preMYB_fwd primer (Table 2.4) was used for Sanger sequencing, and the successfully generated plasmid denoted MYB-RT.

MYB-RT was then used as a vector for insertion of a Halotag sequence, with the final plasmid to be used as a HDR template for CRISPR/Cas9 tagging of MYB with a Halotag. Cloning was carried out using the NEBuilder HiFi DNA Assembly kit (NEB) as described above, but with a Halotag DNA fragment (IDT). MYB_preHT_fwd (Table 2.4) was used for Sanger sequencing, and the successfully generated plasmid denoted MYB-Halotag-RT.

2.2.6.5 Cloning of MYB-HTHAS HDR template plasmid

For generation of a HDR template to tag MYB with a 6xHis-TEV-HA/Strep (HTHAS) tag a restriction enzyme based cloning approach was used.

A plasmid containing the desired HTHAS with flanking homology arms sequence was obtained from ThermoFisher to use as an insert and MYB-RT plasmid was used as the vector. Both the vector and insert contained *Bbs*I and *Hind*III restriction sites.

5 μg of each insert and vector plasmid was digested with 50 units of *Bbs*I-HF (NEB) and 100 units of *Hind*III-HF (NEB) in 10X Cutsmart buffer (NEB) for 1 hour at 37^oC, before heat inactivation at 80^oC for 20 minutes. All digested product was run on a 1% agarose gel (as described in section 2.2.4). The desired insert and vector products were excised from the gel and purified using the QIAGEN gel extraction kit and DNA concentration was measured using the Qubit double stranded DNA broad range (dsDNA BR) kit (Thermofisher).

To prevent the vector from recircularising, the linearised vector was dephosphorylated using CIP (NEB) and purified using the Monarch PCR Purification kit (NEB) according to the manufacturer's instructions.

For the ligation reaction 50 ng vector was combined with insert at varying ratios, including 1:1, 1:2, 1:3 and 1:5 (calculations done using NEBioCalculator tool, NEB) alongside 400 units T4 ligase and 10X T4 ligase buffer (NEB). Optimal ligation

conditions were 30 minutes at room temperature as opposed to overnight, and then ligation reactions were heat inactived at 65 ^{o}C for 10 minutes. 2 μl of each ligation reaction was transformed into *E. coli* (as described in section 2.2.5) and plated onto LB agar plates with ampicillin resistance. Colonies were picked the following day and overnight cultures were grown up for minipreps (as described in section 2.2.6.1). Plasmids produced from the minipreps were sent for Sanger sequencing to confirm correct insertion of the gRNA sequence. To check the correct insertion of the HTHAS tag, CBP-C-T_Fwd2 primer (Table 2.4) was used for Sanger sequencing.

2.3 Mammalian cell culture techniques

2.3.1 Jurkat cell culture

Jurkat cells, Clone E6-1 (American Type Culture Collection (ATCC), TIB-152) were cultured in complete Roswell Park Memorial Institute (RPMI) 1640 media (ThermoFisher)(10% Foetal Bovine Serum (FBS, ThermoFisher), 1% Penecillin/Streptomycin (Pen-Strep, ThermoFisher)) at 37℃ in 5% CO_2.

2.3.2 Passaging of Jurkats

Jurkat cells were cultured in 13 ml complete RPMI (10% FBS, 1% Pen-strep) at 37℃ in 5% CO_2. To passage cells, 12 ml of media were removed and discarded, then 12 ml fresh complete media was added (10% FBS, 1% Pen-strep). To passage cells prior to an experiment, all cells were centrifuged at 100 rcf for 5 minutes. Centrifugation at such a low speed allowed for healthier cells to pellet down, with dead cells remaining in the supernatant allowing for enrichment of a healthier population for experiments. After removal of the supernatant, the pellet was resuspended in 10 ml complete RPMI, and 1 ml of this cell suspension was transferred to a T-75 flask (Starstedt) containing 12 ml pre-warmed complete RPMI.

2.3.3 Cryopreservation and thawing of Jurkats

2.3.3.1 Cryopreservation

For cryopreservation, Jurkats were resuspended in 1 ml freezing media containing RPMI, 10% Dimethyl Sulfoxide (DMSO, Sigma), FBS to a final concentration of 20% (vol/vol), at a number of 2-4x10^6cells/ml. 1 ml aliquots were transferred into labelled cryovials (Sarstedt), and then placed in a Mr. Frosty Freezing Container (ThermoFisher) and stored at -80℃ for at least 24 hours. Cryovials were then transferred to storage in liquid nitrogen (LN$_2$) at -196℃.

2.3.3.2 Thawing

Upon thawing, cryovials were removed from LN$_2$ storage and thawed rapidly in a water bath heated to 37℃. The 1 ml of thawed cells was added to a 50 ml centrifuge tube containing 9 ml complete RPMI and centrifuged at 200 rcf for 5 minutes. The main purpose of this was to remove any DMSO from the cell media. After removing the supernatant, the pellet was resuspended in 13 ml complete RPMI and transferred into a new T-75 flask.

2.3.4 Counting of Jurkat cells

Jurkat cells were counted using the Countess cell counter (ThermoFisher). Cells were resuspended to a single cell suspension so they were not clumped together. 10 μl of cell suspension was mixed with 10 μl Trypan Blue stain (Sigma) and transferred to a Countess slide (ThermoFisher). After inserting the slide into the Countess, cell numbers per ml were calculated and corrected for dilutions.

2.3.5 Transfection of Jurkats using the Neon (ThermoFisher) system

2.3.5.1 Transfection of Jurkats on a smaller scale

Smaller scale transfections were done using the Neon Transfection System 10 μl kit (ThermoFisher). Jurkat cells were counted and 2 x 10^5 cells per reaction were span down at 200 rcf and washed in 1 ml Phosphate Buffered Saline (PBS, ThermoFisher).

Cells were then resuspended in 10 μl buffer R (ThermoFisher), and aliquotted per reaction. 1 μl plasmid or reagent to be transfected was added to each reaction sample. 10 μl of each sample of Jurkat cells were transfected with the electroporation settings 1600 V, 10 ms pulse width and 3 pulses. Each reaction was plated into a pre-warmed 48 well plate, with each well containing 190 μl RPMI (10% FBS), ensuring this was antibiotic free. 24 hours post-transfection media containing antibiotic was added and cells were then expanded into larger volumes for screening or further experiments as required.

2.3.5.2 Transfection of Jurkats on a larger scale

Larger scale transfections were done using the Neon Transfection System 100 μl kit (ThermoFisher). Jurkat cells were counted and 2 x 10^6 cells per reaction were span down at 200 rcf and washed in 10 ml PBS. Cells were then resuspended in 100 μl buffer R (ThermoFisher), and aliquotted per reaction. 10 μl plasmid or reagent to be transfected was added to each reaction sample. 100 μl of each sample of Jurkat cells were transfected with the electroporation settings 1350 V, 10 ms pulse width and 3 pulses. Each reaction was plated into a pre-warmed T-25 flask (Sarstedt), with each flask containing 5 ml RPMI (10% FBS), ensuring this was free of any antibitioc. After 24 hours media containing antibiotic was added and cells were then expanded into larger volumes for screening or further experiments as required.

2.4 Endogenous tagging of proteins

2.4.1 Oligo and RNA design

2.4.1.1 gRNA design

Using the University of California, Santa Cruz (UCSC) Genome Bowser® the DNA sequence of the C-terminal region of the protein of interest was identified and downloaded into an ApE® file (`https://jorgensen.biology.utah.edu/wayned/ape/`). crisprRNA (crRNA) sequences and Protospacer Adjacent Motif (PAM) sites were designed and identified through the UCSC Clustered Regularly Interspersed Palindromic Repeats (CRISPR) targets tool, to target the C-terminus of the target protein. cr-

RNAs were evaluated by UCSC based on the MIT specificity score and the Doench et al. 2016 score (Doench et al. 2016; Haeussler et al. 2016). The two measures complemented each other, as the MIT score encapsulated the calculated off-target effects of the crRNA, whereas the Doench et al. 2016 score reflected the efficiency of a crRNA and how well it was expected to cleave at the target site (Doench et al. 2016; Haeussler et al. 2016).

After selection in UCSC, crRNAs were further evaluated using the Evaluation tool on the E-CRISP site (`http://www.e-crisp.org/E-CRISP/reannotate_crispr.html`). A high specificity score, low annotation score and a medium to high efficacy score indicated a crRNA that was likely to have high efficiency to guide Cas9 to the correct site and induce cleavage correctly and simultaneously unlikely to cleave at off-target loci.

2.4.1.2 Single stranded oligo-deoxyribonucleotide (ssODN) repair template design

Single stranded oligo-deoxyribonucleotide (ssODN) sequences were designed using ApE. The double-strand break (DSB) site was identified as 3 nucleotides 5' of the PAM site. From this, homology arms of 50 nucleotides (nt) flanking the desired insertion sequence were identified, creating the ssODN.

Within the ssODN sequence a silent mutation was introduced to the PAM sequence so that once the edit had been copied in the PAM site would no longer exist, however there would be no change to the amino acid sequence of the protein. Disruption of the PAM site would prevent Cas9 from recognising the same site multiple times and reinserting edits sequentially. Where it was not possible to create silent mutations, a different crRNA was chosen that did allow for PAM site manipulation and maintained the amino acid sequence.

2.4.1.3 Plasmid repair template design

When ssODN templates did not appear to work, a strategy of using a plasmid based repair template with larger homology arms was explored. This was especially important when the size of the tag was increased to include the Halotag. The previous method of identifying the C-terminus, crRNA, PAM and DSB site was followed, how-

ever the plasmid template contained homology arms of approximately 600 nt flanking the desired insertion sequence. The strategy to create silent mutations in the PAM site was also followed in this type of repair template.

2.4.2 Transfecting gRNA with Cas9 enzyme into Jurkat cells

2.4.2.1 Transfecting Jurkat cells with a sgRNA:Cas9 ribonucleoprotein and single stranded oligodeoxyribonucleotide

To form the cr:tracrRNA duplex, 200 μM crRNA (IDT) and 200 μM tracrRNA (IDT) were mixed in nuclease-free TE (10 mM Tris, 0.1 mM EDTA) buffer (IDT) to a final concentration of 44 μM and heated to 95℃ for 5 min. This formed the single guide RNA (sgRNA).

To form the ribonucleoprotein (RNP) complex, 22 pmol crRNA:tracrRNA duplex was incubated with 18 pmol recombinant Cas9 protein (IDT) at room temperature for 20 minutes.

As a control to validate the gRNA, the RNP complex only was transfected into Jurkat cells using the Neon Transfection system 10 μl kit (as described in section 2.3.5.1). The sgRNA:Cas9 RNP was mixed with 10.8 μM Cas9 Electroporation Enhancer (CRISPR-Cas9 control kit, IDT) and the cell suspension and was then electroporated.

To carry out the CRISPR/Cas9 mediated repair Jurkat cells were also transfected with the sgRNA:Cas9 RNP as described above, with the addition of 10.8 μM ssODN (IDT).

2.4.2.2 Transfecting Jurkat cells with a GFP-Cas9 and plasmid repair template

Jurkat cells were firstly transfected using the Neon Transfection system 100 μl kit (as described in section 2.3.5.2), with 11 μl desired pX458 gRNA plasmid to validate the gRNA.

Once the gRNA had been validated, to carry out the CRISPR/Cas9 mediated repair, Jurkat cells were transfected using the Neon Transfection system 100 μl kit (as described in section 2.3.5.2), with 1 μl desired pX458 gRNA plasmid and 10 μl

desired repair template plasmid (as detailed in Table 2.3).

2.4.3 gRNA validations

2.4.3.1 Genomic DNA extraction

For gRNA validation experiments, genomic DNA (gDNA) was extracted using Quick-Extract solution (Lucigen) according to the Manufacturer's instructions.

2.4.3.2 PCR

PCR carried out to amplify the edited C-terminal region of a targeted protein was set up as detailed in section 2.2.3, using Pfu high fidelity enzyme and primers detailed in Table 2.4. Reactions were set up using gDNA from edited and unedited cells as a control.

2.4.3.3 T7 Endonuclease 1 assay

PCR products amplifying the C-terminal region of CBP in both edited and unedited cells was cleaned up using the Monarch PCR cleanup kit (NEB) according to the manufacturer's instructions. DNA concentrations were subsequently measured using the Qubit dsDNA BR kit according to the manufacturer's instructions.

2 µl 10X NEB buffer 2 (NEB) was added to 200 ng DNA and made up to a final reaction volume of 19 μl with H$_2$O. PCR products were then denatured to single strands by heating at 95^oC for 5 minutes. DNA strands were then hybridised by decreasing temperature from 95^oC to 85^oC at a ramp rate of -2^oC/second, then from 85^oC to 25^oC at a ramp rate of -0.1^oC per second.

1 μl T7 Endonuclease 1 (T7E1) enzyme (NEB) was added to each reaction and incubated at 37^oC for 15 minutes. The reaction was quenched with the addition of 1.5 μl of 0.25 M EDTA.

The total reaction volume was run on a 1% agarose gel (as described in section 2.2.4).

2.4.3.4   Inference of CRISPR Edits analysis

To analyse Sanger sequencing data of CRISPR edited DNA sequences, firstly PCR products, as set up as detailed in section 2.4.3.2 were sent for Sanger sequencing and .ab1 files were obtained. These were uploaded to the Inference of CRISPR Edits (ICE, Synthego) tool (`https://ice.synthego.com/#/`, Conant et al. 2022) to compare control vs edited samples.

2.4.4   Flow Cytometry and Fluorescence Activated Cell Sorting

All Fluorescence Activated Cell Sorting (FACS) was carried out at The Medical School Flow Cytometry Core Facility (The University of Sheffield), with technical support being used for training purposes and assistance with gating populations.

2.4.4.1   Preparation of cells for FACS

Prior to cell sorting, cells at the lowest passage possible were required whilst simultaneously being in a healthy growth phase. Any plasmids or reagents that needed to be transfected that would be used for sorting, for example containing fluorescent markers, were transfected 48 hours prior to sorting (as described in sections 2.3.5 and 2.4.2).

On the day of sorting cells were counted and washed twice in PBS. If carrying out a bulk sort to obtain polyclonal lines, cells were resuspended to a final volume of 10 x 10^6 cells in 1 ml PBS. If cells were to be single cell sorted they were resuspended to a final concentration of 5 x 10^6 cells per ml PBS. It was ensured that all suitable controls were also prepared, including a WT cell line and a positive control for fluorescence if required.

Spare aliquots of media (RPMI with 10% FBS, 1% P/S) and PBS were prepared.

If single cell sorting, 96 well plates were prepared with 100 μl pre-conditioned media per well. Preparation of pre-conditioned media involved removal of media from a healthy population of WT Jurkats that had proliferated for 24 hours. Cells were harvested by centrifugation at 400 rcf, media was then aspirated and filtered using a 0.45 μm filter to remove residual cells.

For bulk sorts collection tubes were provided by the Flow Cytometry Core and 1 ml complete media was added to the tube prior to sorting.

To transport to the FACS facility, all tubes and plates were sealed with Parafilm (Sigma) to minimise contamination and placed on ice.

2.4.4.2 GFP bulk sort

Sorting was done using a FACSMelody Cell Sorter (BD Biosciences).

Cells had been transfected with a GFP-Cas9 and gRNA encoding plasmid and a plasmid repair template (as described in 2.4.2.2) 48 hours prior to sorting.

Forward scatter versus side scatter of the WT cells was first measured to set the gates, to obtain single cells. Then GFP positive cells were used to set a gate to distinguish GFP positive cells from the WT, by measuring fluorescence at 488 nm. After this, GFP samples were sorted, where all cells that were GFP positive for each sample were collected into a single collection tube with 1 ml media added to help sustain the cells.

Once finished, the collection tube was placed on ice and returned to the tissue culture facility. Cells were centrifuged at 400 rcf to remove PBS, then resuspended in RPMI (10% FBS, 1% P/S) and finally placed in a flask of appropriate size depending on cell numbers obtained.

Polyclonal cells after sorting were cultured as described in section 2.3.2.

2.4.4.3 Single cell sorting with no fluorescence

During some single cell sorting experiments, such as during the generation of clonal MYB_{HTHAS}:CBP_{FTHH} cell lines (Section 5.4), there were no fluorescent labels or markers in the edit to utilise in single cell sorting. Cells were gated using forward versus side scatter to obtain a stream of single cells and cells were plated in one cell per well across four 96 well plates.

2.4.4.4 Single cell sorting using TMR labelling

For single cell sorting of CBP_{FTHH} cells, where CBP was expressing a Halotag, cells were labelled with a fluorescent Halotag ligand named TMR (Promega) for single cell sorting (Section 4.4.3.1).

10×10^6 cells expressing a Halotag were counted and resuspended in 4 ml complete RPMI and TMR ligand was added to a final concentration of 100 nM. Cells were left to incubate at $37^{o}C$ for 1 hour. Cells were then centrifuged at 200 rcf for 5 minutes and resuspended in 13 ml complete RPMI to wash out any free ligand. Cells were then left to incubate at $37^{o}C$ for a further 2 hours until preparation for FACS.

When labelling with TMR, it was necessary to have a Wild Type (WT) control with no TMR ligand added, as well as a WT with added TMR as a control to assess levels of background labelling to help set gates.

For cell sorting, fluorescence was measured at 561 nm. TMR labelled CBP_{FTHH} cells that fell as expected within the gates set were sorted one cell per well, with 96 well plates set up as described in section 2.4.4.1.

Upon return to the tissue culture facility, plates were placed in an incubator at $37^{o}C$, 5% CO_2.

2.4.4.5   Culturing single cell sorted clones

96 well plates were left to expand for approximately 2 weeks at $37^{o}C$, 5% CO_2, with daily monitoring taking place from the second week. Once single cells had expanded into a visible clump they were resuspended by gentle pipetting up and down to dissociate the clump and encourage expansion of the cells. Once the media began to visibly yellow, media was topped up to a final volume of 200 μl in appropriate wells. As cells became confluent in each well, they were expanded as necessary to larger well sizes and media increased in volume. Once confluent in a 12 well plate, cells were ready for PCR screening.

2.4.4.6   PCR screening of single cell sorted clones

Once clones were confluent in a 12 well plate in 1 ml media, 100 μl cell suspension was harvested for DNA extraction. DNA was extracted using the same methods as used for gRNA validation (as described in section 2.4.3.1).

DNA was amplified using PCR (as described in section 2.2.3) using Taq DNA polymerase enzyme (NEB) in a 25 μl reaction volume to allow for large numbers of clones to be screened at once (Table 2.7). Primers that amplify DNA surrounding the C-terminus were used as detailed in Table 2.4.

PCR products were ran on a 1% agarose gel (as described in section 2.2.4) and the size of bands were used to determine if the tag had been successfully inserted, and whether the clone was homozygous or heterozygous. Any clones of interest that indicated the tag had been successfully inserted were also amplified using PCR but with a forward primer to amplify within the tag. If this was deemed successful, both the C-teminal and within tag PCR products were sent for Sanger sequencing, using the forward primer of the relevant PCR product.

Successful clones were then expanded into larger well plates and flasks and stocks were frozen down (as described in section 2.3.3.1).

2.5 Protein analysis techniques

2.5.1 Western blot

2.5.1.1 Cell lysis

Cells were harvested by centrifugation at 400 rcf and washed in PBS (Sigma). The cell pellet was resuspended in 50 μl RIPA lysis buffer (50 mM Tris-HCl pH 8 at $4^{o}C$ (Sigma), 100 mM NaCl (Melford), 2 mM MgCl$_2$ (Fisher), 1% Triton X-100 (Sigma), 0.1% sodium deoxycholate (Sigma), 0.1% sodium dodecyl sulphate (SDS, Sigma), 1 mM Dithiothreitol (DTT, Melford), 100X Halt Protease and Phosphatase inibitor cocktail (Fisher), with 10 mM sodium butyrate (Sigma) added if looking at CBP).

250 units benzonase (Insight Biotech) was added directly to each cell pellet and incubated at $37^{o}C$ for 20 minutes until viscous chromatin had been digested. Cell debris was pelleted by centrifugation at 20,000 rcf for 10 minutes, and cell lysate concentration was calculated by Pierce BCA assay (ThermoFisher) according to the manufacturer's instructions.

2.5.1.2 SDS-PAGE

50 μg protein per sample was prepared with the addition of NuPAGE LDS Sample Buffer (4X, ThermoFisher) and NuPAGE Sample Reducing Agent (10X, ThermoFisher) and heated to $70^{o}C$ for 10 minutes.

The XCell SureLock Mini-Cell Electrophoresis System (ThermoFisher) tank was

used. The gel tank was set up by placing and securing the gel, removing the comb, filling with running buffer, cleaning out the wells and adding 500 μl NuPAGE Antioxidant (ThermoFisher) to maintain the protein samples in their reduced state.

To aid visualisation of proteins with differing molecular weights, appropriate combinations of varying gel types and running buffers were used. If only CBP was being examined, a 3-8% Tris-Acetate gradient gel (ThermoFisher) was chosen with Tris-Acetate running buffer (ThermoFisher) as this gel type and running buffer allows large molecular weight proteins to run further into the gel. This was ran at 100 V for 120 minutes.

If lower molecular weight proteins such as the transcription factors within the CBP-TAL1 complex were to be examined, a 4-12% Bis-Tris gradient gel (ThermoFisher) was chosen with MOPS running buffer (ThermoFisher), as this combination resulted in proteins of 30-200 kDa to be well distributed through the centre of the gel. This gel was ran at 120 V for 100 minutes.

2.5.1.3 Transfer

The Trans-Blot Turbo transfer system (Bio-Rad) was used to transfer the proteins from the gel to a membrane. Trans-Blot Turbo Mini 0.2 μm Nitrocellulose Transfer Packs (Bio-Rad) were used, and proteins were transferred at 1.3 Amp, 25 Volts for 25 minutes.

2.5.1.4 Antibody incubation

After transfer, the membrane was blocked for at least 40 minutes in blocking solution consisting of 5% milk (Marvel) in 1X Tris buffered saline with Tween (TBS-T, Sigma). If required the membrane was trimmed or cut into desired sized fragments and incubated with a primary antibody of choice diluted in blocking solution (primary antibodies described in Table 2.10) overnight on a rocking platform at $4^{o}C$. The following morning the primary antibody was washed off with three 5 minute TBS-T washes, and then the membrane was incubated with a secondary antibody (secondary antibodies described in Table 2.11) with binding activity of the correct host species of the primary antibody. The secondary antibody used was HRP-conjugated to allow visualisation.

Table 2.10: Primary antibody dilutions for western blotting

Target	Source	Western blotting dilutions
CBP	Cell signalling (7389S)	1:1,000
c-MYB	Abcam (ab45150)	1:1,000
E2A	Cell Signalling (12258S)	1:1,000
GATA3	Proteintech (10417-1-AP)	1:500
RUNX1	Proteintech (19555-1-AP)	1:500
TAL1	Santa Cruz (sc-393287)	1:200
TAL1	Proteintech (55317-1-AP)	1:500
TCF12/HEB	Proteintech (14419-1-AP)	1:500
FLAG tag	Proteintech (66008-3-Ig)	1:1,000
FLAG tag (M2)	Sigma (F3165)	1:1,000
HA tag	Proteintech (51064-2-AP)	1:5,000
GFP tag	Millipore (G6539)	1:1,000
Halotag (polyclonal)	Promega (G928A)	1:1,000
Halotag (monoclonal)	Promega (G9211)	1:1,000
GAPDH	Proteintech (60004-1-Ig)	1:10,000

Table 2.11: Secondary antibody dilutions for western blotting

Target	Source	Western blotting dilutions
HRP conjugated Goat Anti-Mouse IgG (Heavy and Light chain)	Proteintech (SA00001-1)	1:10,000
HRP conjugated Mouse Anti-Rabbit IgG (Light chain only)	Stratech (211-032-171)	1:20,000

2.5.1.5 Membrane visualisation

SuperSignal West Pico Chemiluminescent Substrate (ThermoFisher) was used as a HRP substrate and was incubated with the membrane for 5 minutes before imaging using a G:BOX Chemi-XRQ imager (Syngene).

2.5.2 Co-Immunoprecipitation (Co-IP)

In order to prevent protein degradation, all steps were done in succession until membrane incubation with primary antibody.

2.5.2.1 Cell lysis

200×10^6 Jurkat cells were harvested per co-immunoprecipitation (Co-IP) by centrifugation at 400 rcf for 5 minutes and the pellet was resuspended in 10 ml PBS. Cells were further centrifuged at 2,500 rcf for 5 minutes and the pellet was resuspended in $300\mu l$ CoIP NET buffer (50 mM Tris-HCl pH 7.3, 150 mM NaCl, 0.1 mM EDTA, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 2 mM $ZnCl_2$ (Sigma), 0.1% NP-40 or IGEPAL CA-630 (Sigma), 100X Halt protease inhibitor, 10 mM Sodium Butyrate).

To digest viscous chromatin 10X Turbo DNase buffer (ThermoFisher) was added directly to the pellet as well as 3 μl Turbo DNase (ThermoFisher) and incubated at 37^oC for 30 minutes.

To clear the cell lysate away from chromatin and debris, lysate was centrifuged at 20,000 rcf for 20 minutes and protein concentration of cell lysate was calculated by Pierce BCA assay (ThermoFisher) according to the manufacturer's instructions.

2.5.2.2 Immunoprecipitation

Prior to immunoprecipitation (IP), 50 μl Protein G dynabeads (ThermoFisher) per IP reaction were blocked with three washes of 0.5% (wt/vol) Bovine Serum Albumin (BSA, Sigma) in PBS, to block non-specific binding, and left to rotate at 4^oC in 0.1% BSA for 1-2 hours.

For the IP, 10% total cell lysate was saved as an input sample, then 2 mg of cell lysate for each IP was made up to 300 μl with NET buffer. 4 μg primary antibody per mg cell lysate was added to each IP reaction (antibodies described in Table 2.1) and samples rotated for 1 hour at 4^oC.

2.5.2.3 IP recovery

Blocked beads were washed a further three times in 0.5% BSA, once in 1 ml NET buffer and finally resuspended in 30 μl NET buffer per IP. 30 μl bead suspension was

added to each IP and returned to the rotator at 4^oC for 1 hour.

Beads were washed three times in PBS and switched to a new 1.5 ml microcentrifuge tube on the final wash to prevent co-elution with the plastic.

2.5.2.4 SDS-PAGE and Western Blot

Beads were resuspended in NuPAGE LDS Sample Buffer (4X, ThermoFisher) and NuPAGE Sample Reducing Agent (10X, ThermoFisher) and made up to 30 μl with PBS. Samples were heated to 70^oC for 10 minutes and the supernatant removed from the beads. Samples were then run on a 4-12% Bis-Tris gel with MOPS running buffer (as described in section 2.5.1.2).

Once run, the gel was transferred (as described in section 2.5.1.3) and the membrane blocked and incubated with a primary antibody. The following day the membrane was incubated with a secondary antibody (as described in section 2.5.1.4). The membrane was then visualised to identify if any proteins had co-immunoprecipitated with CBP (as described in section 2.5.1.5).

2.5.3 Developments to the Co-IP protocol

2.5.3.1 Flag Co-IP using anti-Flag magnetic beads

In section 4.5.3.1, Co-IP for the Flag tagged CBP was carried out. All steps for the Co-IP were carried out as described in section 2.5.2 until the immunoprecipitation. 50 μl resuspended magnetic agarose conjugated Flag beads (Pierce) were used per IP and washed three times in NET buffer (as described in section 2.5.2). 500 μg lysate was added to each IP reaction and samples were incubated at room temperature with rotation for 20 minutes. Beads were then washed with PBS three times and resuspended in 500 μl H_2O.

Flag tagged protein was eluted from the beads using a competitive elution with 3x DYKDDDK peptide (Pierce). 100 μl of 3x DYKDDDK peptide (1.5 mg/ml) was added to each IP sample and incubated at room temperature for 10 minutes in a thermomixer at 1400 rpm. Supernatant was removed by placing samples on a magnet stand and the samples were then prepared for SDS-PAGE (as described in section 2.5.2.4).

2.5.3.2   Strep tag Co-IP using conjugated Strep magnetic beads

In section 5.6.1, Co-IP for the Strep tag on MYB was carried out. All steps for the Co-IP were carried out as described in section 2.5.2 until the immunoprecipitation. 20 μl resuspended MagStrep XT beads (Stratech) were used per immunoprecipitation and washed three times in MagStrep wash buffer (1 M Tris-HCl, 1.5 M NaCL, 10 mM EDTA pH 8.0). 1 mg lysate was added to each IP reaction and samples were incubated at $4^{o}C$ with rotation. Samples were washed with 100 μl MagStrep wash buffer for a total of three washes and then beads were prepared for SDS PAGE (as described in section 2.5.2.4).

2.6   RNA analysis techniques

2.6.1   Real Time quantitative Polymerase Chain Reaction (RT-qPCR)

2.6.1.1   RNA extraction

A minimum of 2 x 10^5 Jurkat cells per sample were harvested and centrifuged at 400 rcf for 5 minutes to pellet cells. Cell pellet was resuspended in 500 μl PBS in 1.5 ml microcentrifuge tubes with well-fitting lids and centrifuged again at 400 rcf for 5 minutes. After removal of the supernatant, 1 ml TRIzol (ThermoFisher) was added to each sample and resuspended by gently pipetting up and down. Care was taken at each step to not be too vigorous, to prevent shearing of nascent RNAs. 200 μl Chloroform (ThermoFisher) was added and samples were shaken for 15 seconds, rested for 2 minutes and shaken for a further 15 seconds.

Subsequent centrifugation at $4^{o}C$ for 15 minutes at 13,000 rcf allowed separation of the sample into three distinct phases. The upper aqueous phase was transferred to a new 1.5 ml microcentrifuge tube, with the addition of 1 μl Glycoblue (ThermoFisher). 500 μl of Isopropanol (Sigma) was added to each sample, inverted to mix and samples were then left to precipitate at room temperature for 10 minutes. Samples were centrifuged at 20,000 rcf for 15 minutes at $4^{o}C$ and all isopropanol was pipetted away from the pellet. The pellet was washed in 1 ml 70% ethanol (Sigma), centrifuged at 20,000 rcf for 2 minutes at $4^{o}C$ and ethanol pipetted away from the pellet. The pellet

was then further washed in 1 ml 80% ethanol, centrifuged at 20,000 rcf for 2 minutes at 4^oC and ethanol pipetted away from the pellet. A subsequent centrifugation step at 20,000 rcf for 2 minutes at 4^oC allowed pipetting away of any remaining ethanol. After brief air drying, the pellet was resuspended in 16 μl 1X TBE (0.13 M Tris (pH 7.6), 45 mM boric acid (Sigma), 2.5 mM EDTA) .

The extracted RNA was then DNase treated to remove any remaining DNA. 10X Turbo DNase buffer and 2 μl Turbo DNase was added to each sample and incubated at 37^oC for 30 minutes.

A subsequent Phenol Chloroform extraction was required to purify away the DNase, to prevent any degradation of cDNA after reverse transcription. Samples were brought up to 200 μl with H_2O, and 200 μl acidic Phenol Chloroform was added (pH 4.5, Ambion). Samples were shaken, as opposed to vortexed, to prevent shearing of more unstable RNA species. Samples were centrifuged for 15 minutes at 12,000 rcf at 4^oC, and the upper aqueous phase was taken to a new tube. 200 μl chloroform was added and samples were shaken again and centrifuged at 12,000 rcf for 10 minutes at 4^oC. The aqueous phase was once again taken to a new 1.5 ml microcentrifuge tube and 0.1 Volumes of 3M NaAc (Sigma) and 1 μl GlycoBlue were added. 2.5 Volumes of cold 100% ethanol (Sigma) were added and samples were inverted to combine. RNA was then left to precipitate overnight at -80^oC. After precipitation samples were centrifuged for 30 minutes at 20,000 rcf at 4^oC and the ethanol was pipetted away from the pellet. The pellet was washed in 1 ml 70% ethanol, centrifuged at 20,000 rcf for 2 minutes at 4^oC and ethanol pipetted away from the pellet. The pellet was then further washed in 1 ml 80% ethanol, centrifuged at 20,000 rcf for 2 minutes at 4^oC and ethanol pipetted away from the pellet. A further centrifugation step at 20,000 rcf for 2 minutes at 4^oC allowed pipetting away of any remaining ethanol. After brief air drying, the pellet was resuspended in 12 μl 1X TBE.

2.6.1.2   Reverse transcription

1 μl RNA per sample was taken for quantification using the Qubit RNA BR kit according to the manufacturer's instructions (ThermoFisher).

Reverse transcription was then carried out with between 0.5 μg and 2 μg RNA per sample, with samples for the same experiments made up to the same concen-

tration and samples made up to 10 μl with nuclease-free H_2O. Reverse transcription was carried out using the High Capacity cDNA kit (ThermoFisher) which contains random primers, meaning all RNA species can be reverse transcribed, according to the manufacturer's instructions.

After reverse transcription, samples were diluted to a final volume of 200 μl with Nuclease Free H_2O and stored at -20^oC until preparation of the RT-qPCR plate.

2.6.1.3 Preparation of RT-qPCR plate

Samples were prepared and ran on either 96 well and 384 well plates and Power SYBR reaction mastermix (ThermoFisher) was used.

Table 2.12: Primer list for RT-qPCR

Name of primer	Target	Strand	Usage	Sequence
18s_rRNA_fwd	18S ribosomal RNA	Forward	Gene expression housekeeping gene	GTAACCCGTTGAACC CCATT
18s_rRNA_rvs	18S ribosomal RNA	Reverse	Gene expression housekeeping gene	CCATCCAATCGGTAGT AGCG
CBP_ex30_fwd	CBP	Forward	Gene expression	AGGACCATGTGGCTAA GTGC
CBP_ex30_rev	CBP	Reverse	Gene expression	CAGTGAGAAAGGTCCC CCAC

Table 2.13: Primer list for RT-qPCR 2

Name of primer	Target	Strand	Usage	Sequence
MYB_ex16_ fwd	MYB	Forward	Gene expression	TGTTGCATGGATCCTG TGTT
MYB_ex16_ rev	MYB	Reverse	Gene expression	AGTTCAGTGCTGGCCA TCTT
TAL1_ex4_ fwd	TAL1	Forward	Gene expression	AGGGCCTGGTTGAAGA AGAT
TAL1_ex4_ rev	TAL1	Reverse	Gene expression	AAGTAAGGGCGACTGG GTTT
TAL1_eRNA _S_fwd	TAL1 eRNA plus strand	Forward	Gene expression	TATGGGACTGGGGAGA AGGG
TAL1_eRNA _S_rev	TAL1 eRNA plus strand	Reverse	Gene expression	TGTGTGTCTCCTGAAC GGTG
TAL1_eRNA _AS_fwd	TAL1 eRNA minus strand	Forward	Gene expression	TGCAGCTGAGGACAGT ATTGAT
TAL1_eRNA _AS-rev	TAL1 eRNA minus strand	Reverse	Gene expression	CCACATCAATCTTATG TTGTCCAGG
MYB_6His_ fwd	MYB C-terminus and 6xHis tag	Forward	Gene expression	CGGACGCTAGTCATGA TGGG
HA/Strep_rev	HA/Strep tag	Reverse	Gene expression	TTTTTCGAACTGCGGG TGGC
Flag_fwd	Flag tag	Forward	Gene expression	AGGCGGCAGCGACTAT AAGG

For a 96 well format, 6 μl cDNA sample or H_2O was loaded in triplicate into each well of the 96 well plate (ThermoFisher). Power SYBR mastermixes were set up for each primer set (Table 2.12 for primer list, Table 2.14 for 96 well format) and 14 μl primer mastermix was added to each well. Plates were then centrifuged to spin all reagents to the bottom of each well, and plates were sealed with adhesive film (ThermoFisher).

Table 2.14: Preparation of Power SYBR mastermix for 96 well plate RT-qPCR

	Volume for 1 reaction
2x Power SYBR	10 μl
Forward primer (10 μM)	0.2 μl
Reverse primer (10 μM)	0.2 μl
H$_2$O	3.6 μl
Total mastermix per well	14 μl

For a 384 well format, 3 μl cDNA sample or H$_2$O was loaded in triplicate into each well of the 384 well plate (ThermoFisher). Power SYBR mastermixes were set up for each primer set (Table 2.12 for primer list, Table 2.15 for 384 well format) and 7 μl primer mastermix was added to each well. Plates were then centrifuged to spin all reagents to the bottom of each well, and plates were sealed with adhesive film (ThermoFisher).

Table 2.15: Preparation of Power SYBR mastermix for 384 well plate RT-qPCR

	Volume for 1 reaction
2x Power SYBR	5 μl
Forward primer (10 μM)	0.1 μl
Reverse primer (10 μM)	0.1 μl
H$_2$O	0.8 μl
Total mastermix per well	7 μl

2.6.1.4 Running the RT-qPCR plate

RT-qPCRs were ran on a QuantStudio 12K Flex Real-Time PCR System (ThermoFisher).

qPCR plates were run selecting the "SYBR Green Reagents" and "Standard" options.

For new primers, reactions were set up using 5 standards of combined cDNA and diluted 1:10 in series. Cycle threshold (Ct) values were used to generate standard curves across the serial dilutions and primer pairs that resulted in an R^2 value of over 0.990 were used going forward.

2.6.1.5 RT-qPCR data analysis

Ct scores were exported from the qPCR machine and a mean Ct was calculated using each triplicate reaction.

The change in Ct or delta Ct (ΔCt) was calculated by normalising to the required housekeeping gene, for example *18S* for gene expression data. The relative fold change in gene expression was then calculated as $2^{-\Delta Ct}$. This made any negative fold change values positive, and upon plotting allowed for clearer comparisons of numerical values.

If a further normalisation step was required, for example to compare the change in gene expression between treated and untreated cells, the change in ΔCt was required. The delta delta Ct ($\Delta\Delta Ct$) was calculated by the difference between the ΔCt in the untreated and treated sample. $2^{-\Delta\Delta Ct}$ generated the relative change in gene expression between two samples.

After obtaining the changes in gene expression, data was input into Graphpad - Prism (`https://www.graphpad.com/features`). Where possible, biological replicates were input in triplicate and graphs were generated with the "Summary data" selection, with "Interleaved bars". Error bars were shown as calculated as the Standard Error of the Mean (SEM) and either a linear or a Log_{10} scale for the Y axis was chosen, dependent upon which allowed for the clearest visualisation of the data.

For statistical analysis, this was also carried out in Graphpad - Prism. A paired t-test was carried out between paired datasets that were represented as fold-changes (for example untreated with DMSO, compared to treated with A-485 (section 3.3)), assuming a Gaussian distribution, with a threshold P value set to 0.05. P values were then calculated, with $p > 0.05$ showing no significance, and $p < 0.05$ showing significant differences between the data. To indicate significance on the graphs, ns = not significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$.

2.6.2 Native RNA-Immunoprecipitation (RIP) followed by gel visualisation

In order to prevent protein degradation, all steps were done in succession until RNA isolation.

2.6.2.1 Nuclear extraction and lysis

200×10^6 cells were harvested per RNA-immunoprecipitation (RIP) by centrifugation at 400 rcf for 5 minutes at 4^oC and the pellet was resuspended in 10 ml PBS. Cells were further centrifuged at 400 rcf for 5 minutes at 4^oC and the pellet was resuspended in 5 cell pellet volumes of RIP Buffer A (10 mM Tris pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl (Sigma), 100X Halt Protease and Phosphatase Inhibitor cocktail (ThermoFisher)). Samples were incubated for 10 minutes on ice followed by centrifugation at 400 rcf at 4^oC for 5 minutes.

The cell pellet was then gently resuspended in 2 cell pellet volumes of RIP Lysis Buffer (RIP Buffer A + 0.2% IGEPAL CA-630 (Sigma)) and incubated on ice for 5 minutes to lyse the cells. To pellet the nuclei, samples were centrifuged at 2,500 rcf and at 4^oC for 5 minutes.

To lyse the nuclei the pellet was resuspended in 2 cell pellet volumes of RIP Buffer C (20 mM Tris pH 7.9 at 4^oC, 25% glycerol (Sigma), 400 mM NaCl, 1.5 mM $MgCl_2$, 10 mM EDTA, 0.4 U/μl murine RNAse inhibitor (NEB), 100X Halt Protease and Phosphatase Inhibitor cocktail) and incubated at room temperature for 30 minutes with rotation.

To clear the nuclear lysate, samples were centrifuged at 20,000 rcf for 30 minutes at 4^oC. Nuclear lysate concentration was calculated by Pierce BCA assay (ThermoFisher) according to the manufacturer's instructions.

2.6.2.2 Immunoprecipitation

For immunoprecipitation lysates were diluted to 1 mg/ml in 300 μl RIP IP Buffer (20 mM Tris pH 7.9 at 4^oC, 200 mM KCl, 10 mM EDTA, 0.05% IGEPAL CA-630), with 5% of total lysate being retained for an input sample. 4 μg primary antibody per mg cell lysate was added to each IP reaction (antibodies described in Table 2.1) and samples were rotated for 2-3 hours at 4^oC. Simultaneously, 7 μl Protein G dynabeads (ThermoFisher) per μg antibody used per IP were blocked with three washes of 0.5% BSA (Sigma) in PBS and left to rotate at 4^oC in 0.5% BSA.

2.6.2.3 Recovery of immunoprecipitation

To recover the immunocomplexes, 7 μl of blocked protein G dynabeads per μg antibody used were added to each sample and incubated with rotation at 4^oC for 1 hour. After this incubation, beads were washed for a total of three washes in RIP Wash Buffer (RIP IP Buffer, 10 mM EDTA, 1 mM $MgCl_2$, 0.05% IGEPAL CA-630) and then beads were resuspended in a final volume of 30 μl RIP Wash Buffer.

To eliminate any potential bridging effects of protein-DNA and RNA-DNA interactions, DNase treatment was done on bead. 3 μl 10X Turbo DNase buffer and 3 μl Turbo DNase (Ambion) was added to each sample and incubated at RT for 20 minutes.

Beads were then washed twice in 500 μl RIP Wash buffer. Upon the resuspension in 500 µl, beads were split to take 100 μl for Western Blot preparation and 400 μl was taken for RNA extraction.

2.6.2.4 RNA extraction

All buffer was removed from the beads by placing the samples on a magnet stand. Beads were then resuspended in 1 ml TRIzol (ThermoFisher) and 200 μl Chloroform was added. Samples were shaken vigorously for 15 seconds, rested for 2 minutes and further shaken for 15 seconds. Samples were then centrifuged at 12,000 rcf for 15 minutes at 4^oC. This caused the samples to separate out into an organic phase, an interphase and an aqueous phase.

The upper aqueous phase was removed into a new 1.5 ml microcentrifuge tube and the Monarch RNA Cleanup kit (NEB) was used to clean up the RNA according to the manufacturer's instructions.

The extracted RNA was then DNase treated to remove any remaining DNA. 10X Turbo DNase buffer and 1.1 μl Turbo DNase was added to each sample and incubated at 37^oC for 30 minutes.

2.6.2.5 Denaturing gel electrophoresis

The extracted RNA was combined with 2X Formamide Loading Buffer (1x TBE, 88% Formamide (Sigma), 10 mM EDTA, 5% Glycerol (with blue 0.05% Bromophenol Blue

(Fisher), 0.05% Xylene Cyanol FF (Sigma)) and boiled at 95^oC for 5 minutes, followed immediately by placing on ice for 5 minutes to allow correct folding of the RNA.

The XCell SureLock Mini-Cell Electrophoresis System (ThermoFisher) tank was used. The gel tank was set up by placing and securing the 10% TBE Urea gel, removing the comb and filling with 1X TBE running buffer. Each individual well needed to be washed out immediately prior to loading of each individual sample. The gel was run at 150 V for 70 minutes.

After the gel had run, the gel was excised from the tank and stained in SYBR Gold (ThermoFisher, diluted 1:10,000 in TBE) for 10 minutes. The gel was imaged using the Typhoon FLA 7000 biomolecular imager (GE Healthcare), by excitation at 473 nm, and with a 520 nm filter.

2.6.2.6   Preparation of samples for Western Blotting

Samples taken for Western Blotting had all buffer removed by placing on a magnet stand. Then samples were prepared for Western Blotting as described in section 2.5.2.4.

2.6.3   Native RIP followed by RT-qPCR

In order to prevent protein degradation, all steps were done in succession until RNA isolation.

2.6.3.1   Cell lysis

200×10^6 cells were harvested per RNA immunoprecipitation by centrifugation at 400 rcf for 5 minutes and the pellet was resuspended in 10 ml PBS. Cells were further centrifuged at 2,500 rcf for 5 minutes and the pellet was resuspended in 300 μl RIP NET buffer (50 mM Tris-HCl pH 7.3, 150 mM NaCl, 0.1 mM EDTA, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 2 mM $ZnCl_2$, 0.1% NP-40, 10 mM Sodium Butyrate, 0.4 U/μl murine RNAse inhibitor (NEB), 100X Halt Protease and Phosphatase Inhibitor cocktail). To digest viscous chromatin 10X Turbo DNase buffer was added directly to the pellet as well as 3 μl Turbo DNase and incubated at 37^oC for 30 minutes. To clear the cell lysate away from chromatin and debris, lysate was centrifuged at 20,000 rcf for 20

minutes and protein concentration of cell lysate was calculated using the Pierce BCA assay (ThermoFisher) according to the manufacturer's instructions.

2.6.3.2 Immunoprecipitation

For immunoprecipitation lysates were diluted to 1 mg/ml in 300 μl RIP IP Buffer (20 mM Tris pH 7.9 at $4^{o}C$, 200 mM KCl, 10 mM EDTA, 0.05% IGEPAL CA-630), with 5% of total lysate being retained for an input sample. 4 μg primary antibody per mg cell lysate was added to each IP reaction (antibodies described in Table 2.1) and samples were rotated for 2-3 hours at $4^{o}C$. Simultaneously, 7 μl Protein G dynabeads (ThermoFisher) per μg antibody used per IP were blocked with three washes of 0.5% BSA (Sigma) in PBS and left to rotate at $4^{o}C$ in 0.5% BSA.

2.6.3.3 Recovery of immunoprecipitation

To recover the immunocomplexes, 7 μl of blocked protein G dynabeads per μg antibody used were added to each sample and incubated with rotation at $4^{o}C$ for 1 hour. After this incubation, beads were washed for a total of three washes in RIP Wash Buffer (RIP IP Buffer, 10 mM EDTA, 1 mM $MgCl_2$, 0.05% IGEPAL CA-630) and then beads were resuspended in a final volume of 30 μl RIP Wash Buffer.

To eliminate any potential bridging effects of protein-DNA and RNA-DNA interactions, DNase treatment was done on bead. 3 μl 10X Turbo DNase buffer and 3 μl Turbo DNase (Ambion) was added to each sample and incubated at RT for 20 minutes. Beads were then washed three times in 500 μl RIP Wash buffer.

2.6.3.4 Purification of RNA

Total sample was taken for RNA purification. All buffer was removed from the beads by placing the samples on a magnet stand. Beads were then resuspended in 1 ml TRIzol. RNA extraction then proceeded as described for RNA extraction for RT-qPCR as described in section 2.6.1.

2.7 Protein knockdown and knockout techniques

2.7.1 Use of CBP specific inhibitors

2.7.1.1 A-485

The small molecule A-485 (Stratech) was used as an inhibitor for CBP as it inhibited catalytic activity in this protein.

4 x 10^5 Jurkat cells were plated per well of a 6 well plate in 2 ml complete RPMI. A-485 was added to a final concentration of 5 μM. Untreated cells had corresponding volumes of DMSO added to act as a control.

After 48 hours cells were harvested for RNA extraction and followed by RT-qPCR (as described in section 2.6.1).

2.7.1.2 C646

An additional small molecule C646 (SelleckChem) was used as a CBP inhibitor, as this molecule inhibited histone acetyltransferase activity.

4 x 10^5 Jurkat cells were plated per well of a 6 well plate in 2 ml complete RPMI. C646 was added to a final concentration of 25 μM. Untreated cells had corresponding volumes of DMSO added to act as a control.

After 18 hours cells were harvested for RNA extraction and followed by RT-qPCR (as described in section 2.6.1).

2.7.2 CRISPR/Cas9 mediated knockouts

2.7.2.1 gRNA design

CRISPR/Cas9 was also utilised for knockout experiments, such as the knockout of MYB in Jurkat cells.

Three gRNAs were designed to target early exonic regions of the MYB gene, which would result in knockout of the protein. gRNAs were designed as described previously (in section 2.4.1.1) and are listed in Table 2.5.

A scrambled gRNA with no homology to the Jurkat genome was designed and used as a control.

2.7.2.2 Transfection

The gRNAs were ordered as short oligonucleotides and cloned into the pX458 vector (as described in section 2.2.6.3).

pX458 plasmids containing each gRNA were then transfected into 2 x 10^5 Jurkat cells using the 10 μl Neon electroporation kit (ThermoFisher, as decribed in section 2.3.5.1).

Transfected cells were left to incubate at 37^oC for 96 hours to allow sufficient time for the CRISPR/Cas9 cleavage to have an observable change.

2.7.2.3 RT-qPCR

After incubation for 96 hours, all cells were harvested for RNA extraction and were followed by RT-qPCR (as described in section 2.6.1).

2.7.3 HaloProtacs

2.7.3.1 Generating a MYB-Halotag Jurkat clonal cell line

Within WT Jurkat cells, MYB was tagged with a Halotag using methods as described previously in section 2.4.

2.7.3.2 Addition of HaloProtacs

HaloProtacs (Promega) are a Halotag ligand that induce degradation by targeting the Halotagged protein for degradation by E3 ubiquitin ligase. This was used to induce MYB-Halotag degradation in the Jurkat cells.

3 x 10^6 MYB-Halotag cells were counted and plated into 2.5 ml complete media in a well of a 6 well plate. HaloProtacs were added to each well to a final concentration of 1 μM. For untreated wells, equivalent volumes of DMSO were added as a control.

After 18, 24 and 48 hours, samples were taken for RNA extraction and Western blotting. Western blotting (as described in section 2.5.1) was carried out to examine protein levels within the cell population. RNA extraction was followed by RT-qPCR (as described in section 2.6.1) to examine downstream effects on the cell from MYB degradation.

Chapter 3

CBP forms a regulatory protein complex in Jurkat cells

3.1 Introduction

3.1.1 Action of the TAL1 complex in Jurkat cells

It was previously shown that within the Jurkat cell line, expression of the transcription factors c-MYB, TAL1, GATA3 and RUNX1 was interconnected, as these proteins formed an autoregulatory loop (Sanda et al. 2012). Downregulation of one protein caused downstream downregulation of the other three transcription factors, as shown by knockdown using short-hairpin RNAs (shRNAs) followed by RT-qPCR (Sanda et al. 2012). The proteins found in this positive regulatory loop, alongside the binding partners of TAL1: E2A and HEB (Palomero et al. 2006), were collectively termed the TAL1 complex (Sanda et al. 2012; Mansour, Abraham, et al. 2014).

3.1.2 The TAL1 complex is found at the *TAL1* enhancer

One target of the TAL1 complex has since been identified as the *TAL1* super-enhancer, the key driver of leukemogenesis in Jurkat cells (Figure 3.1, Sanda et al. 2012; Mansour, Abraham, et al. 2014). This regulatory element was found 7.5 kb upstream of the *TAL1* transcriptional start site (TSS) and was responsible for activating high levels of oncogenic *TAL1* transcription (Mansour, Abraham, et al. 2014).

Using Chromatin Immunoprecipitation (ChIP) followed by high-throughput sequencing (ChIP-seq) to examine binding of the TAL1 factors genome wide, it has been shown that the proteins from the TAL1 complex co-localise at the *TAL1* enhancer site (Mansour, Abraham, et al. 2014). In addition, this region was shown to have high levels of the histone mark of acetylation of Lysine 27 on Histone 3 (H3K27ac) (Mansour, Abraham, et al. 2014). This histone mark is indicative of regions of high transcriptional activity and correlates with active enhancer regions (Creyghton et al. 2010). Other proteins that co-localised at this element included the histone acetyltransferase (HAT) and coactivator CREB binding protein (CBP) which was responsible for depositing these acetylation marks onto H3K27 (A J Bannister and Kouzarides 1996), as well as other transcriptional machinery characteristic of enhancers, such as RNA polymerase II and Mediator (Mansour, Abraham, et al. 2014) (Figure 3.1).
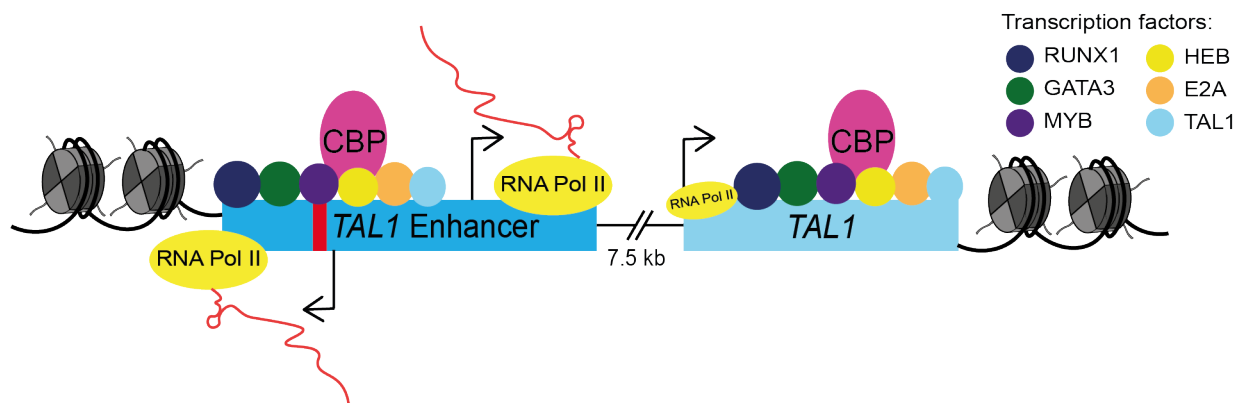
Figure 3.1: The *TAL1* enhancer upregulates oncogenic TAL1 expression in Jurkat cells
The *TAL1* enhancer element found 7.5 kb upstream of the *TAL1* TSS. Red stripe within the *TAL1* enhancer denotes a 12 bp insertion generating a binding site for the transcription factor MYB. The TAL1 complex consisting of MYB, TAL1, GATA3 and RUNX1 (as denoted in key) alongside HEB, E2A and CBP are bound to the *TAL1* enhancer and the *TAL1* TSS. RNA polymerase II (yellow) is also bound to the regulatory element, and transcribes bidirectionally producing sense and antisense eRNAs (red, directionality of transcription indicated by black arrow).

Due to the presence of this transcriptional machinery, enhancers are often transcribed bidirectionally into enhancer RNAs (eRNAs) (De Santa et al. 2010; T.-K. Kim et al. 2010). Both a sense and antisense eRNA have been shown to be transcribed

from the *TAL1* enhancer locus in Jurkat cells (Lidschreiber et al. 2021).

3.1.3 Role of CBP at enhancers
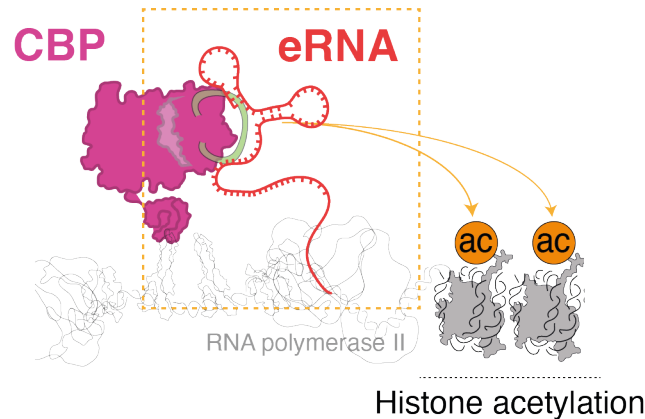
Figure 3.2: eRNA binding activates CBP HAT activity

eRNA (red) binding to the activation loop (green) of CBP (pink), causes conformational change in the activation loop of CBP. This stimulates HAT activity and therefore CBP can deposit acetylation modifications (yellow) onto Histone 3. Figure by Dan Bose.

As previously discussed, the HAT activity of CBP is important for its function, causing deposition of H3K27ac and other activating histone modifications (A J Bannister and Kouzarides 1996). The HAT activity of CBP is sensitive to the presence of local eRNA populations, eRNAs bind to an RNA binding site within the HAT domain of CBP causing a conformational change where the activation loop swings out, resulting in activated CBP which can then go on to deposit acetylation modifications (D. A. Bose et al. 2017)(Figure 3.2).

CBP can also increase gene transcription through many mechanisms further to activating the local chromatin environment (A J Bannister and Kouzarides 1996; Dai et al. 1996). One mode of action is to function as a scaffold to help facilitate protein complex formation (Holmqvist and Mannervik 2013). Such protein complexes can form at enhancer elements, causing an increase in enhancer activity. Interactions between the target promoter and the enhancer enable the delivery of such protein complexes and transcriptional machinery to the TSS of the target gene (Holmqvist

113

and Mannervik 2013). To enable this scaffolding behaviour, the CBP interactome encompasses over 400 proteins, including TAL1 and MYB (Bedford et al. 2010). The MYB-CBP interaction was first characterised almost 30 years ago where CBP was identifed as a coactivator of MYB (Dai et al. 1996).

3.1.4 Dissecting a model enhancer

As the *TAL1* enhancer is large and well characterised, and as enhancer activity can be measured in many ways, including eRNA transcription, as well as *TAL1* gene expression and cell viability, we planned to use this enhancer as a model system to dissect enhancer activity. Furthermore, the TAL1 complex thought to be acting at this enhancer element had also been shown to co-localise with CBP, however CBP had yet to be implicated with TAL1 complex behaviour (Sanda et al. 2012; Mansour, Abraham, et al. 2014). We also aimed to investigate the *TAL1* eRNAs and whether they were interacting with CBP and the TAL1 complex.

This chapter will outline the investigation into the role of CBP within the TAL1 regulatory protein complex and the role of RNA within Jurkat cells and at the *TAL1* enhancer.

3.2 CBP may be forming a complex in Jurkat cells

3.2.1 Investigating protein interactions within the TAL1 complex

The only protein-protein interactions that have been investigated within the TAL1 complex were that of MYB and TAL1 through co-immunoprecipitation (Co-IP) experiments (Mansour, Abraham, et al. 2014) and therefore interactions with other complex components remain uncharacterised. To explore whether there are further protein-protein interactions within the TAL1 complex, or whether these proteins are simply co-localising but not interacting with one another, we firstly sought to show that MYB does co-immunoprecipitate with TAL1 (Figure 3.3).

3.2.1.1   Co-immunoprecipitations

Co-IPs involve the addition of an antibody raised against a protein of interest within a whole cell lysate. The antibody will bind to this protein of interest and can then be extracted from the cell lysate, resulting in an immunoprecipitation. However if the protein of interest is found within any protein complexes or is interacting with any additional proteins either directly or indirectly, these additional proteins will co-immunoprecipitate with the protein of interest. Eluates can be then resolved using sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and subsequent western blotting can shed light as to whether combinations of proteins are co-immunoprecipitating.

3.2.1.2   MYB immunoprecipitates TAL1

MYB was immunoprecipitated from Jurkat cells as indicated by a triple banding pattern observed at approximately 80 kDa in both the input and the MYB IP lane (Figure 3.3). This banding pattern was distinctive in our Western blots for MYB and a likely explanation could be due to the multiple isoforms of MYB: the canonical MYB isoform is 80 kDa, alongside isoforms of 72 kDa, 75 kDa, two isoforms with a mass of 68 kDa and three further isoforms at approximately 85 kDa (Majello, Kenyon, and Dalla-Favera 1986; Westin, Gorse, and Clarke 1990). Therefore within Jurkat cells there could be three MYB isoforms present at sufficient levels to be observed by western blotting.

Blotting for TAL1 revealed a band at 42 kDa in the MYB IP lane, implying that when MYB was immunoprecipitated from Jurkat whole cell lysate, TAL1 was co-immunoprecipitated (Figure 3.3), recapitulating an interaction that had been shown previously (Mansour, Abraham, et al. 2014).

However, a limitation of Co-IP experiments is that it is not possible to infer how these proteins are interacting. For example, whether MYB and TAL1 were directly interacting or whether this protein-protein interaction observed was due to mediation by an additional cofactor such as another protein or RNA molecule.
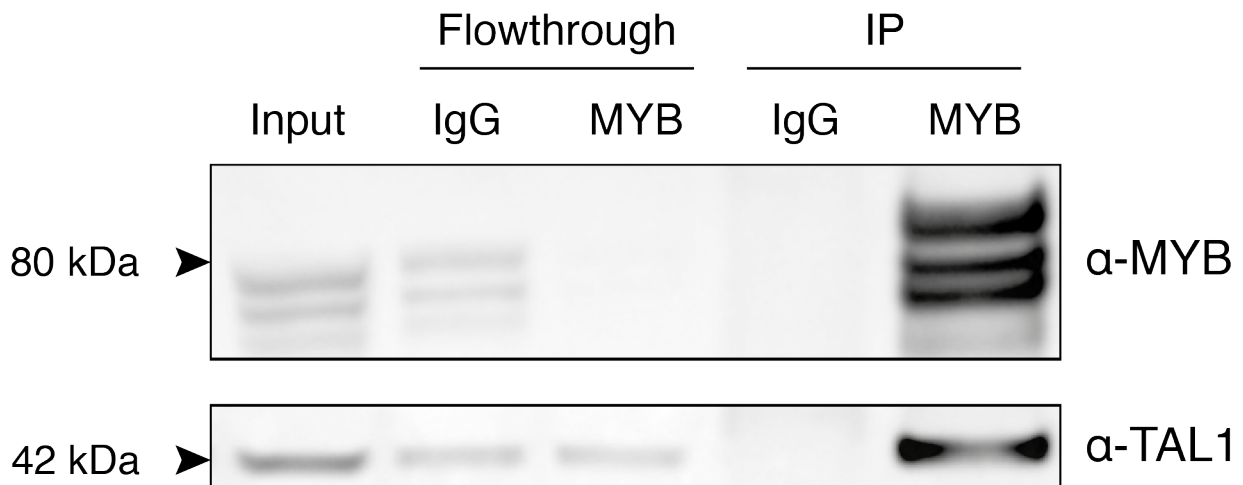
Figure 3.3: MYB co-immunoprecipitates TAL1

Co-immunoprecipitation of MYB from Jurkat cells indicated interaction with TAL1. Input sample contains 10% total lysate. IgG IP controlled for non-specific binding to the antibody. Flowthrough shows lysate unbound to the beads taken from both the IgG and MYB IP. IgG IP shows final elution of immunoprecipitation using the IgG antibody. MYB IP shows final elution of MYB IP, this shows TAL1 being co-immunoprecipitated by a band at 42 kDa. n=2.

3.2.2   CBP may be interacting with the TAL1 complex

After examining interactions between MYB and TAL1, we wanted to investigate the protein-protein interactions within the TAL1 complex further, to explore whether CBP was interacting with the TAL1 complex or co-localising. Consequently, we carried out a Co-IP for CBP in Jurkat cells to identify any protein-protein interactions between CBP and members of the TAL1 complex (Figure 3.4).

Co-immunoprecipitation of CBP appeared to pull down each of the proteins within the TAL1 complex: MYB, TAL1, GATA3, RUNX1, HEB and E2A (Figure 3.2.2). There was a band of expected full-length running height for each factor and there was little to no co-immunoprecipitation of these factors by IgG antibody, used as a control. Despite the band for TAL1 in the CBP IP lane being faint, it was still apparent and we could infer that TAL1, along with the remainder of the TAL1 complex, appeared to be co-immunoprecipitated with CBP.

116
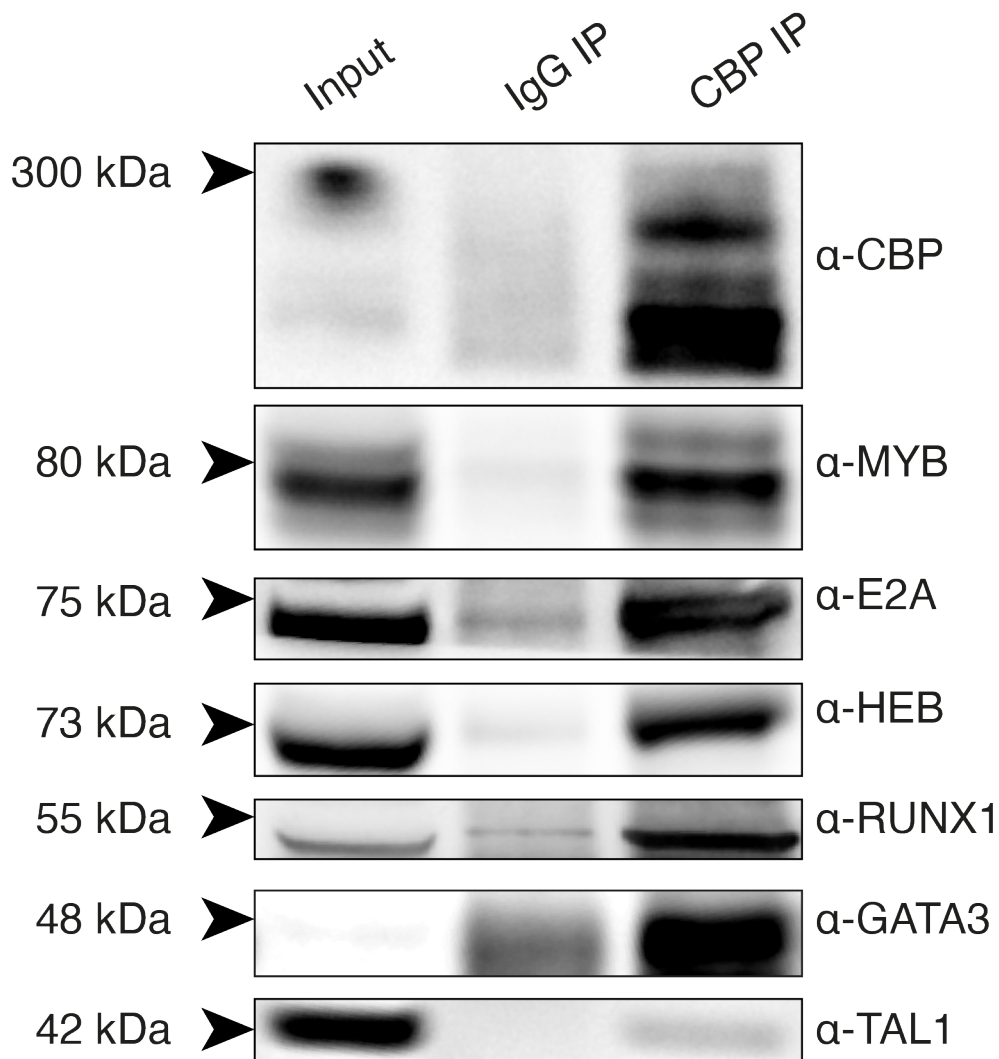
Figure 3.4: CBP Co-IP for the TAL1 complex

Co-Immunoprecipitation in Jurkat cells for CBP shows Co-IP for MYB, E2A, HEB, RUNX1, GATA3 and TAL1. Input shows 10% of input to each Co-IP and IgG Co-IP carried out to control for non-specific binding to the antibody. CBP Co-IP for MYB and TAL1 n=3. CBP Co-IP for GATA3, RUNX1, HEB and E2A n=1.

CBP could be acting as a scaffold within this regulatory protein complex and therefore showing interactions with each transcription factor. However, a limitation of Co-IP experiments is that it is not possible to determine whether these interactions are direct or indirect. Therefore we were not able to draw conclusions from this experiment as to the exact interactions within this protein complex, however we used

this Co-IP to suggest that CBP was interacting with each component in some way. We therefore inferred that as CBP appeared to be interacting with each transcription factor, CBP may also be part of this regulatory protein complex, and henceforth we referred to the complex as CBP-TAL1.

3.3   Examining the role of CBP in Jurkat cells

Our Co-IP experiments suggested that CBP may have been interacting within the CBP-TAL1 regulatory complex. Together with CBP binding being a key feature of enhancer elements, we wanted to greater understand the role CBP HAT activity may play in our system, in particular at the *TAL1* enhancer.

Two different inhibitors of CBP: A-485, a HAT inhibitor (Lasko et al. 2017) and C646, a catalytic inhibitor (Bowers et al. 2010), were added to Jurkat cells to examine downstream effects on the cell. To investigate the role CBP HAT activity plays on expression of CBP-TAL1 complex components and *TAL1* enhancer eRNAs, we followed the addition of the inhibitors for CBP with RT-qPCR, using primers for *CBP*, *MYB*, *TAL1* and the two defined eRNAs transcribed from the *TAL1* super-enhancer.

Addition of CBP inhibitors seemed to alter gene expression of the CBP-TAL1 complex component proteins (Figure 3.5A). When treated with A-485, *MYB* was expressed at a significantly lower level than in the untreated cells. When treated with C646, expression of *MYB* may have been down-regulated, but not to a significant level. Inhibition of the HAT activity of CBP by A-485 and C646 did not result in any significant changes to the expression of *CBP*. *TAL1* expression may have exhibited some weak downregulation after both treatment of A-485 and C646, however not to a level of significance and this is therefore speculative.

Addition of A-485 to Jurkat cells appeared to affect activity of the *TAL1* enhancer (Figure 3.5B). Expression of both the *TAL1* sense and antisense eRNAs was reduced after A-485 treatment compared to control untreated cells, however after C646 treatment the changes in eRNA expression were not significantly different compared to untreated cells.

118
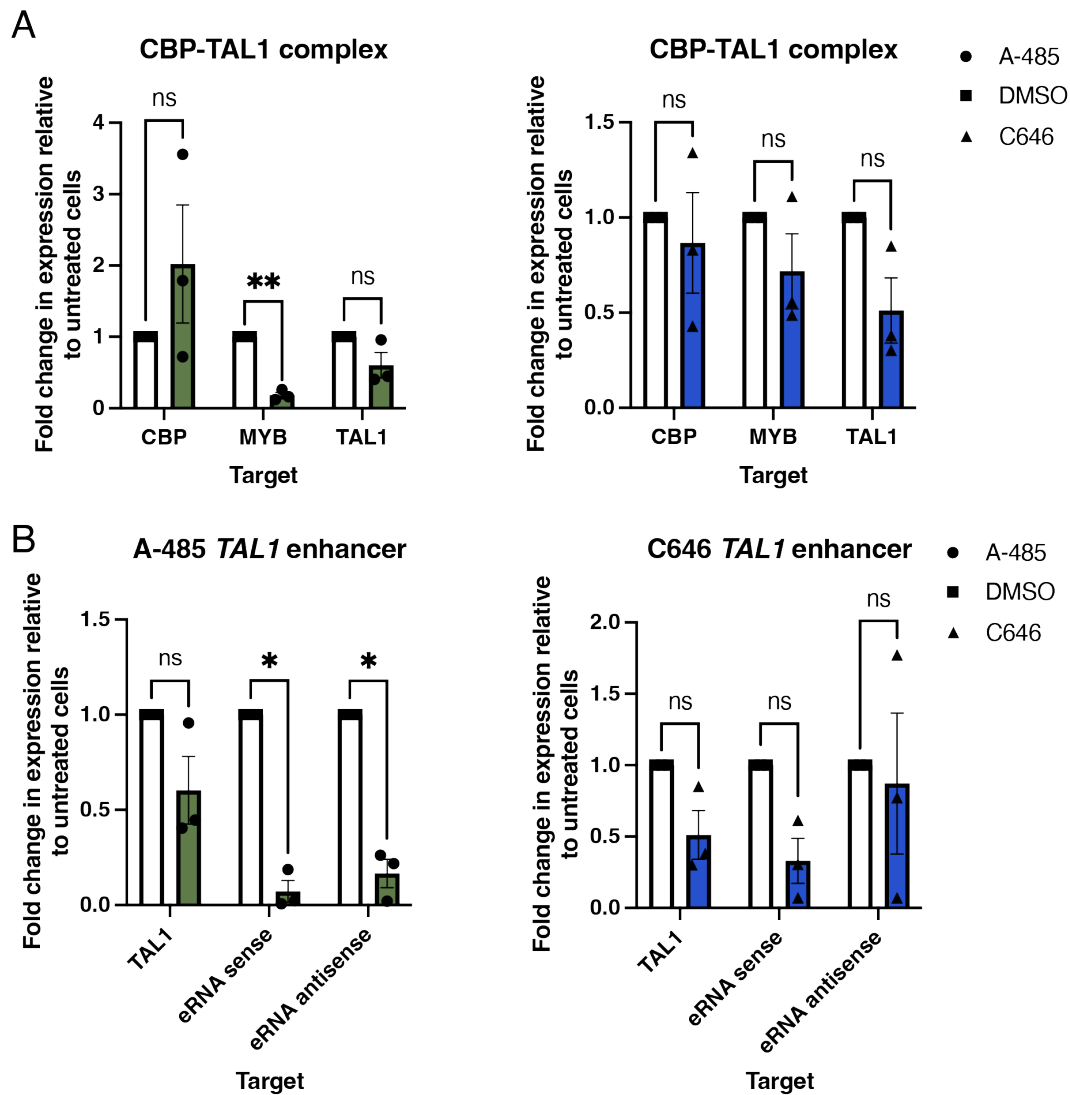
Figure 3.5: Inhibitors of CBP may affect TAL1 complex and *TAL1* eRNA expression RT-qPCR for *CBP*, *MYB* and *TAL1* gene expression as well as levels of the *TAL1* sense and antisense eRNAs after addition of CBP inhibitors. 5 μM A-485 was added and cells were harvested for RNA extraction after 48 hours. 25 μM C646 was added and cells were harvested for RNA extraction after 18 hours. Data was normalised to *18S* expression and expressed as a fold-change compared to DMSO untreated samples. (A) Change in gene expression of the CBP-TAL1 complex factors: *CBP*, *MYB* and *TAL1* after treatment with A-485 (green) and C646 (blue). (B) Change in expression of the *TAL1* enhancer: *TAL1* and the *TAL1* sense eRNA and antisense eRNA, upon treatment with A-485 and C646 (blue). Error bars denote SEM and n=3. Data points from cells treated with DMSO shown as squares, A-485 circles and C646 triangles. Paired t-test used: ns = not significant, * = p<0.05, ** = p<0.01.

The gene expression pattern after C646 and A-485 treatment was different, and this was likely due to the potencies of these CBP inhibitors. C646 was identified first, and despite results showing it was an effective inhibitor of CBP, the potency of C646 was thought possible to be improved upon (Bowers et al. 2010). Consequently, A-485 was developed, and this has since been shown to be a more effective inhibitor of CBP activity (Lasko et al. 2017). Therefore, the difference in gene expression patterns observed was likely due to the difference in effectiveness between the two inhibitors (Figure 3.5).

CBP HAT inhibition altered the gene expression environment within Jurkat cells. By inhibiting the mode of HAT action of CBP, *MYB* expression was reduced. MYB binding has been shown to drive the *TAL1* enhancer element, and the resulting high *TAL1* expression was the cause of leukemogenesis (Mansour, Abraham, et al. 2014). As CBP inhibition was shown to diminish *MYB* levels (Figure 3.5), it was possible that the formation of the CBP-TAL1 complex at the *TAL1* enhancer could have been reduced. This may have reduced *TAL1* enhancer activity and therefore could explain the decreased levels of *TAL1* eRNA expression. A way to test this would be to carry out ChIP–seq or ChIP followed by qPCR to examine chromatin localisation of the CBP-TAL1 complex at the *TAL1* enhancer after CBP HAT inhibition. To determine formation of the CBP-TAL1 complex, after addition of CBP inhibitors, Co-IPs for CBP and MYB could be repeated, which would allow visualisation of any disruption in protein-protein interactions.

Taken together, this suggested that CBP HAT inhibition impacted expression of the CBP-TAL1 complex and activity at the *TAL1* enhancer. This reinforces the role that CBP may be playing, both within the CBP-TAL1 complex and at the *TAL1* enhancer.

3.4   CBP-TAL1 complex and RNA

As CBP HAT inhibition resulted in downregulation of eRNA transcription at the *TAL1* enhancer (Figure 3.5B), coupled with the knowledge that HAT activity of CBP has been shown to be regulated by eRNA binding (D. A. Bose et al. 2017), we wanted to progress this to investigate whether the CBP-TAL1 protein complex as a whole interacted with RNA and if protein-RNA interactions were deduced, whether this

included the *TAL1* eRNAs.

3.4.1 The CBP-TAL1 transcription factors may have RNA binding activity

CBP has been well characterised as having RNA binding properties (D. A. Bose et al. 2017), but we sought to investigate the RNA binding capacity of the transcription factors within the CBP-TAL1 complex. This was done by RNA-immunoprecipitation (RIP) carried out under native conditions, followed by visualisation on a 10% TBE Urea gel.

A RIP involves immunoprecipitation of a protein of interest from cell or nuclear lysate, however after the IP step RNA extractions are carried out. If any RNA is interacting with the protein of interest this will be co-immunoprecipitated and after extraction any RNA can be visualised on a denaturing gel. Alternatively, RNA can be used for RT-qPCR or sent for RNA-sequencing (RNA-seq) to look for specific RNA species.

If the proteins within the CBP-TAL1 complex did have RNA binding properties, any immunoprecipitated RNA that had been extracted would be visualised as a smear or as bands on the denaturing gel.

Native RIP followed by visualisation on a denaturing gel indicated CBP-TAL1 complex proteins may have had RNA binding properties (Figure 3.6). Imaging of this gel revealed bands and smears of RNA on the TBE Urea gel in each IP lane for MYB, TAL1, GATA3, and RUNX1. This implied that when each protein was immunoprecipitated, they could have been binding to RNA under native conditions in the cell.

There were no visible bands for RNA bound to HEB or E2A, this could be due to these transcription factors not binding to any RNA, or it could be due to many other errors within the experiment, such as the antibody quality. Repeating this experiment would shed light as to whether HEB and E2A were able to bind to RNA, and also if we were to harvest part of the sample for western blotting, this could provide information as to whether the protein of interest was being correctly immunoprecipitated.
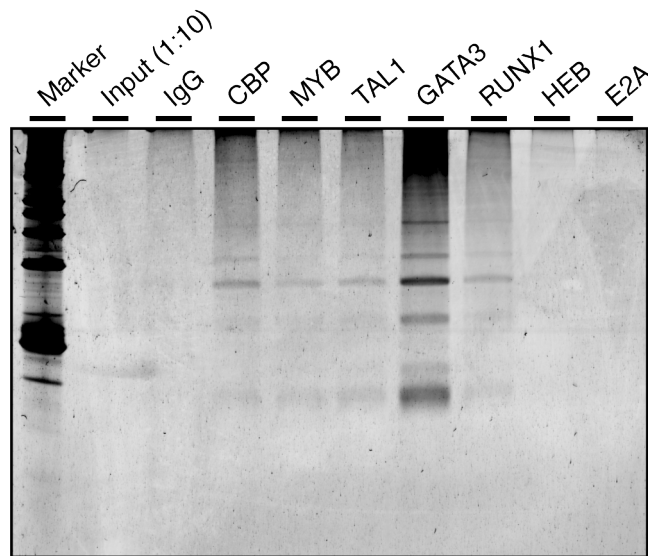
Figure 3.6: CBP-TAL1 transcription factors may have RNA binding properties RNA immunoprecipitation followed by electrophoresis of extracted RNA on a 10% TBE Urea gel. Visualised using a Typhoon FLA 7000 biomolecular imager (GE Healthcare), at 473 nm. CBP RIP, n=3. MYB RIP n=2. TAL1, GATA3, RUNX1, HEB, E2A and HEB RIP n=1.

We cannot make inferences about the strength in RNA binding from this gel, due to lack of loading control. If we were to repeat this, and upon RNA extraction harvest the organic phase after TRIzol addition, we could harvest the protein fraction immunoprecipitated. This would allow us to run an SDS-PAGE gel and stain with coomassie to visualise loading compared between the samples.

CBP, a known RNA binding protein, also appeared to have immunoprecipitated RNA, recapitulating data that had been previously shown in a different cell line (D. A. Bose et al. 2017). Therefore we could infer that CBP retained RNA binding properties in Jurkat cells.

A recent preprint examining the binding of transcription factors to RNA globally within K562 leukemia cells, suggested that TAL1, MYB, RUNX1 and HEB have RNA binding abilities (Oksuz et al. 2022). E2A and GATA3 were not identified as RNA binding proteins, yet other members of the GATA family were: GATA1 and GATA2 (Oksuz et al. 2022). This is the first and only additional evidence within the literature that the transcription factors within the CBP-TAL1 complex have the ability to bind

to RNA, however there were no further investigations as to the species of RNA these factors were binding to or in what contexts.

3.4.2 CBP-TAL1 complex could be binding to the *TAL1* eRNAs

As we had preliminary data indicating that many of the components within CBP-TAL1 had some RNA binding properties on a more global scale (Figure 3.6), we next investigated specific species of RNA. To do this we repeated our native RIP experiments, however after RNA extraction, these samples were reverse transcribed and used for RT-qPCR. We investigated whether the CBP-TAL1 complex was binding to the *TAL1* enhancer RNAs by using primers designed to amplify both the sense and antisense *TAL1* eRNAs (Figure 3.7).

Although due to time constraints, only one replicate was carried out and this would need to be repeated to draw full conclusions, our RIP-RT-qPCR data implied that the *TAL1* eRNAs may have been binding to the individual proteins within the CBP-TAL1 complex.
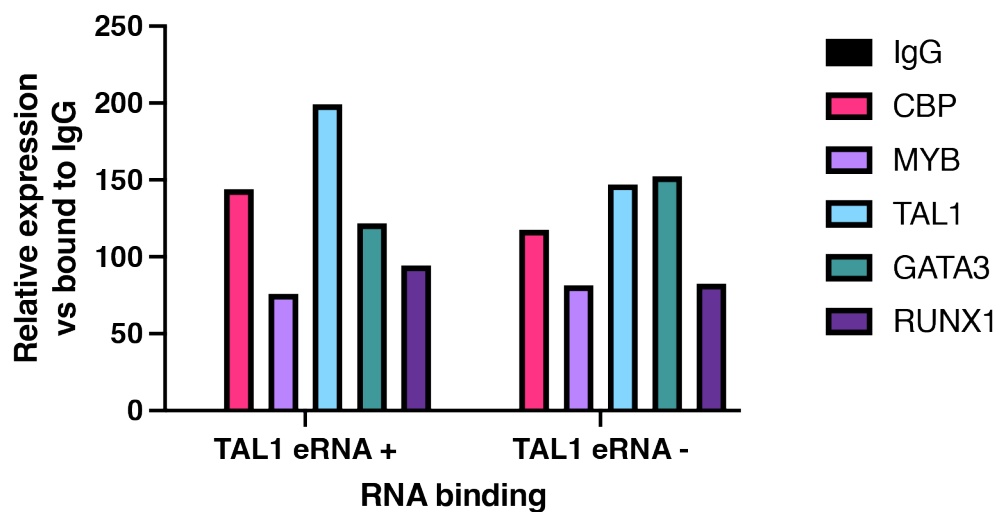
Figure 3.7: CBP-TAL1 proteins could bind to *TAL1* eRNAs
RNA immunoprecipitation followed by RT-qPCR implies binding to the eRNAs transcribed from the plus and minus strand of the *TAL1* enhancer. n=1.

Native RIP-RT-qPCR using primers designed to amplify the *TAL1* eRNAs tran-

scribed from the sense and antisense strand of the *TAL1* enhancer indicated that CBP, MYB, TAL1, GATA3 and RUNX1 could have been binding to both eRNAs (Figure 3.7). Interestingly, each of the proteins appeared to bind to each of the eRNAs at different levels. For example, TAL1 seemed to bind to the sense strand eRNA at much higher levels than the antisense strand, however for GATA3 the converse was true.

CBP has been shown to bind eRNAs and this eRNA-CBP interaction has been shown to regulate its HAT activity (D. A. Bose et al. 2017). Therefore, we anticipated being able to show that CBP did bind to both the sense and antisense *TAL1* eRNAs and in our one replicate, CBP did appear to bind to both of the eRNAs transcribed from our enhancer of interest (Figure 3.7).

3.5 Discussion

In this chapter we have explored the mode of action of the CBP-TAL1 regulatory protein complex which has been shown to form at the *TAL1* enhancer.

3.5.1 The chromatin environment of the *TAL1* enhancer

Within Jurkat cells, the *TAL1* enhancer is responsible for the aberrant upregulation of oncogenic TAL1 and causes leukemogenesis (Mansour, Abraham, et al. 2014). Due to the nature of this enhancer being an extremely large region, with H3K27ac levels stretching over 15 kb at this locus (Mansour, Abraham, et al. 2014), and there being many measurable outputs of this enhancer's activity, such as *TAL1* transcription, eRNA transcription, transcription factor and coactivator binding and cell viability, we have used this as a model enhancer to further explore the interactions between CBP and eRNAs.

We have used this enhancer to begin to dissect the role of regulatory protein complexes further. Firstly we have suggested that CBP be included within the regulatory protein complex that forms at this locus to upregulate transcriptional activity.

3.5.1.1 The TAL1 complex

The existence of the TAL1 complex was first suggested following the description of an autoregulatory loop between TAL1, MYB, GATA3 and RUNX1 in Jurkat cells

(Sanda et al. 2012). Due to the nature of TAL1 forming heterodimers with HEB and E2A to function as a transcription factor (Hsu, Cheng, et al. 1991; Hsu, Wadman, and Baer 1994; Palomero et al. 2006), HEB and E2A were also assumed to be part of this protein complex (Sanda et al. 2012).

Upon investigation of the *TAL1* enhancer, transcription factor binding motifs for each of these proteins were discovered close to the initiating MYB binding site, with subsequent ChIP-seq data confirming that this group of proteins were co-localising at this enhancer locus (Mansour, Abraham, et al. 2014). Reciprocal Co-IPs for MYB and TAL1 confirmed protein-protein interactions between these two proteins and it was therefore concluded that TAL1, MYB, GATA3, RUNX1, HEB and E2A were working together within a single protein complex (Sanda et al. 2012; Mansour, Abraham, et al. 2014). Our replicates of MYB Co-IP experiments corroborated these findings (Figure 3.3). However, no further protein-protein interactions within the TAL1 complex had as yet been investigated.

3.5.1.2   CBP is also a member of the TAL1 regulatory complex

When the ChIP-seq data from Jurkat cells was analysed on a more global level to investigate MYB binding sites genome wide, there was a high level of overlap between CBP and MYB binding sites (Mansour, Abraham, et al. 2014). It was also shown that TAL1, RUNX1 and GATA3 binding sites overlapped with these MYB binding sites (Mansour, Abraham, et al. 2014). Therefore, it was likely that CBP, MYB, TAL1, GATA3 and RUNX1 proteins were co-localising, not only at the *TAL1* enhancer site but throughout the genome.

CBP-MYB interactions have long been established and acetylation of MYB by CBP has been shown to aid in the binding of MYB to enhancer elements (Dai et al. 1996; Sano and Ishii 2001). CBP has also been shown to previously interact with TAL1, MYB, E2A and HEB (Bedford et al. 2010). CBP is known to function as a scaffold protein and has been shown to aid in the formation of regulatory protein complexes, helping to recruit different transcription factors and hold them at enhancer regions, enabling interactions with further transcriptional machinery such as RNA Pol II (Holmqvist and Mannervik 2013). Therefore, it would seem likely that CBP could be interacting with the TAL1 complex and acting as a scaffold to aid in protein

complex formation within Jurkat cells. Despite this, there is no evidence within the literature for CBP being part of the TAL1 protein complex. The co-localisation of MYB and CBP has been explained due to CBP's role as a coactivator of MYB and due to its role as a HAT and being a marker of active enhancer regions (A J Bannister and Kouzarides 1996; Creyghton et al. 2010; Mansour, Abraham, et al. 2014). However, it seemed to us possible that CBP was functioning as a scaffold within the TAL1 complex and was aiding in holding this group of proteins together at a single locus, helping to together activate transcription.

Therefore, to investigate our hypothesis of CBP behaviour with the TAL1 complex within Jurkat cells, we carried out a CBP Co-IP to identify which, if any, of the TAL1 complex proteins were interacting with CBP. Our data indicated that immuno-precipitation of CBP resulted in Co-IP of MYB, E2A, HEB, RUNX1, GATA3 and TAL1 (Figure 3.2.2). Due to CBP appearing to exhibit some interaction with each of the TAL1 complex proteins, coupled with the co-occupancy of CBP and the TAL1 complex throughout the Jurkat genome, this suggested that CBP could also be part of this regulatory protein complex. A drawback of Co-IP experiments is they are unable to distinguish between direct and indirect interactions, therefore it was not possible to directly conclude that CBP was scaffolding the remaining members of the protein complex directly from this CBP Co-IP experiment. Cross-linking mass spectrometry would enable any directly interacting peptides to be cross-linked together (O'Reilly and Rappsilber 2018). After degradation of non cross-linked peptides, these samples could be analysed by mass spectrometry and this could provide further information as to whether CBP is interacting directly or indirectly within the CBP-TAL1 complex (O'Reilly and Rappsilber 2018). Additionally, carrying out experiments *in vitro*, could provide further information as to whether CBP was interacting directly with each transcription factor. If CBP and each of the transcription factors was to be expressed and purified using affinity purification techniques, different combinations of these proteins could be mixed together. These could then be passed down a fractiona-tion column, and any proteins that elute within the same fractions could be inferred as directly interacting. Protein-protein crosslinking agents, such as glutaraldehdye, could also be utilised to crosslink any directly interacting peptides to improve fractionation (Migneault et al. 2004).

We proceeded with our investigations as one whole protein complex, referring to

126

it as the CBP-TAL1 complex to highlight the inclusion of CBP within this regulatory protein complex.

3.5.1.3 CBP-TAL1 complex at other loci in Jurkat cells

The CBP-TAL1 complex not only regulates the *TAL1* enhancer within Jurkat cells, but also at additional enhancers and promoters throughout the genome. As previously discussed, TAL1, GATA3, RUNX1 and CBP binding sites have been shown to overlap with MYB binding sites genome wide (Mansour, Abraham, et al. 2014). In addition, TAL1, GATA3, RUNX1 and CBP have been shown to bind at the same site within a canonical enhancer for each of *TAL1*, *GATA3* and *RUNX1* (Sanda et al. 2012). This provided further evidence that this regulatory protein complex was working at many loci throughout Jurkat cells to determine transcriptional behaviour.

Additionally, a knockdown screen was carried out within Jurkat cells to investigate downstream effects of the TAL1 complex (Leong et al. 2017). In this case, CBP was not included in their investigation. Knockdown of each transcription factor by shRNAs, coupled with ChIP-seq analysis, led to identification of AT-rich interactive domain (ARID) protein 5B (ARID5B) as a downstream target of the TAL1 complex (Leong et al. 2017). ARID5B has been shown to be important for the differentiation from progenitor cells to B lymphocytes (Whitson, T. Huang, and Itakura 1999) and also interacted with histone deacetylases (Joshi et al. 2013). Through ChIP-seq data, a super-enhancer was identified 135 kb upstream of the *ARID5B* TSS, and the CBP-TAL1 complex was shown to co-localise at this enhancer element (Leong et al. 2017). Similar to the investigations previously carried out into the *TAL1* enhancer (Sanda et al. 2012; Mansour, Abraham, et al. 2014), the TAL1 complex was shown to co-localise at this *ARID5B* enhancer, along with CBP. This provided more evidence that CBP was acting within the CBP-TAL1 regulatory protein complex within Jurkat cells and showed that this regulatory protein complex was important for controlling transcriptional networks throughout Jurkat cells. Determining whether CBP is necessary for CBP-TAL1 complex formation at enhancers genome wide would require depletion experiments coupled to ChIP-seq, or targeted ChIP-qPCR.

127

3.5.2 Inhibitors of CBP affect transcriptional activity of the *TAL1* enhancer

To further investigate the role of CBP within the CBP-TAL1 complex and within Jurkat cells, we treated wild type Jurkats with two different small molecule inhibitors of CBP HAT activity and followed this with RT-qPCR to examine changes to gene expression.

A-485 is a more potent inhibitor than C646 (Bowers et al. 2010; Lasko et al. 2017), so this was the most likely explanation for why we observed a larger effect on the CBP-TAL1 complex and activity at the *TAL1* enhancer upon A-485 treatment compared to C646 (Figure 3.5). It was reassuring to be able to observe that upon CBP inhibition, we measured decreases in eRNA transcription at the *TAL1* enhancer, implying that CBP HAT activity is required for transcriptional activation at this enhancer element.

3.5.2.1 Optimising HAT inhibitor experiments

Upon addition of A-485 to the Jurkat cells, a concentration of 5 μM was used, which despite being the recommended concentration from the manufacturer, is high. Therefore there is the possibility that the effects observed on transcriptional activity within the treated cells were due to the toxicity of the compound to the cell as a whole, rather than the inhibition of CBP. Carrying out a concentration curve, where increasing amounts of A-485 would be added to WT Jurkat cells, followed by an assay of HAT activity and RT-qPCR would allow determination of the optimal concentration of A-485 to inhibit HAT activity without being toxic to the cells. Therefore, we could be sure that changes to transcriptional activity observed were due to HAT inhibition.

3.5.2.2 HAT inhibitors affect both CBP and p300

The paralogues CBP and p300 are highly similar, and in most of our experiments we were able to distinguish between CBP and p300 with specific antibodies and RT-qPCR primers recognising differences within the N- and C-termini. However, the two proteins show 90% sequence conservation within the HAT domain (Henry, Kuo, and Andrews 2013). Therefore, addition of inhibitors to target the HAT domain of CBP will also inhibit p300 and therefore any affects to the cell after inhibition of CBP

HAT activity must also have the caveat that p300 HAT activity is also being affected (Lasko et al. 2017).

3.5.2.3   Follow up experiments after CBP inhibition

After showing that we were able to inhibit CBP to achieve a profound effect at our enhancer of interest, this allows scope for additional investigations. We could use inhibition of CBP to further dissect the action of the *TAL1* enhancer and also globally examine activity at enhancer elements along with the transcriptional environment across Jurkat cells. If any of these further experiments were to be carried out, this would be done using A-485 as we observed a more potent effect on CBP activity (Figure 3.5).

As A-485 is an inhibitor of CBP HAT activity (Lasko et al. 2017), not only would histone acetylation be reduced, but further acetylation substrates would exhibit reduced acetylation (A J Bannister and Kouzarides 1996). For example, in order for CBP to work as a coactivator of MYB, CBP is required to acetylate MYB (Dai et al. 1996; Fuglerud, Ledsaak, et al. 2018). Therefore, follow up experiments could determine how the lack of CBP acetyltransferase activity could affect complex formation and activity at the *TAL1* enhancer.

Carrying out transient transcriptome sequencing (TT-seq), when comparing cells treated and untreated with A-485, would allow us to capture the nascent RNAs transcribed across the Jurkat genome (Schwalb et al. 2016). This RNA population would therefore encompass eRNAs, as they are short-lived and very rapidly transcribed (T.-K. Kim et al. 2010). As we have observed that CBP inhibition caused a decrease in eRNA transcription at the *TAL1* enhancer (Figure 3.5), coupled with the knowledge that CBP is found at nearly all enhancer elements (T.-K. Kim et al. 2010), TT-seq in CBP inhibited cells would allow us to determine if this pattern of decreased eRNA transcription was representitive of every enhancer and whether this was indeed a global knockdown across Jurkat cells.

In addition, if we were to follow A-485 treatment with ChIP-seq or ChIP-qPCR for the CBP-TAL1 complex at the *TAL1* enhancer, we would be able to elucidate further the impact that CBP inhibition would have on this regulatory protein complex. As we hypothesise that at the *TAL1* enhancer, CBP is recruited by MYB and then acts as a

scaffold to aid in recruitment of the remainder of the complex, it would be interesting to observe CBP-TAL1 complex binding patterns after CBP inhibition. This could provide more detail about how CBP is functioning as a scaffold within this protein complex, and may shed light as to the protein-protein dynamics of this complex. Importantly, this would also allow us to test whether CBP activity is required for complex assembly genome wide.

To complement the investigations of how the CBP-TAL1 complex forms after CBP inhibition, Co-IPs could be repeated to examine if the protein complex is still able to form. Carrying this out on a larger scale by co-immunoprecipitating each protein within the CBP-TAL1 complex would allow us to determine the extent of an effect CBP has within the complex, as we could measure if any further protein binding was disrupted as a result of CBP inhibition.

3.5.3   The CBP-TAL1 complex and RNA

CBP is a known RNA binding protein, with binding of eRNAs to CBP resulting in conformational changes of the protein (He et al. 2016; D. A. Bose et al. 2017). Additionally, RNA species, particularly eRNAs, have been shown to play a role with the tethering of transcription factors to specific loci within the genome, particularly at enhancers (Sigova et al. 2015). We therefore wanted to explore whether there was a role for RNA in the CBP-TAL1 complex and if this did include eRNAs.

3.5.3.1   Transcription factors and RNA binding

The group of proteins known as transcription factors have the ability to bind to DNA, at specific binding sites for each factor (Vaquerizas et al. 2009). However, evidence is growing for many transcription factors to exhibit RNA binding behaviour also. Over 20 years ago, evidence for transcription factors exhibiting both DNA and RNA binding ability was suggested (Cassiday and Maher 2002). In a recent preprint, it was suggested that 41% of transcription factors within murine embryonic stem cells had RNA binding abilities (Oksuz et al. 2022), and as previously discussed eRNAs have been suggested to tether transcription factors to chromatin (Sigova et al. 2015).

Therefore, we investigated whether the transcription factors within the CBP-TAL1 complex had the ability to bind to RNA. We showed preliminary data that TAL1,

MYB, GATA3 and RUNX1 may show RNA binding abilities and could also bind to the eRNAs transcribed from the *TAL1* enhancer (Figures 3.6 and 3.7). This could therefore corroborate previous findings surrounding transcription-factor RNA and eRNA binding.

This also raises many questions as to whether each protein within the CBP-TAL1 complex is binding to the eRNAs individually, or whether there could be some level of cooperative binding across the complex. Therefore, decoding how the *TAL1* eRNAs are interacting within the CBP-TAL1 protein complex as a whole would shed light on CBP action within regulatory protein complexes, transcription factor-RNA and eRNA binding behaviour and a deeper insight into eRNA behaviour and function.

RNAs have been shown to promote the formation of phase separated condensates, and particularly at enhancers phase separation is increasingly being implicated as a model for assembly of transcriptional machinery (Banani et al. 2016; Hnisz, Shrinivas, et al. 2017). eRNAs have also been thought to play a scaffolding role within condensates at enhancers, helping to organise chromatin by tethering proteins to the enhancer, or by aiding in enhancer-promoter interactions (Sigova et al. 2015; Banani et al. 2016; Nair et al. 2019). Therefore there is large scope for speculation as to a potential mode of action of condensates forming at the *TAL1* enhancer, with the CBP-TAL1 complex and the *TAL1* eRNAs. It could be that the CBP-TAL1 complex is helping to form a condensate and the complex collectively binds to many copies of the eRNAs that are packed densely together in the separated phase. Or the eRNAs could be playing a tethering role, where they are helping to hold the complex at the enhancer locus, causing activating gene transcription. This could begin to be investigated by the targeted knockdown of the eRNAs, followed by ChIP-qPCR at the TAL1 locus, to examine if the CBP-TAL1 complex is able to bind at this enhancer without eRNAs.

3.5.4 The CBP-TAL1 complex is active at the *TAL1* enhancer and binds to the *TAL1* eRNAs

In this chapter we have begun to indicate that CBP may be interacting within the protein complex acting at the *TAL1* enhancer and therefore have suggested CBP could also be included within this protein complex. We hence referred to this complex as

131

CBP-TAL1 to highlight the inclusion of CBP. To investigate RNA interactions we have implied the factors within CBP-TAL1 may have some RNA binding properties and could bind to both the *TAL1* sense and antisense eRNA. Taken together, we suggest a development to the current model of the *TAL1* enhancer, that the CBP-TAL1 complex and the *TAL1* eRNAs may be together aiding in the formation and action of this protein complex to drive enhancer activity and downstream transcriptional activity of the *TAL1* target gene.

Chapter 4

Establishing a stable tagged CBP cell line for studying RNA binding

4.1 Introduction

In the previous chapter we investigated the CBP-TAL1 regulatory protein complex and implied that the transcriptional coactivator CBP and a series of transcription factors (TFs) comprising of c-MYB, TAL1, GATA3, RUNX1, HEB and E2A could interact to form a protein complex (Section 3.2.2). Further to this we indicated that the transcription factors found within CBP-TAL1 may have RNA binding capacity (Section 3.4).

CBP has previously been shown to bind to a range of RNA species, including enhancer RNAs (eRNAs) which are bidirectionally transcribed from active enhancers (De Santa et al. 2010; T.-K. Kim et al. 2010; D. A. Bose et al. 2017). Upon eRNA binding to a specific RNA binding region within the histone acetyltransferase (HAT) domain of CBP, the activation loop of the protein is released. This exposes the active site and enables HAT activity, resulting in high levels of H3K27ac being deposited at enhancer regions (D. A. Bose et al. 2017). The *TAL1* super-enhancer where the CBP-TAL1 complex is known to bind and highly upregulate target gene transcription is bidirectionally transcribed into both sense and antisense eRNAs (Danko et al. 2015; Lidschreiber et al. 2021). This raises a question of whether the *TAL1* super-enhancer eRNAs are binding to CBP.

4.1.1 Transcription factors bind to RNA

As we had previously indicated that all of the individual components within CBP-TAL1 may have had RNA binding properties (Section 3.4), we wanted to examine RNA activity within this protein complex as a whole. Data establishing transcription factor-RNA binding was first identified many years ago, with evidence still continuing to grow for transcription factors as a group to have RNA binding properties (Cassiday and Maher 2002). For example, the transcription factor SOX2 which is key for maintaining pluripotency in human embryonic stem cells has been shown to have direct RNA binding properties that are important for its function (Holmes et al. 2020). Additionally, the transcription factor Yin-Yang 1 (YY1) binds to different species of non-coding RNAs (ncRNAs), including eRNAs. Upon downregulation of these ncRNAs across the cell, YY1 binding is reduced across the genome. This implicates the ncRNA-YY1 interaction as a method of tethering transcription factors to regulatory elements and helping to upregulate transcriptional activation (Sigova et al. 2015).

4.1.1.1 Limitations of crosslinking and immunoprecipitation

Traditionally, Crosslinking and Immunoprecipitation (CLIP) experiments, and the many subsequent developments, have been used to examine protein-RNA interactions (Ule et al. 2005). For example Photoactivatable ribonucleoside-enhanced CLIP (PAR-CLIP) was used to examine CBP-RNA binding activity and this data was built upon to demonstrate the relationship of how CBP activity changes when in the presence of eRNAs (Hafner et al. 2010; D. A. Bose et al. 2017). CLIP experiments have been used to investigate global binding of a specific protein to RNA across the cell. In our system we wanted to look more specifically, to examine RNA binding when proteins are in complex together. As the transcription factors found alongside CBP within the CBP-TAL1 protein complex are key transcription factors that have multiple roles throughout the cell, it is likely that only a small proportion of our target transcription factors are incorporated into the CBP-TAL1 complex at any one time. Therefore, CLIP experiments that focus on single transcription factors will not differentiate RNA bound to that transcription factor independently or at other loci, from those bound in the context of the whole CBP-TAL1 complex.

4.1.2 Tandem-Affinity and Crosslinking Analysis of cDNAs

To investigate RNA binding within the CBP-TAL1 complex, we sought to understand whether changes in the protein composition of the complex resulted in changes within the bound RNA population. With a focus on CBP, we wanted to determine if changing the transcription factors bound to CBP altered the bound RNA population. As we had previously shown that the CBP-TAL1 complex had a role at enhancer regions, in particular at the *TAL1* enhancer, we planned to focus on the identification of changes to eRNAs, as this would give greater insight into eRNA activity. To characterise the RNA population binding to the transcription factors only when the transcription factors are found within the CBP-TAL1 regulatory protein complex specifically, we turned to Tandem Affinity Crosslinking and Analysis of cDNAs (TA-CRAC), a development of CRAC (Granneman, Kudla, et al. 2009; Thoms et al. 2015). TA-CRAC allowed investigation of the RNA species bound to transcription factors only when found within this specific regulatory protein complex.

4.1.2.1 Crosslinking and analysis of cDNAs

CRAC requires the protein of interest to have two endogenous tags. The two affinity tags separated by a Tobacco Etch Virus (TEV) protease cleavage site allow for sequential affinity purification steps. In our case CBP was to be endogenously tagged with a Flag-TEV-6xHis tag (Figure 4.1A). For CRAC, firstly RNA is crosslinked to protein by UV at 254 nm (Figure 4.1B). The Flag tag on CBP is then used to isolate tagged protein and bound RNAs under native conditions. Competitive elution using 3xFlag peptide will elute CBP from the beads and allow for a second purification step. The second affinity tag is a 6xHis tag, which is able to bind to nickel on a column or on beads. This interaction forms a strong covalent bond that does not require any protein folding (Hochuli et al. 1988). This means that the bound His-tagged protein of interest can undergo very strong denaturing washes in high salt conditions that disrupt all protein folding, but additionally will abrogate any protein-protein interactions. This results in only CBP remaining bound to the nickel beads alongside crosslinked RNA, eliminating any non-specific binding. Therefore, the RNA species that are subsequently extracted and sequenced are very specific in their protein binding activity and we are able to determine the binding of these species to CBP

(Granneman, Kudla, et al. 2009; Thoms et al. 2015).

4.1.2.2 TA-CRAC investigates RNA binding with two interacting proteins

The CRAC approach can be widened to Tandem-Affinity CRAC (TA-CRAC), which includes a second protein of interest with a different combination of affinity tags, that is known to be in complex with the first protein of interest (Figures 4.1A and B). A first affinity purification step isolates all tagged species, including those found within protein complexes. As the second protein of interest is known to be in complex with the first, there are some species bound to the first protein of interest. TEV cleavage elutes these proteins from the pulldown and subsequently releases a 6xHis tag found on the second protein. A following second affinity pulldown step can then occur under strong denaturing conditions, disassociating any interactions that are not between the 6xHis tagged protein bound to the beads or the cross-linked RNA. This ensures that only the protein of interest and the RNA species that are directly crosslinked will be pulled down. After crosslink reversal the bound RNAs can then be reverse transcribed to cDNA and sequenced (Granneman, Kudla, et al. 2009; Thoms et al. 2015).

4.1.2.3 Investigating MYB-CBP bound RNA

To investigate CBP-TAL1 as a protein complex, a second protein of interest to CBP is required: one of the six transcription factors found within CBP-TAL1. As we have previously indicated, MYB may have RNA binding properties and we have preliminary data demonstrating binding of MYB to the *TAL1* eRNAs (Section 3.4). As MYB is key for determining and driving generation of the *TAL1* enhancer, we sought to investigate the role played by MYB alongside CBP in determining RNA binding within this complex (Sanda et al. 2012; Mansour, Abraham, et al. 2014).

4.1.2.4 TA-CRAC

In order to use TA-CRAC to identify the RNA species that are bound only to MYB when it is found in protein complexes with CBP, we planned to further tag MYB with tandem affinity tags: 6xHis-TEV-HA-Strep (Figure 4.1A) within the tagged CBP cell line.

136
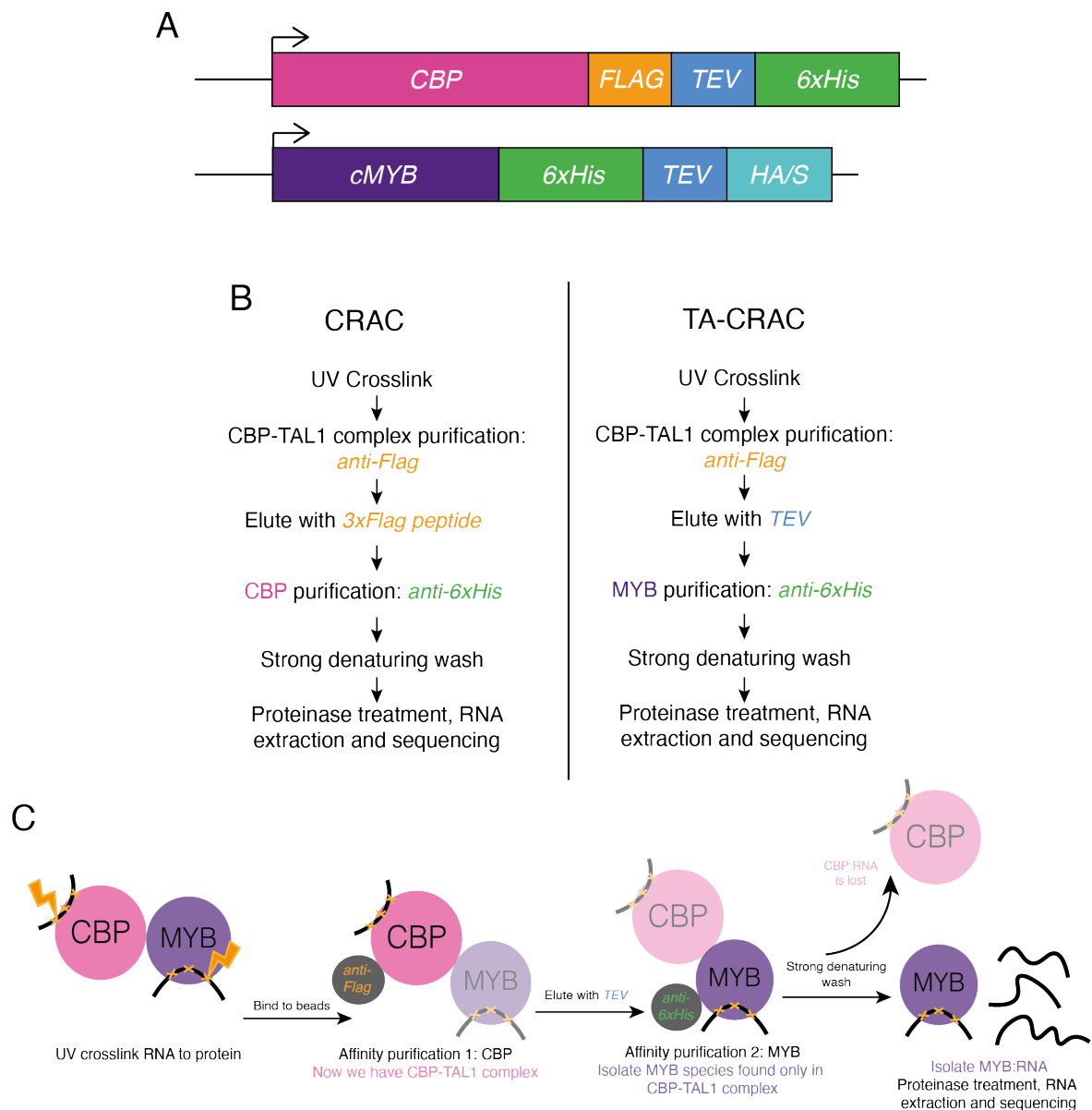
Figure 4.1: Tandem Affinity Crosslinking and Analysis of cDNAs to investigate the CBP-TAL1 RNA binding activity
(A) Tagging strategy. CBP will have a Flag-TEV-6xHis tag on the C-terminus and MYB will have a C-terminal 6xHis-TEV-HA/Strep tag (B) Strategy for CRAC and TA-CRAC. Colours match elements of the affinity tags to be inserted onto CBP and MYB in A. (C) CBP-MYB TA-CRAC.

By UV-crosslinking RNA to protein, we can isolate all CBP species by a first

affinity purification step under native conditions, but this will include the CBP-TAL1 complex isolated on the beads. By doing a second affinity purification step to target the different tags on MYB under denaturing conditions, we can then isolate only the MYB species that are bound to CBP within the complex. At this point we can extract out the RNAs directly bound to MYB and sequence. This will produce an RNA binding map of MYB, but only when in complex to CBP (Figure 4.1C).

To reach the point of carrying out TA-CRAC, the cell line containing tagged CBP and tandem tagged MYB require generation. Firstly the Flag-TEV-6xHis (FTH) tag will be inserted endogenously into CBP using CRISPR/Cas9, generating a tagged CBP Jurkat line, we term CBP-Flag-TEV-6xHis (CBP$_{FTH}$) (Figure 4.1A). Next, this CBP$_{FTH}$ cell line will be taken further and the same methodologies will be used to insert a 6xHis-TEV-HA/Strep (HTHAS) tag onto MYB, generating the MYB-6xHis-TEV-HA/Strep (MYB$_{HTHAS}$:CBP$_{FTH}$) cell line.

This chapter will outline the establishment of CRISPR/Cas9 and Homology Directed Repair (HDR) as a method used to endogenously tag CBP in the form of a clonal cell line. This will include strategy design and a review of the experimental form in which the planned tagging would occur. Subsequently, establishing CRISPR/Cas9 and HDR experiments requires many levels of validation: from gRNA activity to the generation of clonal lines, which will be outlined. Finally, tagged CBP will be validated to confirm tag expression and tagged protein behaviour.

4.2 Designing a tagging strategy

CBP DNA and RNA construct design
Design of gRNA and ssODN

CBP gRNA validation
T7E1 assay and ICE analysis

Evaluation and
strategy change

Repair template validation
T7E1 assay, restriction digests and ICE analysis

Generation of clonal lines for tagged CBP
Single cell sorting, restriction digests and ICE analysis

CBP CRAC

Use tagged CBP line to tag MYB
Repeat same validation steps as for CBP

MYB TA-CRAC

Figure 4.2: Workflow of tagging CBP

Workflow showing the steps to be taken to correctly knock in CBP with the desired tag, including each step of validation. Through evaluating and revising our tagging strategy early on, we can adjust our approach as needed to ensure successful tag insertion. Then we can proceed to steps to generate a clonal cell line and use of this cell line to carry out CRAC. It also shows that after generation of the CBP tagged line, this line will be used to further tag MYB and then tandem affinity CRAC (TA-CRAC) can be carried out. ssODN - single stranded oligodeoxyribonucleotide.

4.2.1 CRISPR/Cas9 mediated genome editing

To firstly tag CBP endogenously and go on to further tag MYB, CRISPR/Cas9 was chosen as a method of DNA editing. CRISPR/Cas9 mediated genome editing involves Cas9 endonuclease forming a complex with a guide RNA (gRNA). This gRNA is designed to contain sequence homology with a specific short 20 bp region of DNA. The gRNA will then guide the Cas9 to the recognition sequence and Cas9 cleaves the DNA strand upstream of the protospacer adjacent motif (PAM) site (Figure 4.3)(Jinek et al. 2012).

After generating the double-strand break (DSB) within the DNA, various repair mechanisms can be employed by the cell. In particular the homology-directed repair (HDR) pathway requires a template for DNA repair machinery to copy in to the DSB, so specific DNA sequences can be supplied to be used as HDR templates, resulting in "knock-in" of specific sequences (Figure 4.3B). This can be done for single base substitutions or large insertions and deletions, for example deletions of over 65 kilobase pairs (kb) and insertions of over 5 kb have been reported (L. Zhang et al. 2015). The HDR pathway is not very efficient as the cell most often tries to repair DSBs through non-homologous end-joining (NHEJ) (Ran et al. 2013). NHEJ is error-prone, as this is a catch all mechanism to try and repair DNA quickly, without being so careful of sequence specificity. NHEJ most often results in various indels, causing downstream frame-shifts in amino acid encoding sequencing, often resulting in the generation of premature stop codons (Figure 4.3B). For this reason harnessing the NHEJ pathway for Cas9 mediated DSB repair can result in efficient "knockout" of specific protein-coding genes or regulatory elements within the genome (Mali et al. 2013; Ran et al. 2013).

140
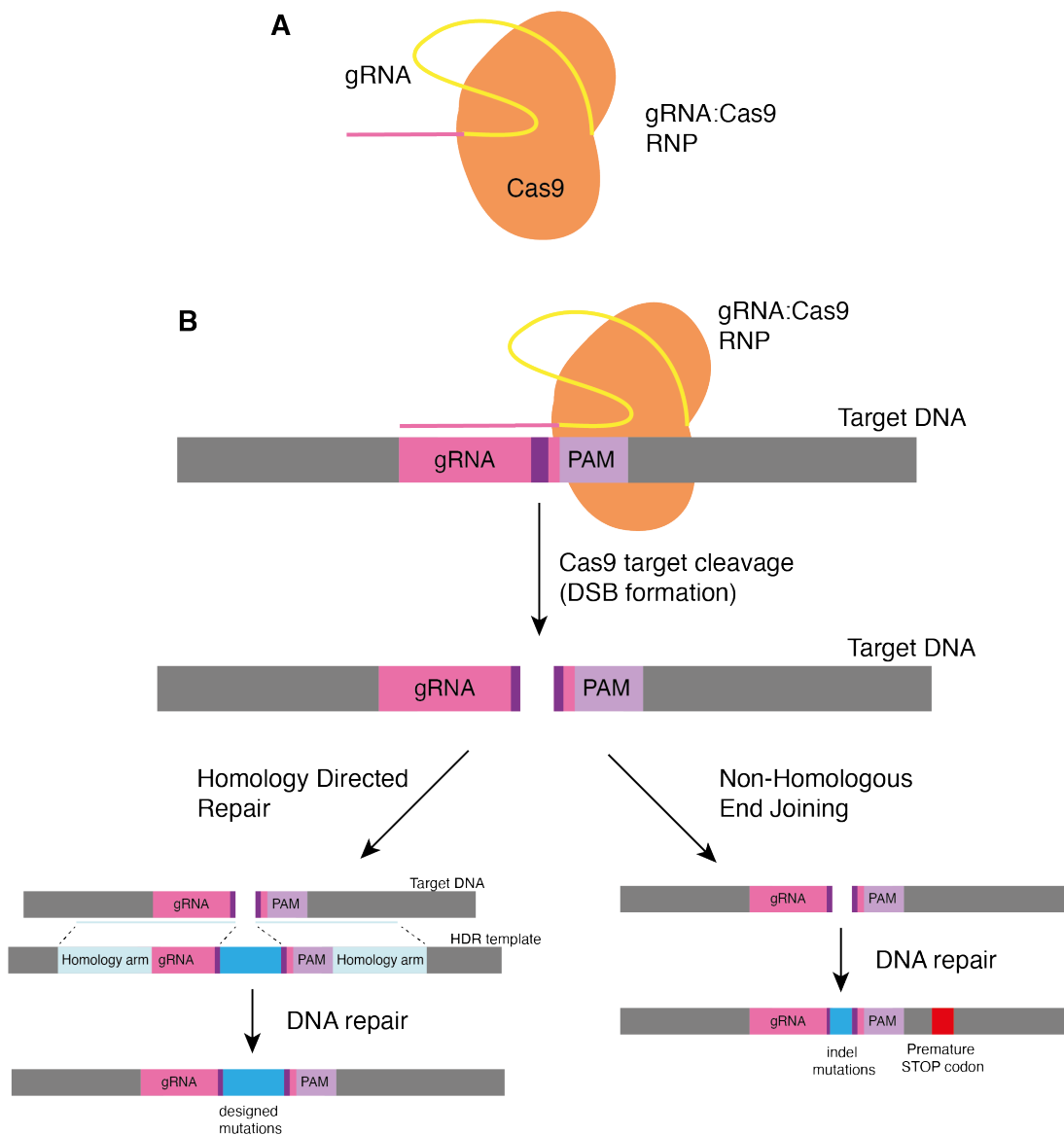
Figure 4.3: CRISPR/Cas9 generates precise double strand breaks in DNA sequences (A) Cas9 (orange) forms a complex with a gRNA (yellow and pink). (B) Cas9:gRNA complex is guided to designed cut site through homology between gRNA and recognition sequence within the DNA sequence (pink). DNA is cleaved by Cas9 upstream of the protospacer adjacent motif (PAM, light purple) and generates a double strand break (DSB, dark purple). DNA repair mechanisms are then employed by the cell to repair the DSB. Homology Directed Rpair (HDR) on the left shows how use of precise HDR templates creates specific designed mutations to the target DNA sequence. Non-Homologous End Joining (NHEJ) generates many indels to repair DSBs. Figure adapted from Ran et al. 2013.

4.2.2 CBP DNA and RNA construct design

There is an abundance of choice available for the strategy and methods used to carry out a CRISPR/Cas9 endogenous tagging experiment.

4.2.2.1 gRNA design

To carry out CRISPR/Cas9 mediated gene editing, it is possible to use plasmid encoded Cas9 with a corresponding encoded guide RNA (gRNA) or use a recombinant ribonucleoprotein (RNP) complex formed from recombinant Cas9 and gRNAs *in vitro*. Recombinant Cas9 used as an RNP complexed to a gRNA has a shorter half-life in the cell. Though this appears to be a negative for this strategy, as the Cas9 is in the cell for a decreased amount of time there is therefore less time for the Cas9 to cut in sites with decreased homology to the gRNA and therefore this strategy can reduce the frequency of off-target editing. It can also reduce the chances of the Cas9 repeatedly cutting at the designed PAM site, which can lead to errors in the designed knock-in (Okamoto et al. 2019).

However, plasmid encoded Cas9 and gRNAs do also offer benefits, especially in terms of cost. Once plasmids are cloned, large volumes can be prepared cost-effectively and with ease using scalable plasmid purification kits. Use of plasmids allows for selection of cells successfully containing expressed Cas9 or the HDR template to enrich the population for successful editing, either by utilising selection from the plasmid encoding the Cas9 and gRNA or the plasmid encoding the HDR template. Methods of selection for positively edited cells are very useful in the cases of cell populations with low editing efficiency, as they can be employed to select for cells with the correct edit, increasing the proportion of correctly edited cells within the population. It is possible to use an antibiotic selection based approach, for example using a Cas9 plasmid containing a puromycin resistance casette (Ran et al. 2013). After antibiotic addition, only cells expressing plasmids with resistance casettes will survive, meaning that the cell population would therefore also be expressing Cas9, thus enriching the population for successful editing. However previously in our lab, we had found that Jurkat cells struggled to survive well under antibiotic selection and where a different method was possible it was best to explore alternatives (Celadova 2022). One such other option would be to utilise fluorescence, for example using a Cas9 fused to a green

142

fluorescent protein (GFP). Cells expressing GFP-Cas9, could be taken for Fluorescence Activated Cell Sorting (FACS) to collect all GFP positive cells. This would then be a different method to also collect a population expressing Cas9 in the cell.

4.2.2.2   Repair template design

There is also choice available in terms of repair template, which contains the desired edit and flanking regions of homology to the target sequence, termed homology arms (Figure 4.3B). Size of insertion sequence appears to be the determining factor in terms of homology arm length. Insertions of <130 bp can use small homology arms of 50 bp, whereas larger insertion sequences are recommended to use over 500 bp of homology flanking the desired edit. Due to the varying size, these small homology arm repair templates can be used as synthesised single-stranded oligonucleotides (ssODNs) and can offer increased efficiency with shorter insertion sequences. Larger repair templates are most commonly used in the form of a double stranded plasmid, as longer homology arms are favoured when inserting larger sizes of insertion sequences (Paix, Schmidt, and Seydoux 2016).

As we are adding tags to our proteins of interest, there is also possibility to add in additional tags that may aid in selection of positively edited cells. For example, a Halotag (Promega) is a 33 kDa protein tag with a wide variety of ligands that are commercially available. When these ligands bind to the Halotag they bind covalently meaning that they are bound irreversibly. This can offer huge advantages in terms of selection as ligands with fluorescent dyes can be added to the population of cells. FACS could then be used to sort positively labelled cells and the population produced would only be those expressing the Halotag (Promega).

4.2.2.3   Other experimental considerations

There is also choice in the method of transfection of reagents into the cell. For both plasmid encoded and recombinant forms, Cas9, gRNAs and repair templates can be transfected into the cell using electroporation or chemical transfection. Previously in our lab we had noticed that chemical transfection methods lead to poor survival rates in Jurkat cells and much greater success was yielded using electroporation based methods (Celadova 2022).

4.2.3 gRNA validation

4.2.3.1 T7 Endonuclease 1 assay

The T7 Endonuclease 1 assay is often the first method employed to validate that Cas9 is cleaving DNA in the cell.
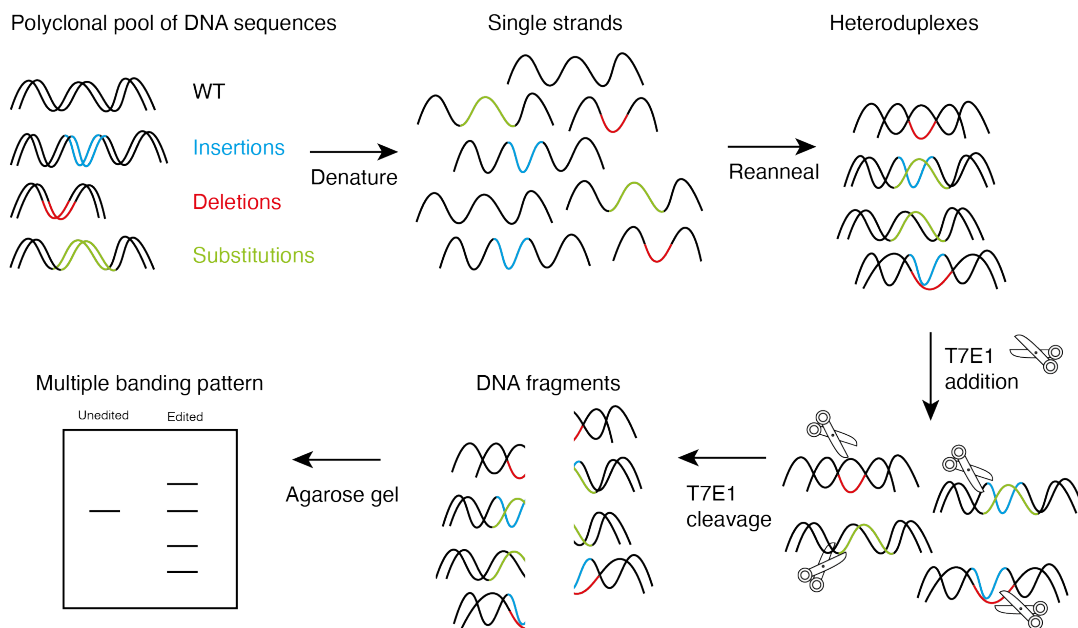
Figure 4.4: T7 Endonuclease 1 assay

T7E1 assay. Mixed and polyclonal pool of DNA sequences containing WT sequences (black), sequences with insertions (blue), sequences with deletions (red) and sequences with substitutions (green) are denatured to single strands. Upon heating strands re-anneal, with mismatches found within heteroduplex strands. T7E1 enzyme recognises and cleaves the mismatches and after running on an agarose gel produces a pattern of multiple bands.

Cells are first transfected with a Cas9 and gRNA with DNA being extracted after 48 hours to allow time for the Cas9 to cleave at the target site (Figure 4.4). The target region is amplified using PCR with primers designed asymmetrically around the double strand break (DSB) site. The PCR products are then heated to denature the DNA and cooled to reanneal, forming heteroduplex DNA strands. Where Cas9 has cleaved the target site, it is very likely to have been repaired through the NHEJ pathway, introducing a range of mutations, especially indels. Therefore, due to the variety of

144

mutations and there being some wild-type (WT) sequence remaining, the heteroduplex strands of DNA are mismatched. The T7E1 enzyme cleaves these mismatches, with data from successful T7E1 assays demonstrating either multiple bands or a smear of bands on an agarose gel (Sentmanat et al. 2018).

4.2.3.2   Inference of CRISPR Edits (ICE) analysis

The same PCR products generated for the T7E1 assay can be sent for Sanger sequencing and the traces can be analysed. Inference of CRISPR Edits (ICE, Synthego, Conant et al. 2022) software analyses .ab1 files produced from Sanger sequencing of a control sample and an edited sample, alongside the sequence of the gRNA. The software compares the two Sanger traces to determine if successful cleavage has occurred. Using the gRNA sequence the DSB site can be identified, enabling comparison of the homogeneity of the sequencing trace after the designed cut site in the test sample against the control. If the two traces are identical, this implies that no successful Cas9 cleavage has occurred. Whereas if there is a large amount of variance in the edited sample, this indicates Cas9 has cleaved successfully and either erroneous mutations have been introduced due to NHEJ, or a designed edit may have been knocked in.

4.2.4   Repair template validation

After successful gRNA validation, we can proceed with the addition of a repair template to introduce the tag to be knocked in.

After transfection of the cells with Cas9, gRNA and ssODN, the T7E1 assay and ICE analysis can be repeated to identify if any successful edits have occurred. PCR using a forward primer binding directly within the tag will produce bands on an agarose gel if there is any amount of the designed tag present in the extracted DNA, and these bands can be extracted and sent for Sanger sequencing.

This population of cells are likely to be polyclonal as they would be a heterogeneous mix of WT, successfully edited and edited cells with mutations.

4.2.5 Generation of clonal lines for tagged CBP

To use our tagged CBP line for CRAC and go on to develop it for TA-CRAC, it will be necessary to generate a clonal line to ensure that all of the CBP species are tagged in the same way and to ensure we have enough tagged copies of CBP to be used in the pulldown experiments.

4.2.5.1 Single cell sorting

The polyclonal population will be single cell sorted using a BDFACSMelody into 96 well plates. Four 96 well plates will be sorted, allowing the potential for 384 clones to survive. The cells will then be allowed to grow for two weeks until they are of a sufficient colony size that they are ready for expansion. Clones will then be expanded into increasingly larger sized well dishes until sufficient material can be obtained for DNA extraction and screening.

4.2.5.2 Clone screening

DNA can then be extracted using QuickExtract soluction (Lucigen) according to the manufacturer's instructions for each clone. This can then be used for screening by PCR (Section 2.2.3) using primers to amplify the C-terminus of CBP (Table 2.4). If the tag is inserted correctly we will be able to see a shift of band size corresponding with tag size, and the number of bands will allow us to determine the number of alleles that have been successfully edited. If two bands are present, one running at the height of the WT sequence and a second having shifted to be the size including the tag, this indicates a heterozygous clone. But if only one larger band corresponding with the correct size for tag insertion is visible, this is indicative of a homozygotous clone. These bands can then be excised from the gel and subsequent Sanger sequencing will allow visualisation of the DNA sequence to identify if the designed edits have been faithfully knocked-in.

4.2.6 Validation of clonal lines for tagged CBP

After identifying clones of interest that are positive for tag insertion, several rounds of validation need to be carried out. This is to ensure the tags are being expressed

correctly and are not impeding protein function or health of the cell.

4.2.6.1 Validation of DNA sequence

It is necessary to rescreen the clones to confirm the tag has been knocked in. PCR will be repeated using primers to amplify the C-terminal region and visualised on an agarose gel to check for the correct banding pattern. Subsequently this PCR product will also be sent for Sanger sequencing to check that the correct edit has been knocked in and there are no unwanted mutations

4.2.6.2 Validation of tag expression

To determine that the tags are being correctly expressed onto the protein of interest, reverse transcriptase quantitative PCR (RT-qPCR) with primers designed within the tag sequence will confirm the tags were being expressed on an mRNA level. RT-qPCR data could be complemented by western blotting, to determine expression of full-length tagged protein. Blotting for the protein should reveal a small shift in size with the tagged compare to WT untagged protein and blotting using antibodies against the inserted tag should reveal whether full length protein is expressing the desired tag.

4.2.6.3 Validation of protein function

After confirmation that the tag is in place and being expressed correctly, it is necessary to determine that the tags are not having any affect on protein function, protein complex interaction or impacting the growth and health of the cells.

Firstly, using RT-qPCR expression levels of CBP can be compared in tagged and untagged WT cells to determine if the tag is having any adverse affects on protein expression.

After this, Co-IP of CBP and also using the tags would confirm that tagged CBP is still able to form the CBP-TAL1 complex and that the tags are not impeding protein-protein interactions.

ChIP-qPCR to compare tagged and untagged protein would additionally show us whether the chromatin binding activity of tagged CBP is affected.

Taking this RT-qPCR, Co-IP and ChIP-qPCR data together will show that the tags on CBP are not impeding function of the protein in any way and that complex

formation is not impeded and neither is its ability to bind to chromatin.

4.2.6.4 Validation of no off-target editing taking place

Whilst validating these clones, it is also important to check that there are no unwanted off-target edits within the cell.

The method we will use to examine off-target editing is to return to the original gRNA design software (section 2.4.1.1). The UCSC gRNA software details not only the intended gRNA recognition site, but additionally the sequences with the next highest homology to the gRNA sequence. These would therefore be the regions most prone to off-target activity, as they are the sites most likely to be recognised by the gRNA and therefore have Cas9 recruited and cleave. PCR and Sanger sequencing of these regions will identify if any mutations have been introduced as a result of off-target editing.

4.3 Generating a tagged CBP cell line using Cas9 RNP and ssODN

Initially we sought to tag CBP with a Flag-TEV-6xHis tag and to do this we chose to utilise recombinant Cas9-gRNA RNP, as it has been shown that editing efficiency was high through use of an RNP (Okamoto et al. 2019). It has also been shown that the RNP method works well with use of an ssODN repair template, and when the two methods are used in conjunction, this greatly increases the editing efficiency of the desired knock-in (Liang et al. 2017). As our insertion sequence was 90 bp, we chose to use an ssODN for a repair template with 50 bp of flanking homology. This was also because the RNP complex and ssODN could be easily transfected into Jurkat cells using the Neon electroporation system (Invitrogen).

After validating the RNP and ssODN strategy, we would then be able to review our approach and switch to plasmid based methods or alternative methods for selection if necessary.

4.3.1 gRNA design

We targeted the C-terminus of CBP to minimise disruption to the protein that could arise due to frame shift mutations from the CRISPR editing. A gRNA was designed with the best on and off-target scores using multiple design and evaluation tools (as described in 2.4.1.1) as close to the C-terminus of CBP as possible. This targeted the genomic locus of chr16:3,727,726 (Figure 4.5A).
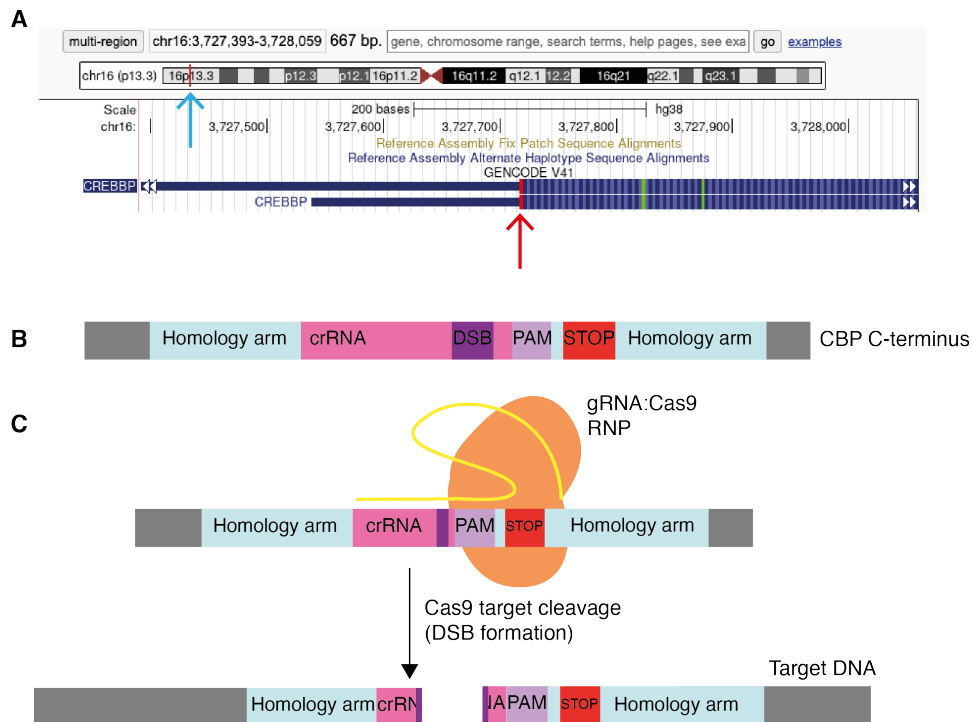
Figure 4.5: gRNA targets CBP C-terminus
(A) UCSC genome browser view of the genomic locus chr16:3,727,726. Blue arrow indicates the locus on the chromosome, with the red arrow indicating the locus at the C-terminus. Co-ordinates listed in Appendix 8.1 (B) Schematic of CBP C-terminal region, note schematic shows 5'-3' direction. Sequences of note are highlighted including crRNA homology site (pink), the double strand break (DSB) site (dark purple), protospacer adjacent motif (PAM) site (light purple) and stop codon (red). (C) Cas9 (orange) complexed with the gRNA (yellow) recognises target DNA, through the crRNA binding to a homologous region of DNA sequence, and tracrRNA binding to the PAM site. This allows the Cas9 to cleave the DNA.

149

4.3.2 gRNA validation
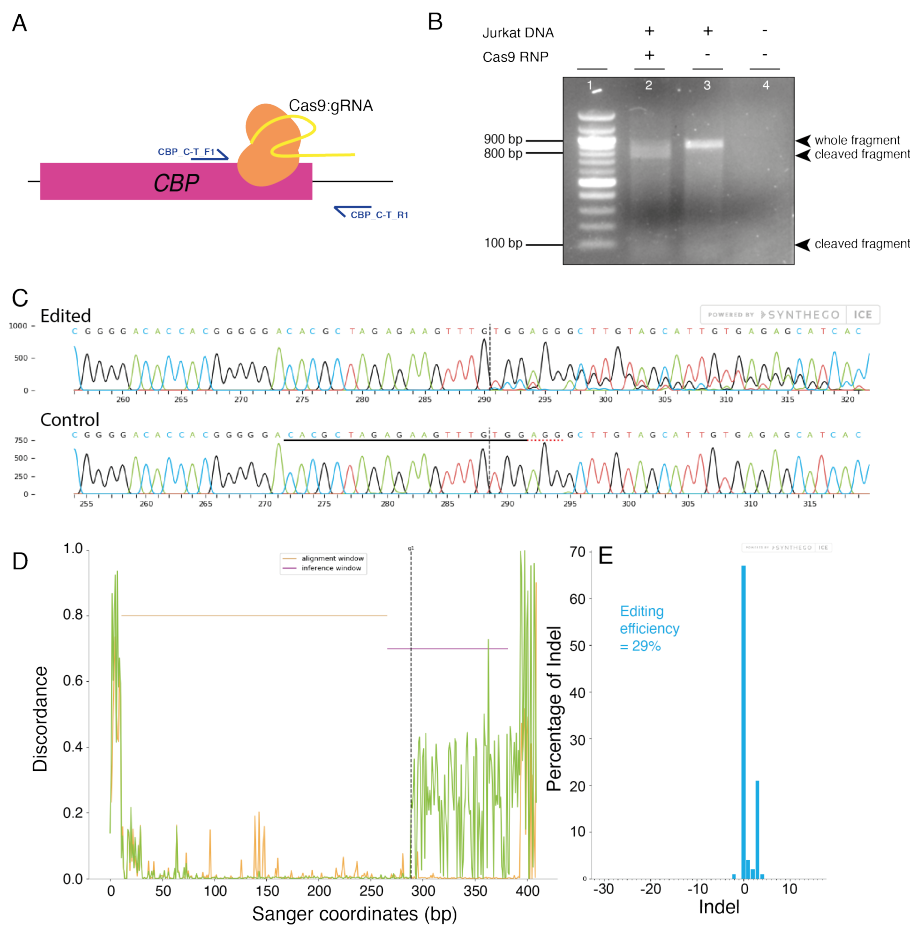
Figure 4.6: CBP gRNA validation

(A) Cas9:gRNA RNP complex targets CBP C-terminus, with PCR primers (blue arrows) amplifiying the C-terminal region of CBP and the UTR (B) 1.5% agarose gel showing T7E1 assay in Jurkat cells. DNA from edited Jurkat cells runs as a smear compared to unedited cells. Arrows indicate bands of interest, with corresponding sizes marked upon the ladder. Marker is 100-1500 bp ladder. (n=1) (C) Comparison of Sanger traces for the CBP edited sample (upper trace) against the CBP control sample (lower trace). The crRNA sequence (black bold line) has been aligned, indicating the PAM site (red dashed line) and the DSB site (black dashed line) (ICE software, https://ice.synthego.com/#/, n=1). (D) Discordance plot showing the difference in signal between the bases found in the sanger traces of the control (orange) against the edited sample (green). After the cut site (black dashed line) the discordance between the samples increases. (E) Distribution of size of indels (bp) in the edited sample and displays an editing efficiency of 29%.

150

4.3.2.1 T7 Endonuclease 1 assay

The gRNA was complexed with Cas9 recombinant protein to form an RNP and this was transfected into Jurkat cells using the Neon electroporation system. The cells were left to grow for 72 hours and were then harvested for DNA extraction.

To check that Cas9 was being transfected into the cell and was actively cleaving DNA, the T7E1 assay was carried out (Figure 4.6B).

PCR was used to amplify the C-terminal region of CBP in edited and unedited cells, with these primers designed asymmetrically around the DSB site. The total PCR product with no edits was expected to be 900 bp, with the DSB site located 800 bp into this. Therefore, use of the T7E1 assay in successfully edited cells should demonstrate cleavage of this PCR product at 800 bp.

There was a distinct smear visible between 800-900 bp in the PCR product from edited cells, after having been denatured and reannealed with the addition of T7E1 enzyme (Lane 2, Figure 4.6B). This indicated that mismatches were present 800 bp into this PCR product as a result of error-prone repair mechanisms trying to heal the DSB. Compared with PCR product from DNA extracted from cells that were not transfected with the Cas9 RNP, there was a band of 900 bp only, as this was the size of the amplified region (Lane 3, Figure 4.6B). This was the expected result as the induced DSB should be 800 bp into the 900 bp PCR product, which indicated that the Cas9 RNP had induced DSBs into our region of interest successfully.

4.3.2.2 Inference of CRISPR Edits (ICE) analysis

After implying that DSBs were being induced correctly in the cell, we next wanted to check that the correct sequence was being targeted and the DSB site was as designed.

Sanger sequencing of PCR products of the target area were analysed using ICE software (Synthego, Figure 4.6C, D and E). Figure 4.6C showed the comparison of the two Sanger traces, the upper showing the edited and the lower showing the control sequence. When the two traces were compared, in the edited sample after the DSB site the Sanger sequencing trace became very messy with many combinations of nucleotides being present at each position, whereas the WT sequence remained clean with only one nucleotide having a peak. This was likely due to Cas9 having cleaved at the designed DSB site and in order for the cell to combat the DSBs, error-prone repairs have been

made inducing many combinations of erroneous mutations including frameshifts which result in misaligned sequencing reads downstream of the cut-site. This suggested that our gRNA was cleaving at the correct site and generating DSBs where designed.

Figure 4.6D showed another ICE output, measuring the discordance in sequence between the edited and control sample. In the edited sample (green) after the cut site the discordance in sequence between the WT sequence (orange) increases, suggesting that many mutations had been introduced as a result of DNA error-prone repair pathways.

Figure 4.6E showed a histogram of the varying insertion and deletion (indels) mutations that the ICE software can infer from the Sanger traces. The histogram showed the distribution of the different sizes of edits that have been introduced at this site, and the proportion of total sequence with each size of mutation. Almost 70% of the trace and sequence did not appear to differ between the edited and the control sequences, meaning no mutations had been introduced in nearly 70% of the population. Therefore, the software had calculated an editing efficiency of 29%.

4.3.2.3   gRNA validation conclusions

Taken together the T7E1 data and the ICE analysis implied that Cas9 was being guided to the designed DSB site, and erroneous mutations were detectable using these methods. Therefore, we hoped that these methods would be able to validate tag insertion and to analyse CRISPR/Cas9 editing efficiency in downstream experiments, not solely for gRNA validation.

Taking this forward, the next step was to add in the repair template and to attempt to edit in the Flag-TEV-6xHis tag.

4.3.3   ssODN design

After the gRNA had been shown to be cutting and at the correct site, the previous protocol of transfecting Jurkat cells with a Cas9 gRNA RNP was repeated, with the addition of a homology directed repair template in the form of a single stranded oligo-deoxyribonucleotide ssODN. An ssODN was chosen for the form of repair template as this was the most efficient for short sequences and worked well with RNP Cas9 and gRNA mediated editing (as discussed in Section 4.2.2.2, Okamoto et al. 2019).

The ssODN repair template was designed using 50 bp of homology flanking the DSB site with the Flag-TEV-6xHis tag sequence present prior to the stop codon to ensure the tag would be included with transcription of CBP. A silent mutation was also included within the PAM site in the ssODN, this meant that upon successful insertion of the tag the nucleotide sequence of the PAM would change, meaning that the Cas9 would not be able to cleave it again and reinsert the tag multiple times. As it is a silent mutation it would not change the amino acid sequence of CBP (Figure 4.7)(Okamoto et al. 2019), (ssODN sequence described in Appendix section 8.1.2.1).
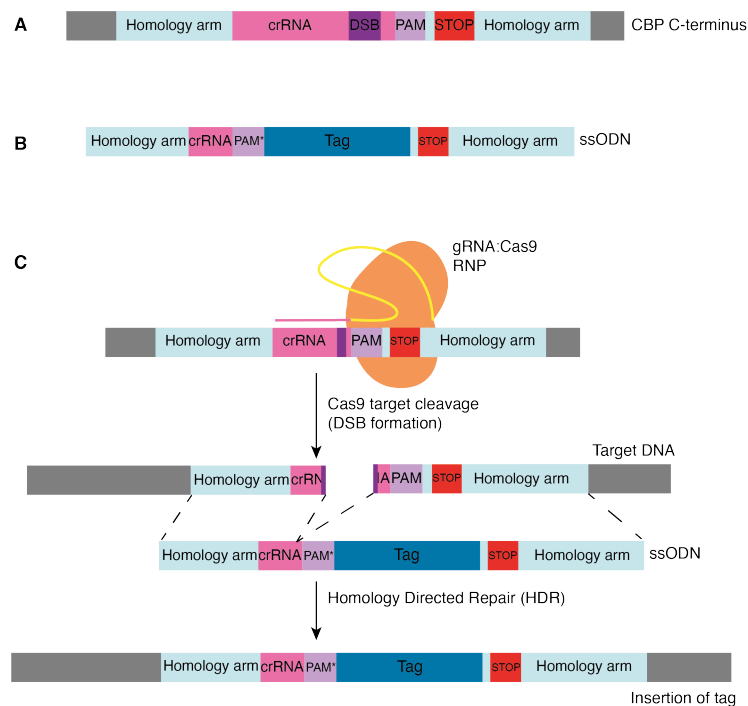
Figure 4.7: CBP targeting gRNA, ssODN and experimental approach design
(A) CBP C-terminal region, highlighting regions of note including crRNA (pink), the double strand break (DSB) site (dark purple), protospacer adjacent motif (PAM) site (light purple) and stop codon (red). (B) ssODN design. PAM* indicates silent mutation within the PAM. (C) Cas9 (orange) complexed with the gRNA (tracrRNA yellow, crRNA pink) recognises target DNA through homology in the crRNA sequence and tracrRNA recognising the PAM site. After cleavage by Cas9, the ssODN aligns to the complementary region of DNA, homology regions (yellow) are recognised and through the HDR pathway, the tag (dark blue) is copied and inserted into the DNA sequence.

4.3.4 Validating the ssODN repair

4.3.4.1 T7E1 assay

After co-transfection of the Cas9 RNP and ssODN, the T7E1 assay was repeated to determine if any successful editing of the tag into the C-terminus of CBP was determinable (Figure 4.8). PCR amplification across the C-terminus of CBP would produce a band of 900 bp if amplified from WT sequence. As the tag was 100 bp in length, PCR amplification of successfully edited cells would produce a band of 1000 bp (Figure 4.8A). Treatment with T7E1 enzyme would then produce 800 bp fragments for both WT and edited cells, however for edited cells there would be a visible fragment at 200 bp (Figure 4.8A).

PCR product from untransfected cells gave a band of expected height of 900 bp and addition of T7E1 resulted in no difference in the sizes of PCR products, suggesting no editing had occurred. PCR products from cells transfected only with the Cas9-gRNA RNP, not treated with T7E1 also showed a single band at 900 bp, whereas with the addition of T7E1 enzyme showed two distinct bands at 800 bp and 900 bp. This recapitulated the data in Figure 4.6B when validating our gRNA. When DNA was extracted from cells that were edited with the ssODN as well as the Cas9-gRNA RNP, a similar pattern was observed. Treatment with T7E1 showed two distinct bands at 800 bp and 900 bp, showing that T7E1 was cleaving at the designed DSB site, implying that some level of editing had taken place. However, in both lanes with ssODN, there was no visible band at 1000 bp to indicate a full-length PCR product containing the inserted tag and after T7E1 treatment there was no visible band at 200 bp. (Figure 4.8).

Therefore, the T7E1 assay at this stage implied that addition of the ssODN to the RNP transfection reaction was not affecting Cas9 activity, and Cas9 was active in the cell population. However, there was no indication that the FTH tag had been successfully edited into the cell population.
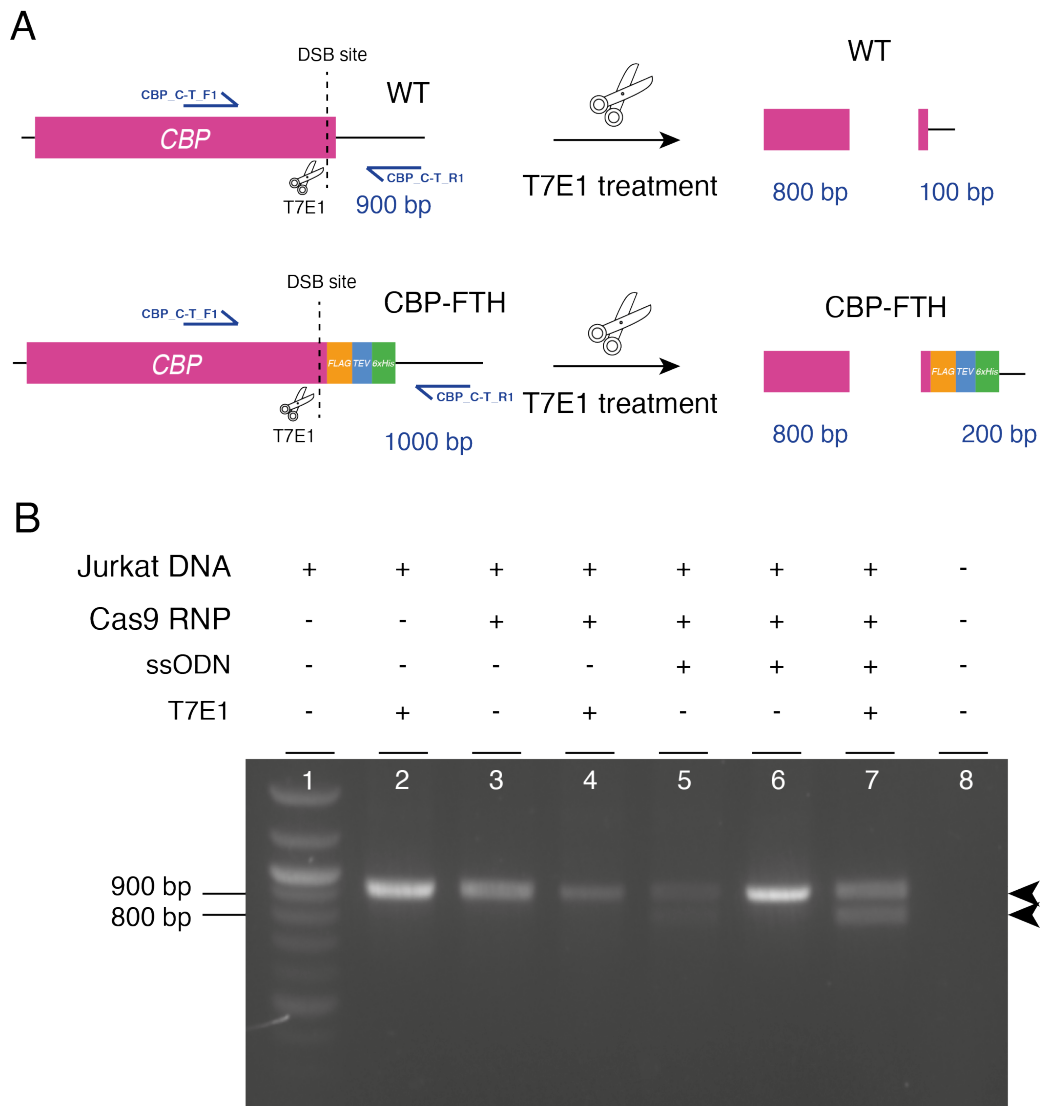
Figure 4.8: CBP ssODN and gRNA validation.
T7E1 assay in polyclonal edited Jurkat cells. (A) T7E1 assay with WT (upper) and successfully tagged DNA (lower). PCR products indicated by purple arrows, with fragment sizes denoted accordingly. T7E1 treatment results in different sized fragments (right hand side). (B) PCR products ran with and without T7E1 treatment, for non-edited cells in lanes 2 and 3, cells edited only with the RNP in lanes 4 and 5, and cells with the RNP and the ssODN in lanes 6 and 7. 200 ng DNA loaded into each lane. Arrows indicate bands of interest, with corresponding sizes marked upon the ladder. Marker is 100-1500 bp ladder, n=1.

4.3.4.2 ICE analysis

The T7E1 assay was followed by repeating Sanger sequencing of the edited areas and analysing traces by ICE. At this stage we were testing whether any proportion of cells transfected with Cas9, gRNA and ssODN contained the desired edit, or if there were any noticeable differences between RNP only and RNP + ssODN treated cells indicating any affects on Cas9 cleavage or on repair mechanisms utilised by the cell.

From the ICE analysis, looking firstly at the comparison of the Sanger traces from cells edited with the RNP only and cells edited with the RNP and an ssODN repair template, the traces appeared similar (Figure 4.9A and 4.9B). For both edited samples after the DSB, there were more peaks of each base at each Sanger coordinate, and it was observable that there would be difficulty producing a consensus sequence due to the complexity of the trace. Although, as our cells were polyclonal, we were unable to gain any indication if any successful editing had occurred.

If the tag had been successfully edited into the sequence using the ssODN as a HDR template, we would have expected to observe a large amount of discordance in sequence between the edited and control cells because at the site of tag introduction these two sequences would be completely different. However, there was a greater amount of discordance in DNA sequence between unedited and edited cells when treated with RNP only compared to those treated additionally with the ssODN (Figure 4.9C and 4.9D). This was not what was expected and implied that within the RNP only treated cells there were more error-prone repair mechanisms being utilised by the cell therefore introducing more erroneous mutations at this site as a consequence, compared to cells treated with both the RNP and ssODN. This could have implied that the ssODN had been having an effect on repair of the DSBs induced by the Cas9. Potentially due to the homology arms of the ssODN binding either side of the DSB and bringing the two cleaved pieces of DNA together, more efficient faithful repair of the DNA break was occurring. As the discordance between control and edited cells within the RNP and ssODN treated cells was lower compared to that observed within the RNP only treated cells, this did not provide any indication that the ssODN was causing correct editing of the designed tag sequence.

156
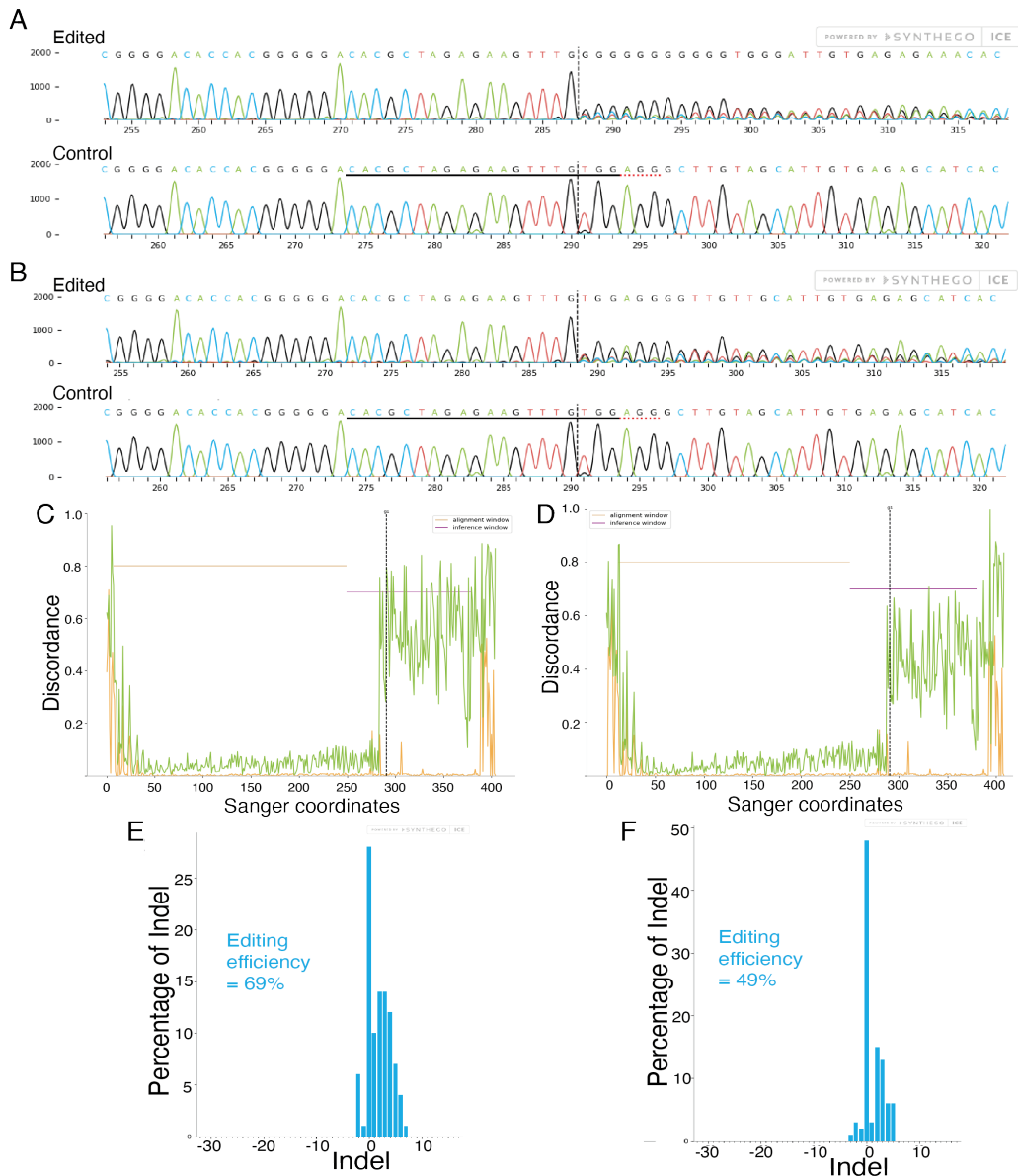
Figure 4.9: ICE analysis of CBP edited cells with and without repair template (A and B) Comparison of Sanger traces between edited and control samples. (A) RNP only against a control unedited sample. (B) RNP and ssODN repair template against a control unedited sample. (C and D) Discordance plots showing the difference in signal between the bases found in the sanger traces of the control (orange) against the edited sample (green). Traces taken from cells edited with RNP only are shown in (C) and RNP with ssODN repair template are shown in (D). (E) Distribution of indels in RNP only edited sample. (F) Distribution of indels in RNP and ssODN edited sample. (RNP only n=2, RNP + ssODN n=1).

When examining the distribution of base pair sizes for insertion and deletion (indel) mutations introduced in the edited samples, we found that editing with the RNP resulted in a distribution of indels ranging from -2 to +8 bp. Furthermore, a much greater percentage of sequences contained an indel compared to no mutation or an indel size of 0 bp, resulting in an editing efficiency of 69% (Figure 4.9E). However when analysing the RNP and ssODN sample, indels ranged in size from -3 to +5 bp, but the percentage of the pool containing no mutation or an indel size of 0 bp was almost 50% (Figure 4.9F). Therefore the sample with ssODN was resulting in a lower editing efficiency, as fewer error-prone repairs had been introduced in the DNA.

If this analysis was correct, this implied that the ssODN was successfully being incorporated into the cells, reaching the DSB site within the nucleus and potentially having some form of effect on repair of DSBs. However, it was not possible to infer that the ssODN had knocked-in the desired edit.

4.3.4.3   ssODN validation conclusions

Taken together, the T7E1 and ICE data showed upon ssODN addition, Cas9 was still able to reach and target the correct site for cleavage. However, there was no evidence that the desired edit of a Flag-TEV-6xHis tag had been inserted in any way.

A recent study using CRISPR/Cas9 to knock-in a similar length edit into Jurkat cells observed only 2% of clones containing the desired edit (Borowicz et al. 2020). Our cell population at this stage was polyclonal, and this suggested that there could be cells containing the desired edit, but these were at a very low proportion.

To proceed we narrowed down the polyclonal population to try and identify any cells that did contain the edit through single cell sorting to obtain clonal cell lines. Although this should have produced many clones that required screening, if 2% of clones did contain the desired edit, these could be identified, screened and would allow much easier visualisation of successful tag insertion. It is necessary to obtain clonal lines, as to be able to carry out TA-CRAC, a homozygous tagged CBP is required to ensure that we can identify RNA binding to all CBP proteins throughout the cell.

4.3.5   Single cell sorting of edited cells to obtain clones

The polyclonal population of Jurkat cells previously transfected with Cas9 and gRNA RNP plus ssODN for editing were single cell sorted using Flow Cytometry. This sorted one single cell into individual wells of a 96 well plate, allowing expansion into clonal lines for screening to identify if any editing of the desired tag was taking place (Figure 4.10A).

After single cell sorting, clones were grown up and expanded until sufficient number of cells could be used for DNA extraction. 384 single cells were sorted but viability was poor, with only 20 surviving to develop into a clonal population. As the tag did not contain any marker for selection, clones were screened using a PCR and restriction enzyme based approach. By amplifying the tag using PCR, the amplification products could be digested using a restriction enzyme BstXI, as a BstXI restriction site was found only within the inserted Flag-TEV-6xHis tag. As a lower number of clones survived to this stage than expected, all were DNA extracted and PCR amplified and digested with BstXI (Figure 4.10B and C).

Clones that did not contain the tag were expected to have a band at 900 bp. If the clones did contain the tag, the undigested PCR product would run at 1000 bp, and if digested by BstXI, there would be a band at 900 bp and a band at 100 bp (Figure 4.10B). The polyclonal population shown in lanes 2 and 3 had an identical single band at 900 bp implying that BstXI digestion had little to no effect (Figure 4.10C). Clones 1-6 were shown with and without BstXI digest. There were no lanes with multiple bands visible and the pattern that should be observed in an edited clone would be a higher band in the undigested sample compared to the digested sample, and this was not an identifiable pattern for any of the clones.

In the expanded clones we had been unable to visualise our desired edit. We considered repeating our single cell sorting and screening to be able to screen more clones, as if the editing effiency was similar to 2% as described in other studies, we would need to screen at least 50 clones to obtain a positive clone (Borowicz et al. 2020). However, rather than persist with optimising this approach, we decided to adjust our editing strategy to enrich for positive editing efficiency.

159
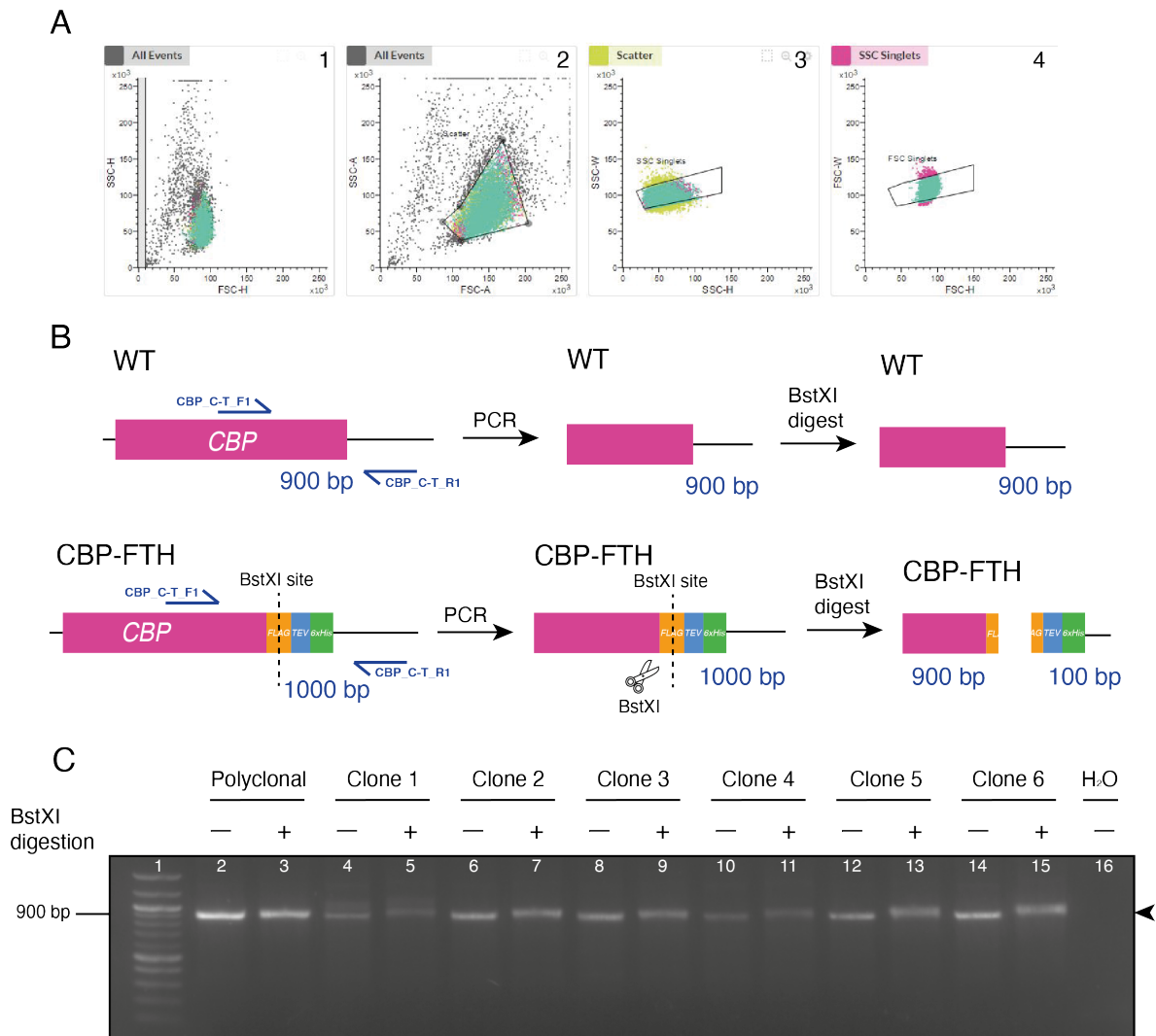
Figure 4.10: Obtaining and screening clonal lines.
(A) Single cell sorting of RNP + ssODN transfected population. Panels from left to right indicate increasingly smaller populations of cells, to obtain the single cell population. Panels 1 and 2 show side scatter against forward scatter for height and area and are used to gate for the viable and live cells. Panel 3 shows side scatter width against height and panel 4 shows forward scatter width against height. These were used to isolate the single cells. Mixed polyclonal Jurkat cells were sorted using a BDFACs Melody. (B) Expected PCR fragment sizes and size of fragments after BstXI digestion for WT sequence (upper) and CBP-FTH (lower) (C) PCR products with and without BstXI digest for the polyclonal population and 6 clones obtained from single cell sorting. 450 ng DNA loaded into each lane. Arrows indicate bands of interest, with corresponding sizes marked upon the ladder. Marker is 100-1500 bp ladder, n=1.

4.4 A strategy change: Using a plasmid based approach with GFP-tagged Cas9

As we were unable to validate that the tag was being edited into the Jurkat cells as designed, we explored changing our tagging approach. If we were to pivot to using a plasmid based approach, selection markers could be used that would enable us to select for either the plasmid encoding the Cas9 and gRNA or the plasmid encoding the HDR template. We considered using an antibiotic selection based approach, for example using a Cas9 plasmid containing a puromycin resistance casette (Ran et al. 2013). However previously in the lab we had found that the Jurkat cells struggled to survive well under antibiotic selection and where a different method was possible it was best to explore alternatives (Celadova 2022).

Use of a plasmid expressing a gRNA alongside Cas9 fused to GFP via a Thosea asigna virus 2A (T2A) self cleaving peptide allows successful transfection of host cells to be monitored using GFP signal (Ran et al. 2013). This approach allows co-transfection of the HDR template plasmid alongside the GFP-Cas9-gRNA plasmid and subsequent enrichment of a polyclonal cell population successfully transfected with GFP-Cas9 (Figure 4.11A and B). This would not give us a population of successfully edited cells, as we would not be applying any selection to the repair template plasmid, however it would enrich the population for Cas9 and therefore the proportion of successfully edited cells in the polyclonal population should increase (Figure 4.11C).

We also reconsidered HDR template design. An additional tag could be added to be used in enriching the population of edited cells and in the identification of positively edited cells. The addition of a Halotag (Promega) to the Flag-TEV-6xHis tag was chosen (Figure 4.11D and E), as the Halotag offers a wide range of commercially available ligands that covalently bind to a Halo-tagged protein. Halo-tagged CBP could be labelled through addition of fluorescent ligands and these cells could be single cell sorted using FACS. This would allow identification of positively tagged cells whilst screening fewer cells.

The Halotag is a large tag, with its addition bringing the size of the edit to be inserted to 1005 bp. As this is so large, it was no longer possible to use an ssODN repair template, therefore we used a plasmid based HDR template and increased our

homology arms to 600 bp to work in combination with the plasmid based Cas9-gRNA (Figure 4.11D). Larger length homology arms in a plasmid increase editing efficiency with larger insertion sizes and homology arms of >500 bp are recommended (Paix, Schmidt, and Seydoux 2016).

A
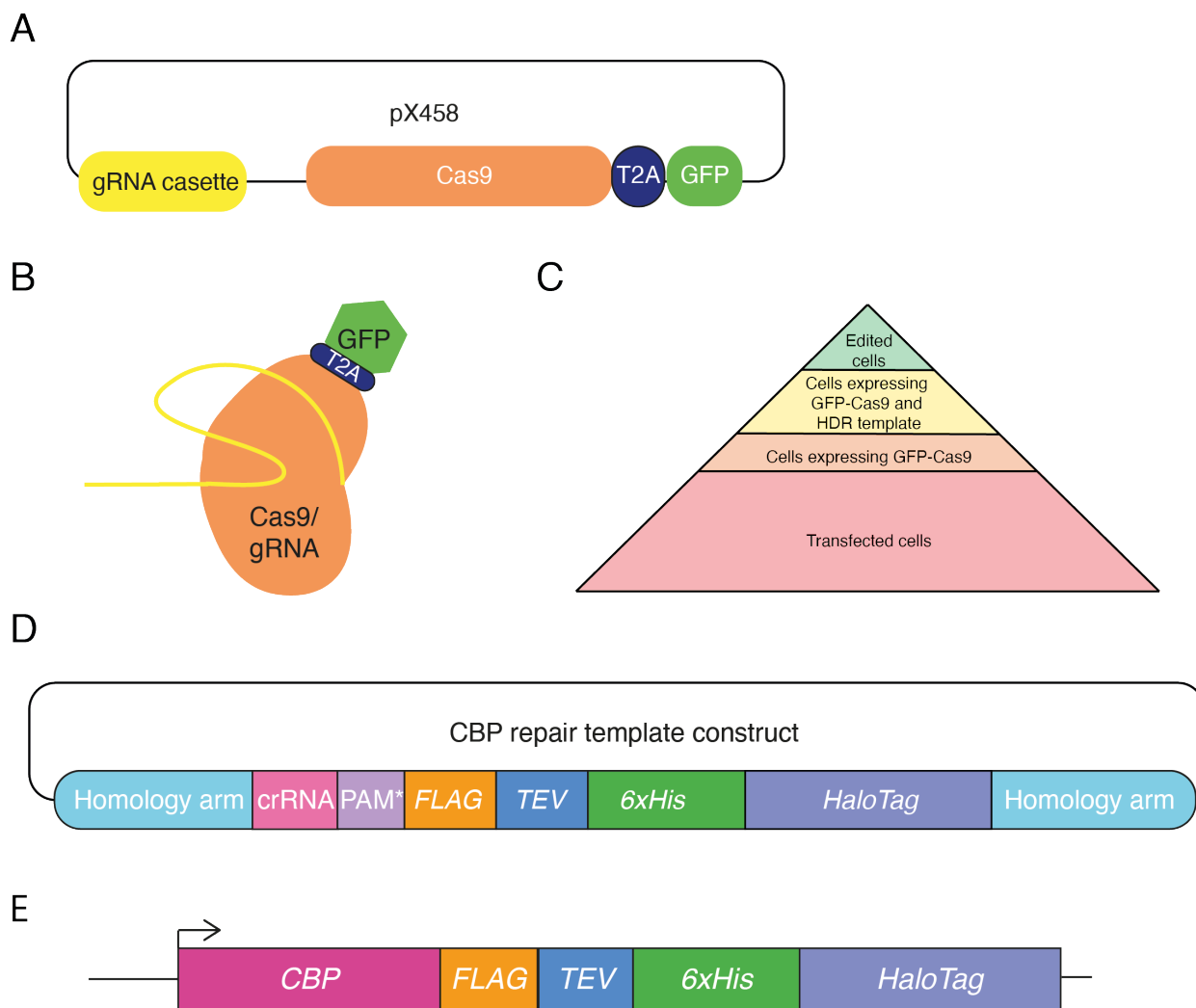
Figure 4.11: Use of a GFP tagged Cas9 to enrich for a population of edited cells (A) Composition of pX458, the plasmid used to encode a GFP-tagged Cas9 (gRNA yellow, Cas9 orange, T2A navy, GFP green) (B) Expressed pX458 as recombinant Cas9, with gRNA, T2A and GFP. (C) Pyramid depicting enrichment of transfected cells that contain Cas9 after sorting by GFP. (D)Composition of the HDR template used to tag CBP with a Flag-TEV-6xHis-Halotag. (E) CBP will have a Flag-TEV-6xHis-Halotag.

Therefore, we chose to proceed with a new strategy of a plasmid based approach for both the Cas9-gRNA and HDR template. The repair template plasmid we moved forward with was designed to edit into the C-terminus of CBP a Flag-TEV-6xHis-Halotag (CBP-FTHH, Figure 4.11D and E), with this plasmid denoted CBP-FTHH-RT (Table 2.3).

4.4.1   Validation of a plasmid encoded Cas9 and gRNA

To begin to design and plan our new CRISPR tagging approach, we used the same gRNA sequence validated previously and this was cloned into the pX458 GFP-Cas9 plasmid, denoted pX458-CBP (section 2.2.6.3, Table 2.3). We repeated the gRNA validations to check that this form of Cas9 and gRNA was still being expressed correctly and was able to cleave where it was designed to.

The T7E1 assay was repeated, with the edited cells in lane 3 producing two distinct bands at 800 bp and 900 bp (Figure 4.12C). DNA from unedited cells without T7E1 treatment in lane 2 and with T7E1 treatment in lane 4 show a single band at 900 bp. This replicated the data we had previously shown for our T7E1 assays using this gRNA in the form of an RNP (Figure 4.6B). This indicated that within the plasmid encoded format, Cas9 could cleave at the target site within the transfected cell population.

The same PCR products used were also sent for Sanger sequencing followed by analysis using ICE software (Synthego) (Figure 4.11D). The ICE software also recapitulated the RNP gRNA validation data (Figure 4.6C). The unedited control sample trace appeared very clean across the trace, however the gRNA sequence and DSB site was successfully identified at the correct locus. After the DSB site in the edited cells there was a very messy and complicated looking trace, showing that there was a large variety of sequence within this population due to error prone repair mechanisms trying to repair the DSB site.

Taken together, the T7E1 assay and ICE analysis data echo that shown previously of the RNP gRNA validation, and indicated that we were getting successful delivery to the cells of our gRNA-Cas9 plasmid. The Cas9 was then being guided correctly to the DSB site with this form of gRNA and was successfully cleaving at the correct site.

Therefore, it appeared the plasmid based delivery system of Cas9-gRNA was working efficiently and we were then able to proceed with the addition of the HDR template

to the transfection reaction to carry out the designed edit.
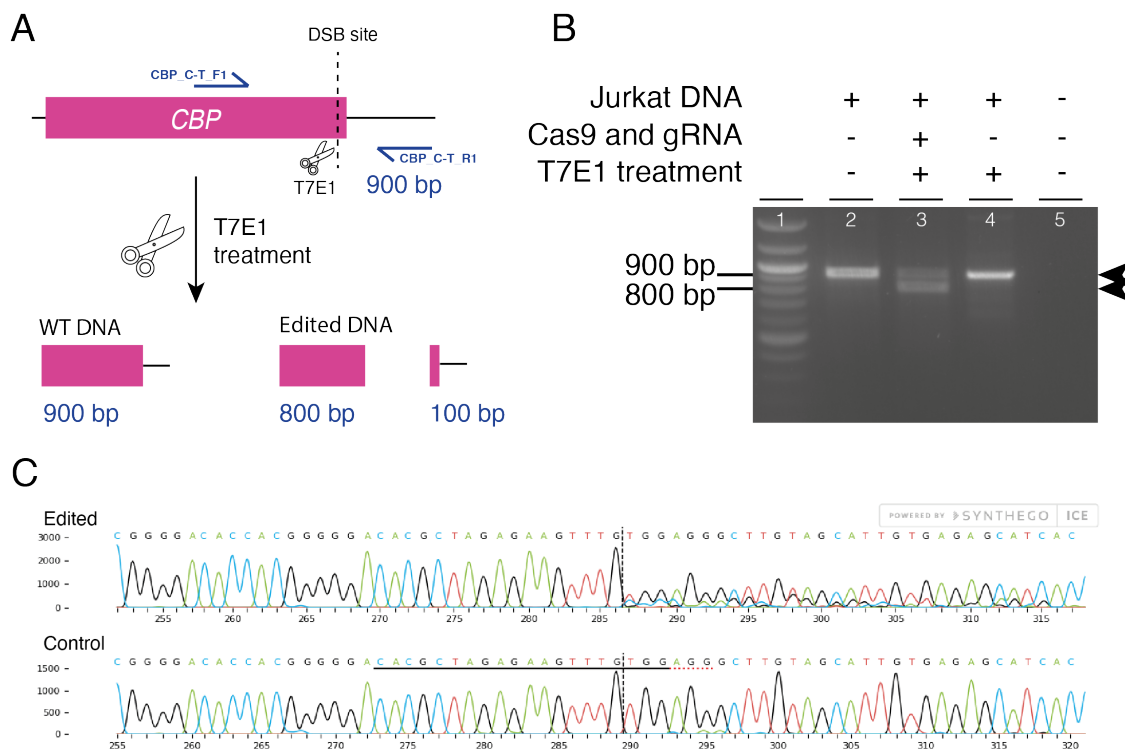
Figure 4.12: Flag-TEV-6xHis-Halotag repair template schematic
(A) T7E1 assay. PCR amplifies a 900 bp region, if T7E1 cleavage occurs this generates fragments of 800 bp and 100 bp. (B) T7E1 assay of pX458 encoding Cas9. Lanes 2 and 3 contain PCR product from edited DNA without and with T7E1 treatment respectively, lane 4 shows WT DNA treated with T7E1 as a control. (C) ICE comparison of Sanger traces between cells edited with GFP-Cas9 (upper) and control cells (lower). n=1.

4.4.2 Generating a polyclonal tagged CBP line

To carry out the CRISPR/Cas9 editing with the HDR template, Jurkat cells were co-transfected with pX458-CBP and CBP-FTHH-RT plasmids. using the Neon electroporation system (Invitrogen, section 2.3.5.2). 48 hours post-transfection, cells were sorted by FACS using the GFP signal to enrich for successfully transfected cells (Figures 4.13A and B). The resulting polyclonal population of cells was positive for Cas9 expression and therefore should contain a higher proportion of correctly edited cells.

The sorted cell population was then screened for successful editing of the insertion via PCR using primers that amplified directly within the tag. If there were any cells that did contain the tag, a band would be visible at 1100 bp as the positive control amplification of the HDR template plasmid produced a 1100 bp band (Figure 4.13C, Lane 2). PCR product from DNA extracted from the polyclonal cell population showed a band at 1100 bp, indicating that some proportion of the polyclonal population at contained the desired edit. This was the first observed evidence that there was some level of our designed editing occurring in these cells.
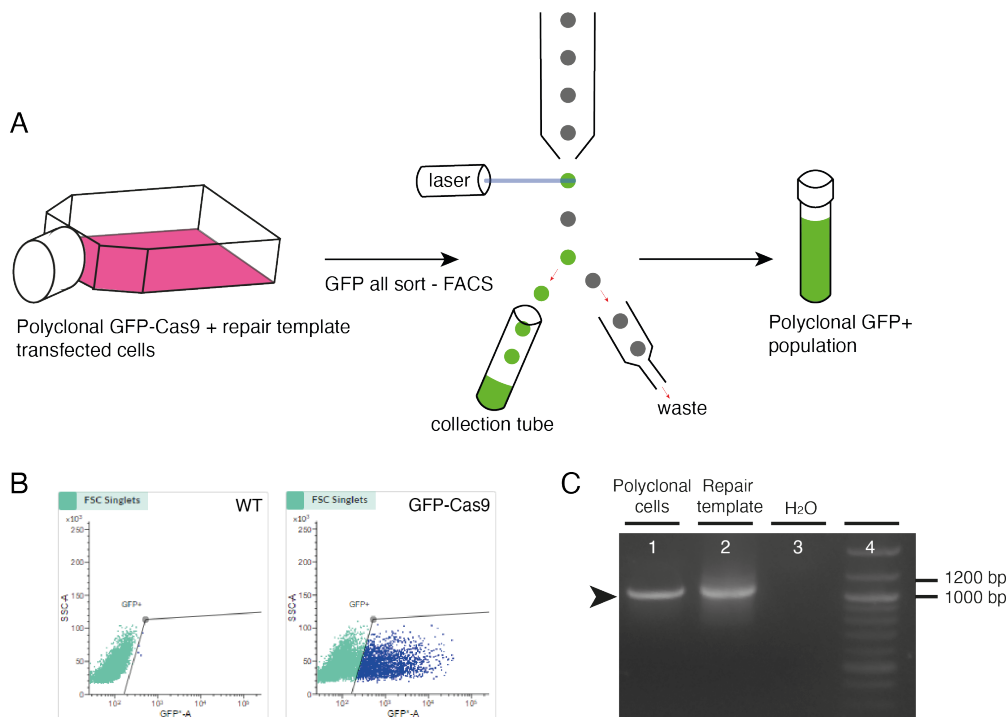
Figure 4.13: Generating a polyclonal tagged CBP cell population using FACS (A) Fluorescence Activated Cell Sorting (FACS). Jurkat cells co-transfected with the CBP-FTHH-RT and pX458-CBP were passed through the sorter in a droplet stream with a single cell per droplet. The laser at 488 nm shined at each droplet and cells expressing GFP fluoresced and were sorted into a collection tube. Cells without GFP expression were sorted into the waste. (B) Scatter plot of WT cells on left panel, gated to show no GFP expression. Right panel showing the collected population of GFP positive cells. (C) PCR using primers to amplify directly within the tag of edited cells and CBP-FTHH-RT plasmid as a positive control. n=1.

165

4.4.3 Generating clonal lines

The polyclonal population of tagged CBP cells allowed us to identify that the CRISPR editing was working as designed and that the tag had been successfully inserted onto the C-terminus of CBP. However, a clonal population of cells with homozygous tagged CBP was required for CRAC and TA-CRAC to obtain sufficient tagged material to carry out the purifications, generated through single cell sorting and screening.

4.4.3.1 Single cell sorting

To aid in our single cell sorting, we added a fluorescent Halotag ligand TMR (Promega), to bind covalently to Halo-tagged CBP, to only label cells with the desired tag. By single cell sorting based upon fluorescence for the dye, we could obtain single cells that have successfully tagged CBP. This would reduce the number of cells to be screened after clonal expansion (Figure 4.14A).

Firstly, polyclonal cells containing CBP with a FTHH tag: denoted CBP_{FTHH}, were analysed at 488 nm to check that there was no background GFP expression. WT cells were used for gating, to ensure cells without any fluorescence were not being sorted.

To control for background fluorescence, TMR was added to WT cells. Prior to sorting the dye was washed out of the cells, however during the sorting process a substantial population of cells appeared to fluoresce at 562 nm. This suggested that there were high levels of background TMR fluorescence, where the cells had taken up the TMR but it had not bound to any tag, as there was no Halotagged protein present in these WT cells.

This meant we proceeded with caution when observing the CBP_{FTHH} + TMR population. There was a small shift in the CBP_{FTHH} + TMR population compared to the WT + TMR population, so it was difficult to gate between background and correctly labelled CBP_{FTHH}. To ensure there were enough cells to sort and there was a population healthy and large enough to screen, the gates used were not as strict as they could have been. This increased the likelihood that fewer cells would be positive for CBP-FTHH than expected from the experimental design (Figure 4.14B). CBP_{FTHH} + TMR cells were single cell sorted into four 96 well plates for expansion and screening.
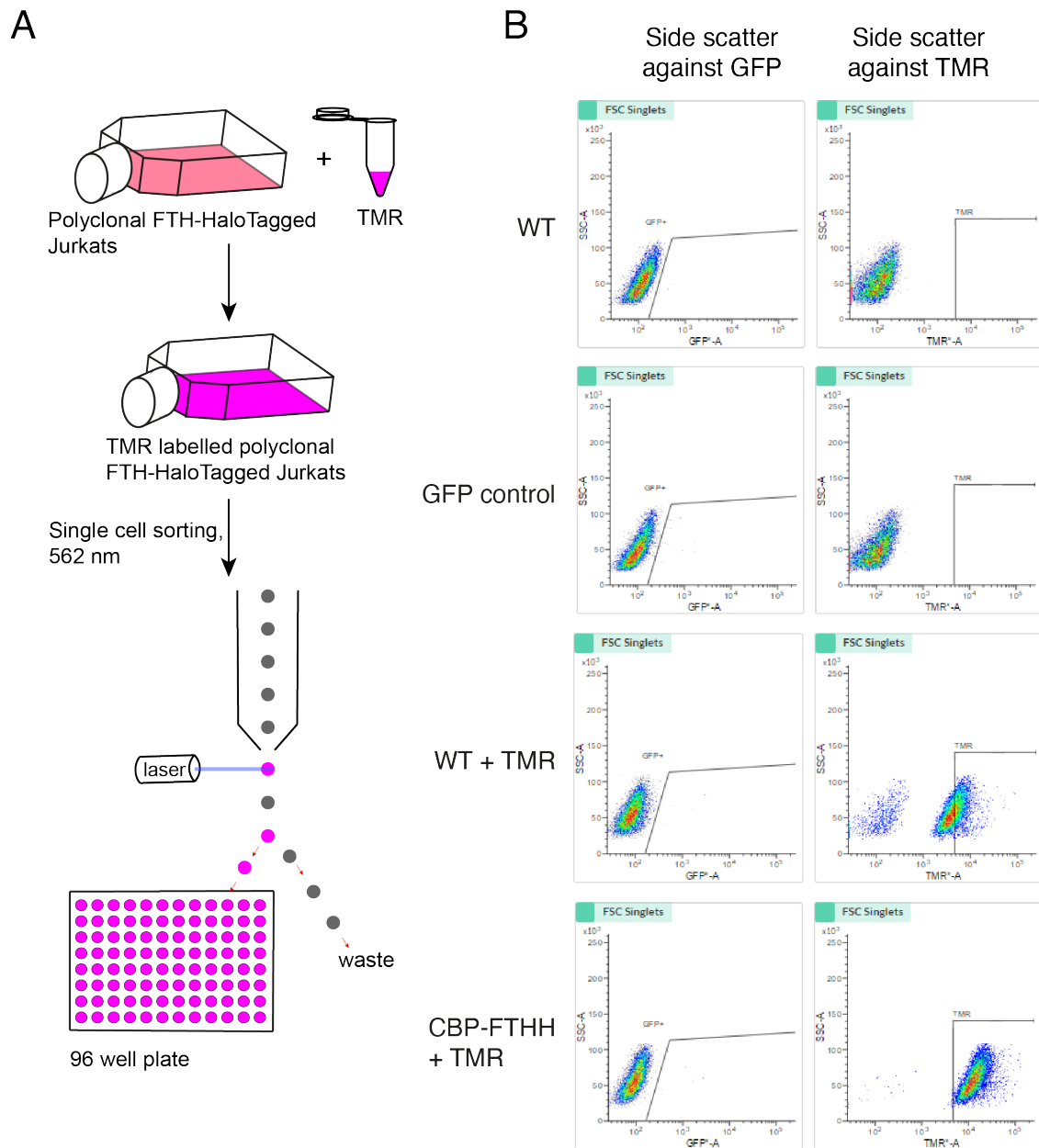
Figure 4.14: Single cell sorting based upon TMR labelling

(A) Single cell sorting into 96 well plate based upon TMR fluorescence at 562 nm. (B) FACS data for CBP_{FTHH} TMR cells. Left hand plots show side scatter against GFP for Cas9 signal, and right hand plots show side scatter against TMR signal. Upper panel shows WT cells, and second panel shows CBP_{FTHH} cells with no TMR added. Third panel shows WT cells labelled with TMR to control for background labelling and lower panel shows CBP_{FTHH} cells labelled with TMR. n=2.

4.4.3.2 Screening of clonal lines

After single cell sorting, CBP_{FTHH} clones were grown up and expanded over two weeks until the populations were sufficient for screening. Of the possible 384 cells that were sorted, 227 clones survived for screening.
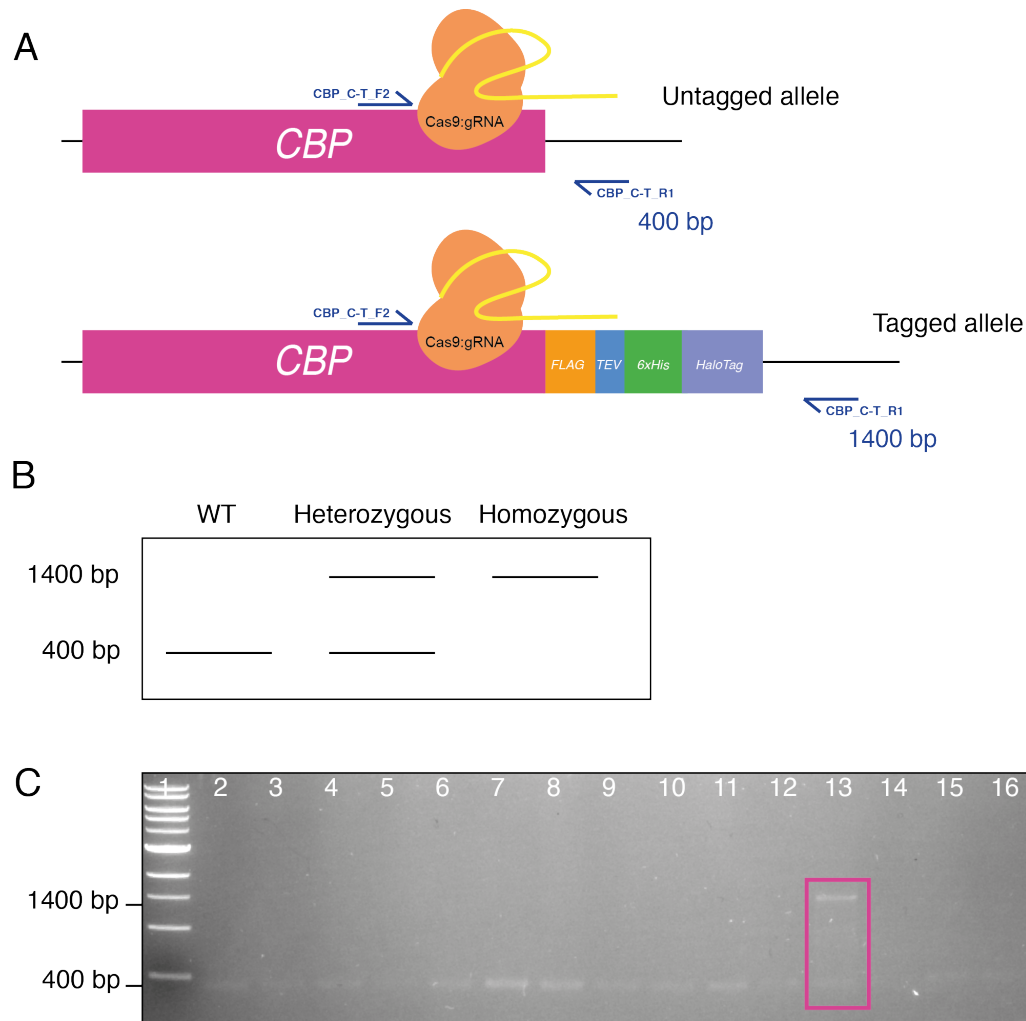
Figure 4.15: Screening of CBP tagged clones
(A) Tagged versus untagged CBP alleles. Purple arrows highlight the position of PCR primers (B) Expected band sizes for WT cells, heterozygous and homozygous clones. (C) Representative screening results. A double band at 400 bp and 1400 bp identified a single heterozygous tagged clone (clone CBP_{FTHH} B), indicated by pink box (Lane 13). n=1.

To screen colonies, PCR was carried out using extracted genomic DNA to amplify a region of DNA surrounding the DSB site (primers listed in Table 2.4, Figure 4.15A). We chose to use primers that amplified across the whole region rather than priming within the tag. This approach allowed us to confirm that our PCR reactions were successful, as every lane should contain a PCR product: a 400 bp band for WT alleles and a 1400 bp band for edited alleles. Otherwise there would be little way to distinguish between a failed PCR and a WT cell. Amplifying around the C-terminus also allowed us to distinguish between heterozygous and homozygous tagged CBP, as if the population was homozygous there would be a single band at 1400 bp, whereas if it was heterozygous it would have a WT and tagged allele and therefore two bands of 400 bp and 1400 bp (Figure 4.15B).

After screening the 227 clones, there were two clonal lines that contained bands at 1400 bp, but they also had bands at 400 bp, indicating they were heterozygous for tagged CBP. These were clones 16 and 43, henceforth denoted CBP$_{FTHH}$ clone A and B respectively, corresponding with the order in which they were screened. Lane 13 in Figure 4.15C, highlighted in pink, shows the screening of clone CBP$_{FTHH}$ B. This indicated we had successfully been able to generate clonal heterozygous lines for tagging of CBP.

To obtain further clones the same editing, single cell sorting and screening steps were repeated. No further tagged clones were identified, so we therefore proceeded to further screening and validation steps of these two heterozygous tagged CBP lines.

4.4.4 Limitations of a heterozygous cell line

Although we had initially aimed for homozygous tagged CBP, we were only able to generate heterozygous tagged CBP. Even with the additional enrichment steps put in to our workflow including GFP-Cas9 enrichment and TMR labelling, the efficiency of tagging this protein in this cell system was extremely low. CBP is a large protein at 300 kDa and it binds to a large number of proteins and at nearly all enhancer elements (Bedford et al. 2010). This means that CBP is critical for cell function and any disruption to the protein can have negative consequences to the cell, such as a reduction in cell viability.

Therefore we chose to proceed with heterozygous tagged CBP as we would have a

WT allele to compensate for any disruption the tagged allele may cause, however we would still have a significant proportion of the CBP species in the cell with our desired tags. This will still allow us to carry out planned CRAC and TA-CRAC experiments.

The next step in establishing these lines was to validate the tags were being expressed, as well as confirming that protein and cell function was not being disrupted due to the tag. Validation data is subsequently shown for CBP_{FTHH} clone B.

4.5   Validation of tagged CBP cell line

After identifying two heterozygous clones, we sought to confirm the edits were faithful in a number of ways. i) The DNA needed to be sequenced to confirm the tags had been correctly edited in and the sequence was not mutated. ii) Expression of the proteins also needed to be confirmed by RT-qPCR and Western blot. Due to the addition of the Halotag, labelling with a fluorescent dye and screening via FACS or imaging could also be used to confirm the tag was correctly knocked in.

4.5.1   Validation of DNA sequence

To validate the DNA sequence was as designed, we repeated DNA extractions and PCR using the same primer pairs as used in section 4.4.3.2. After confirming the PCR products were the correct size for a positively edited cell (Figure 4.16A), PCR products were extracted and sent for Sanger sequencing to confirm there were no mutations.

Analysis of the Sanger traces using available software proved difficult due to the size of the tag. For example, the ICE analysis used in validation of gRNA efficiency can also be used for HDR knock-in experiments, but is limited to a maximum 270 bp insertion sequence with flanking regions of >15 bp of homology required. Therefore it was only possible to analyse our CRISPR edit in without providing the software with a HDR template. When the tagged CBP trace was run against a WT control, positive insertions were identified, although it was not possible to determine that the 1005 bp edit had been introduced. 36% of cells contained a 4 bp insertion and there were 3% of cells that contained an insertion of 16 bp (Figure 4.16B).
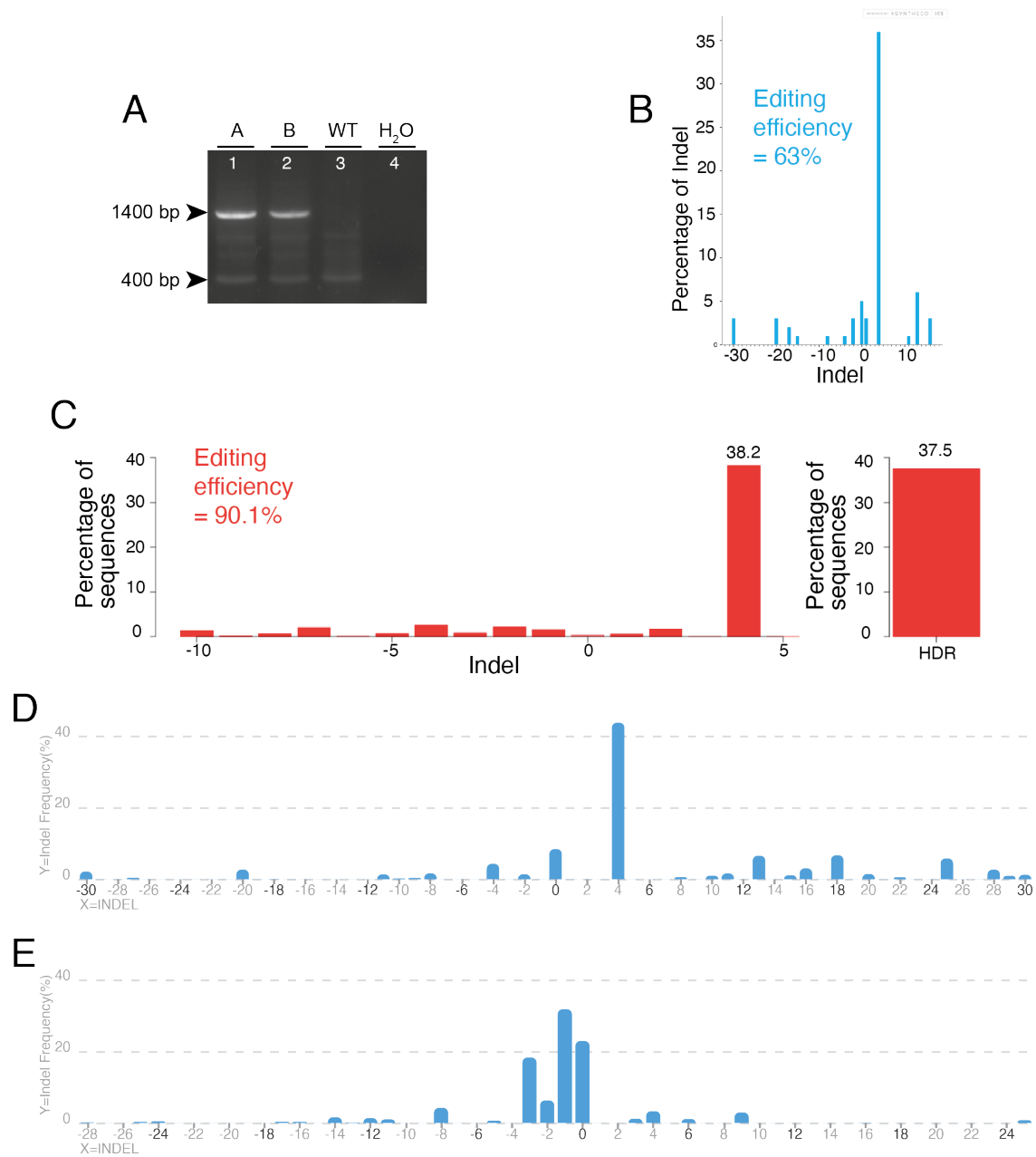
Figure 4.16: Investigation of CBP tag by Sanger sequencing and analysis
(A) PCR of heterozygous clones A and B, WT DNA and water control using higher
fidelity polymerase. (B) Distribution of indels visualised as a histogram in clone B
sanger trace compared to a WT only control (ICE, Synthego, Conant et al. 2022). (C)
Percentage of insertion by HDR (TIDER, Brinkman, Kousholt, et al. 2018). Distri-
bution of indels in clone B sanger trace compared to a (D) WT only control and (E) a
positive control HDR template ("SeqScreener Gene Edit Confirmation" app from the
ThermoFisher cloud). n=1.

As the ICE software was limited in the size of knock-in that could be analysed, we chose to look at other options available. An alternate software named TIDER (Brinkman, Kousholt, et al. 2018), a development of the Tracking of Indels by DEcomposition (TIDE, Brinkman, T. Chen, et al. 2014) software developed specifically for HDR experiments was used, however this can only measure a maximum size of 5 base insertions. This software requires a clean trace from a homozygous tagged clone as a positive control. We used a Sanger sequencing trace from the HDR template plasmid, as our positive control trace. When the clone trace was analysed against the WT control and the positive control, it identified 37.5% of cells as having HDR occurring and the edit being copied in, but only up to a maximum of 5 bp (Figure 4.16C).

A third software, the "SeqScreener Gene Edit Confirmation" app from the ThermoFisher cloud was also used and this was able to analyse up to 30 bases of insertion. Firstly the Sanger traces of the edited CBP clone were analysed against the WT trace, as carried out in the ICE analysis (Figure 4.16D). This identified that >40% of cells contained insertions of 4 bp and there were significant percentages of cells that contained up to 30 bp of insertions. Additionally, the tagged CBP clone was analysed against the positive control trace generated for the previous TIDER analysis. This identified that >20% of the CBP tagged cells did not contain any indels as compared to the positive control sequence, meaning that they had identical sequences (Figure 4.16E).

Each of these softwares did indicate that some proportion of HDR had taken place and the correct tag had been inserted.

Despite each software having limitations, it is useful for us to have analysed our tagged CBP in this way as it provides more evidence that within our cell line there are a significant portion of cells that do contain the tag. As our tagged cell lines were heterozygous, it was possible that the heterogeneity within the PCR sequence was adding to the difficulty of sequencing trace analysis, meaning it was more difficult to measure the true percentage of actual HDR. Additionally there may be biases towards one particular allele in the PCR or the sequencing. Especially given that the tagged allele is over 1000 bp, the PCR extension time chosen had to favour either the tagged or untagged allele and this may affect the proportion of tagged vs untagged alleles in the final PCR product.

As the tag was 1000 bp, PCR amplification of the C-terminus followed by visualisation on an agarose gel allowed distinct separation of the tagged and untagged allele. Gel extraction followed by Sanger sequencing of the tagged allele allowed improved analysis of Sanger traces. This removed heterogeneity between the tagged and untagged allele and allowed us to visualise if the tagged allele was homogenous and whether any mutations were present that had not been previously identified (Figure 4.17A).
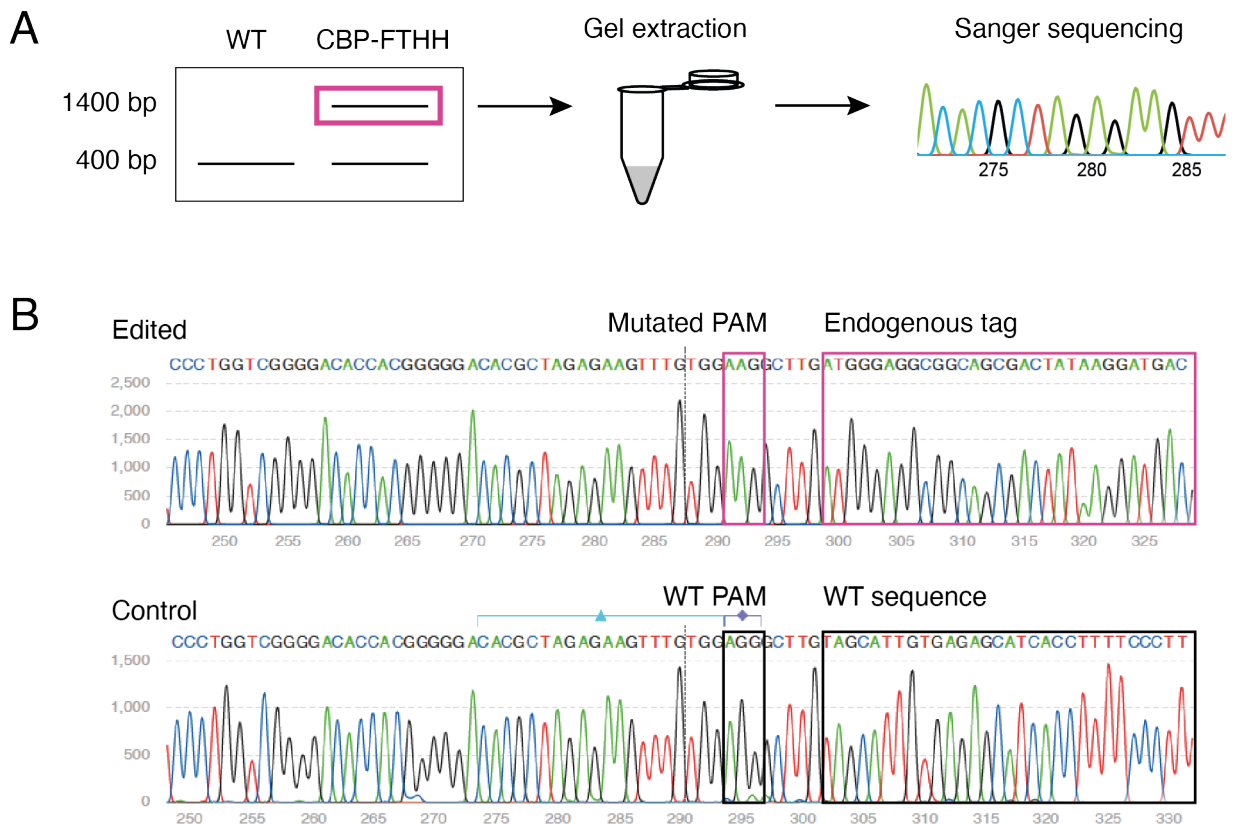
Figure 4.17: Sanger sequencing of tagged CBP allele

(A) Schematic showing separation of tagged vs untagged alleles amplified by PCR then separated by gel electrophoresis. The tagged allele of a band size of 1400 bp was isolated from the gel and extracted. The DNA was then sent for Sanger sequencing. (B) Comparison of tagged vs WT allele Sanger sequencing traces. Upper trace shows the tagged allele, demonstrating the mutated PAM site and then the tag has been successfully inserted (pink boxes). Lower trace shows the WT allele, with WT PAM sequence and 3' UTR shown in black boxes. n=1.

173

Analysis of the Sanger sequencing traces of the gel extracted tagged allele was done using the "SeqScreener Gene Edit Confirmation" app from the ThermoFisher cloud. Comparison of the edited vs unedited traces firstly showed that the tagged allele had the PAM site mutated correctly, preventing sequential cleavage and repair (Figure 4.17). It also showed the tag sequence appeared to have been correctly inserted, without any mistakes, and indicated that the cell line is truly clonal as the Sanger trace is clean and there is no heterogeneity visible in the sequence.

4.5.1.1 Off-target editing

A concern with CRISPR/Cas9 mediated genome editing is the unknown off-target effects that can occur within the genome. There is the possibility that the designed gRNA can still guide Cas9 to a site with decreased homology and induce cleavage at an undesirable locus. We began to screen the loci within the genome with the highest amount of homology to the CBP C-terminal gRNA. This would allow us to Sanger sequence the regions that had the highest chances of off-target editing, to determine whether any mutations were present. As the majority of our loci with high homology to this gRNA were found within regions of highly repetitive sequence, such as on the X chromosome, it was very difficult to obtain a distinct PCR product that could be sent for Sanger sequencing. Therefore, due to time constraints we were not able to confirm that no off-target editing had taken place.

4.5.2 Validation of protein expression

After confirming correct insertion of the tag sequence, the next step was to check that the tagged protein was correctly expressed using RT-qPCR and western blotting.

4.5.2.1 RT-qPCR

First we used Real Time quantitative PCR (RT-qPCR) to check for protein expression at an RNA level. A forward primer was designed to amplify directly within the Flag-TEV-6xHis tag and the reverse at the C-terminus of the Halotag to examine expression levels (Figure 4.18A).

The Flag-TEV-6xHis-Halotag was clearly expressed in the CBP tagged cell line,

but not in WT cells (Figure 4.18B). Examining CBP expression levels confirmed that addition of the tag was not reducing CBP mRNA expression.
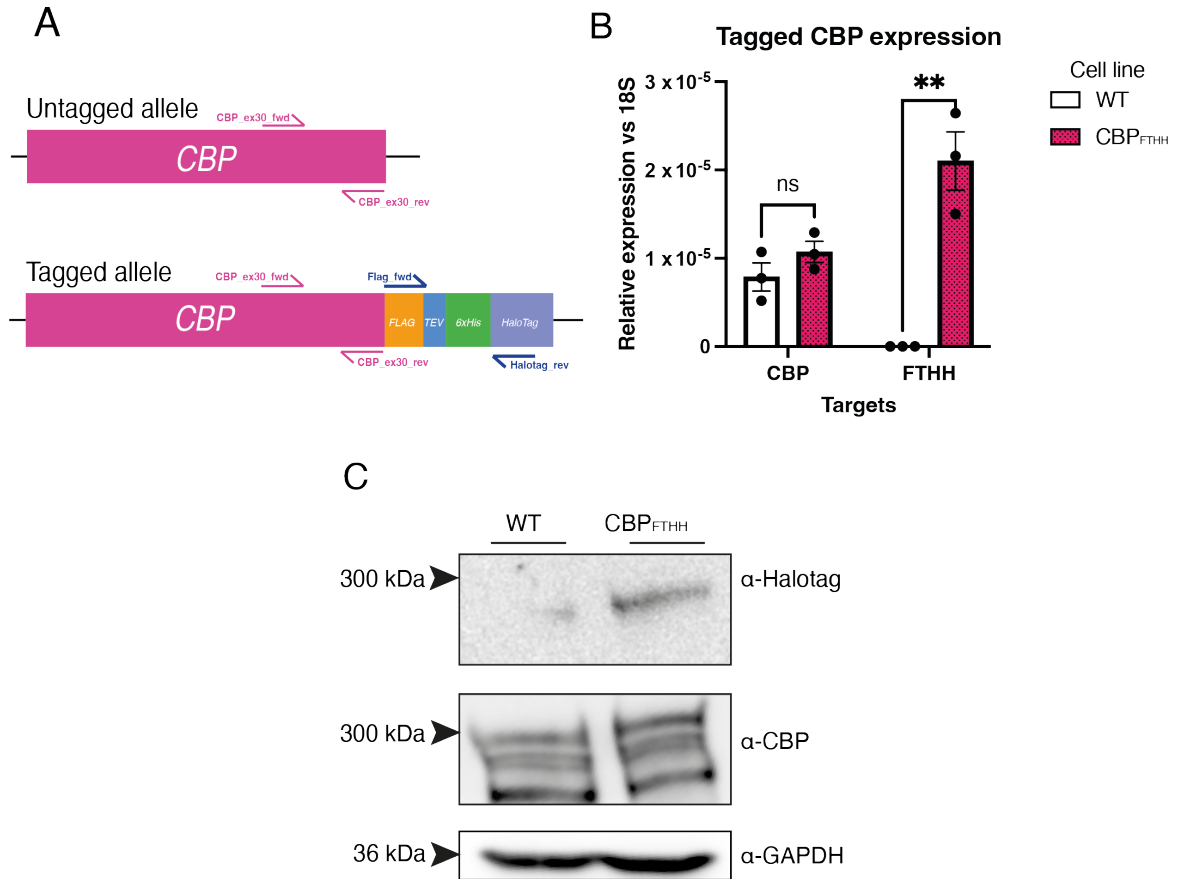
Figure 4.18: CBP is expressing a Halotag

(A) Primer binding within tagged and untagged CBP for primers targeting CBP (pink) and primers targeting the FTHH tag (blue). (B) RT-qPCR showing FTHH tag expression in the CBP_{FTHH} tagged cell line, expression is normalised to *18S*. Primers correspond with those shown in (A). Error bars denote SEM, n=3. Unpaired t test was used: ns = no significance, * = p<0.05, ** = p<0.01. (C) Western blot showing CBP has a Halotag in tagged Jurkat cells, n=1.

4.5.2.2 Western blot

To confirm expression of full-length, tagged protein we carried out a western blot against the incorporated Halotag. Clear bands were visible at the expected molecular weight for full length CBP in the tagged cell lines, but not the WT control, indicating

the correct expression of the full length tagged protein (Figure 4.18C).

Taken together, the Western blotting and RT-qPCR and data added evidence that the FTHH tag was being correctly expressed in these cell lines.

4.5.3 Validation of protein activity

After confirmation that the affinity tags on CBP were correctly edited into the cell, it was also very important to check that the affinity tags were not impeding CBP function. The Flag and 6xHis tags along with the protease cleavage sites adds 4 kDa, bringing the molecular weight of only the tags in total to 37 kDa with the Halotag. This is a large addition to CBP and it is possible these tags could interfere with CBP protein function. CBP is a transcriptional coactivator that activates many important transcription factors to increase transcriptional activity (Bedford et al. 2010). Should protein-protein interactions be disrupted this could abrogate the transcriptional coactivator role of CBP. Additionally, to function at enhancers and act as a histone acetyltransferase, CBP must be able to be recruited to chromatin and be able to acetylate histones and other non-histone substrates (A J Bannister and Kouzarides 1996; Holmqvist and Mannervik 2013). Therefore it is important to show that our affinity tags are not affecting protein function for the CRAC and TA-CRAC experiments.

4.5.3.1 Protein-protein interaction

To confirm that the endogenous tags knocked-in to CBP were not affecting protein-protein interactions within the CBP-TAL1 complex, a Co-IP was carried out using the Flag tag (Figure 4.19). A Flag Co-IP rather than a CBP Co-IP was done to isolate the tagged population of CBP within the cell only and reflects the approach we will take in our TA-CRAC experiments.

The Flag Co-IP implied that when we pulled down Flag, MYB and TAL1 were co-immunoprecipitated, suggesting that CBP-TAL1 complex formation was still retained (Figure 4.19). This further suggests that addition of the FTHH tag to CBP did not impede formation of protein-protein interactions between CBP and MYB and TAL1.
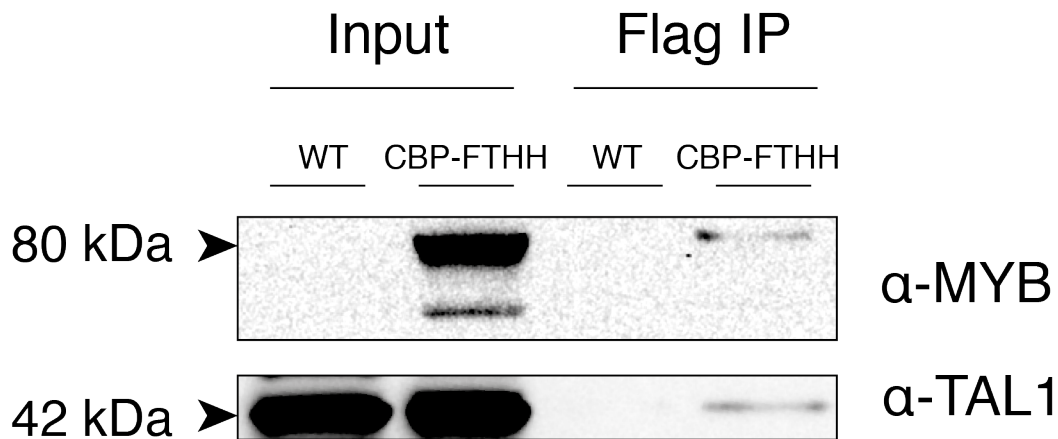
Figure 4.19: Tagged CBP retains protein and chromatin binding activity Immunoprecipitation using anti-Flag magnetic beads in WT and CBP-FTHH Jurkats shows Flag pulls down MYB and TAL1 in CBP-FTHH cells only. n=1.

4.5.3.2 ChIP

The chromatin binding activity of CBP is key to its role as a coactivator and being able to bind at a large number of enhancer sequences (Bedford et al. 2010). Moreover, if chromatin binding is disrupted it is possible that our tagged proteins are no longer binding at known binding sites and in particular the enhancer sequences of interest in our project. We were unable to complete ChIP analysis followed by either high throughput sequencing or qPCR, however this would be a key experiment to confirm tagged protein function. For example, if tagged CBP is no longer binding at the *TAL1* enhancer locus, then it is unlikely that CBP will interact with eRNAs transcribed from this locus and thus could affect the downstream CRAC and TA-CRAC experiments.

4.5.3.3 CBP catalytic activity

It was also important to check that the Flag-TEV-6xHis-Halotag did not disrupt the catalytic activity of CBP. We were unable to do this due to time constraints, however activity could be determined *in vitro* using immunoprecipitated tagged CBP in a fluorescent HAT assay (Abcam) or via western blotting using a nucleosomal substrate (D. A. Bose et al. 2017). Additionally activity could be checked for using ChIP-seq or ChIP-qPCR for H3K27ac.

4.6 Discussion

4.6.1 Establishment of the tagged CBP line

In order to carry out CRAC on CBP, a cell line containing endogenously tagged protein was required. In this chapter we outlined the steps taken to design, generate and validate a cell line containing affinity tagged CBP.

4.6.2 Choice of CRISPR/Cas strategies

There was a wealth of choice available in the strategy employed to tag our proteins. Over the last ten years since the first identification of programmable Cas enzymes for use in genome editing, the field of genome editing has grown significantly, alongside the tools and technologies available (Jinek et al. 2012).

4.6.2.1 Choice of Cas9 enzyme

Jinek et al. first identified the Cas9 enzyme from *S. pyogenes* (spCas9) as a tool for genome editing and there now exists a wide variety of Cas enzymes that have been validated for use in CRISPR experiments. This includes enzymes originating from varying bacterial species or enzymes that have been engineered in different ways to increase flexibility or efficiency when targeting PAM sites. Enzymes from certain species can change and often minimise the requirements for the PAM, increasing experimental design flexibility. For example, Cas9 originating from *S. Canis* recognises an NNG PAM sequence (Chatterjee, Jakimo, and Jacobson 2018; Chatterjee, Jakimo, J. Lee, et al. 2020). As a result this greatly increases the number of PAM sites available, offering more choice to the targets that can be used. In other cases, various mutations have been introduced into spCas9 that altered the specificity of the enzyme. For example a variant named spCas9-HF1 requires a much higher level of specificity for the crRNA to recognise the corresponding DNA sequence, which therefore reduces the number of off-target binding and errors (Kleinstiver, Pattanayak, et al. 2016).

Although various enzymes are available, we chose to use spCas9 without any further modifications as there was growing evidence that although increasing the flexibility of PAM sites allowed access to an increased number of target sites, the efficiency of

the cleavage at these sites can be reduced (Kleinstiver, Prew, et al. 2015; Kleinstiver, Pattanayak, et al. 2016). We were able to design high scoring gRNAs at our target sites within the NGG PAM constraints of spCas9 and therefore we did not need to utilise Cas enzymes with increased PAM flexibility (section 4.3.1). Additionally sp-Cas9 is the most widely used Cas9 for traditional genome editing and therefore the large number of online tools available for design and validation are mainly used with parameters for spCas9 editing. Also this lended itself to providing flexibility when it came time to delivering the editing machinery into the cell.

4.6.2.2   Form of gRNA, Cas9 and HDR templates can vary

There are different choices available for the form of Cas9, gRNA and repair template used in the cell. Cas9 and gRNA encoding plasmids are a frequently used strategy as it is easy to generate large amounts of plasmids, and therefore reduces cost of the editing experiment. It also makes it cheaper to try various gRNAs as multiple guides can be cloned into the Cas9 plasmids with ease. Recombinant Cas9 protein can also be used when complexed to the chosen gRNA and although this increases the cost of the experiment it has been shown to improve editing efficiency. This was behind our reasoning to begin with a Cas9/gRNA RNP complex for our first round of genome editing (Okamoto et al. 2019).

The use of Cas9/RNP complex had been shown to be at its most efficient when combined with an ssODN repair template. As we initially sought to only introduce a Flag-TEV-6xHis tag, this was an insertion of 110 bp. As it was just over the recommended limit of 100 bp, we decided to proceed with using an ssODN form. It had been reported that ssODNs increased editing efficiency in Jurkat cells, particularly as they are difficult to transfect and manipulate (Okamoto et al. 2019; Borowicz et al. 2020).

There is possibility that once the Cas9 has cut at the designed DSB site and the tag has been introduced by HDR, that the Cas9 and gRNA could recognise the same site a further time and cut once again, inducing errors and mutations. We sought to prevent this by introducing a silent mutation into the PAM sequence in the repair template, meaning that once the Cas9 had generated a double strand break and the edit had been inserted correctly by HDR mechanisms in the cell, the PAM site no

longer had an NGG sequence and therefore Cas9 could not cleave it again. This would prevent multiple rounds of cleavage and would decrease the chance of errors induced at this site (Okamoto et al. 2019).

With our method of using an RNP we were successfully able to validate that our gRNA was able to cleave at the designated cut site meaning this was an effective method of Cas9 cleavage in our cell line (section 4.3.2). However, we were unable to produce any evidence that editing of our insert was occurring using the ssODN as a repair template (section 4.3.4). Previous studies using ssODNs in Jurkat cells had found 1% of cells were edited using these methods, so a low editing efficiency was expected (Borowicz et al. 2020). However, at this point if any succesful editing was occurring in our system it was at a percentage that was so low it was unidentifiable. Therefore methods to enrich this population for editing in the cell were required.

4.6.3   A turn to plasmid based methods

To provide a larger number of options for enriching for editing, we explored plasmid based methods. Plasmids can be used in both repair templates and Cas9 delivery and can contain many different markers and additional tools that can be used singularly or in combination, to aid in visualisation of successful editing.

4.6.3.1   A plasmid encoded Cas9 and gRNA

Antibiotic resistance markers were available on Cas9 plasmids, meaning that after transfection of Cas9 through antibiotic selection a single population that were expressing Cas9 could be generated. However, previously in the lab we had found that Jurkats did not tolerate antibiotic selection well. After selection the proportion of cells that did survive grew slowly and it therefore made downstream experiments more difficult (Celadova 2022).

Therefore we chose to not use antibiotic based selection and turned to fluorescent marker based methods. We chose a GFP-tagged Cas9 and to utilise Fluorescence Activated Cell Sorting (FACS) to positively sort all cells expressing GFP, therefore generating a polyclonal population where all cells should be expressing Cas9.

The same validation steps used for validating the gRNA in RNP form (section 4.3.2) were repeated. Despite the gRNA sequence being the same, it was important

to repeat these validations to check that the plasmid was correctly encoding the Cas9 and it functioned as designed (section 4.4.1).

After carrying out the FACS a high proportion of cells survived, showing that the Jurkat cells tolerated the sorting well (section 4.4.2). This could be due to the fact that Jurkat cells grow in suspension, meaning that they were more tolerant to passing through the stream and being sorted and this did not induce as much stress to the population as it might do to strictly adherant cells. This made an effective method of enrichment and the polyclonal population generated appeared healthy meaning that there were no issues with population size for downstream experiments.

4.6.3.2   A plasmid encoded HDR template

With rethinking our strategy for Cas9 delivery we also explored if there were any further changes we could make to our HDR template. As we had been unable to visualise any successful HDR editing with the ssODN and were turning to a plasmid based Cas9, we also examined a plasmid based repair template. The homology arms were extended to 600 bp as this would be tolerated in the plasmid based system. We hoped this would increase efficency as it would give larger regions of homology.

As the size of the homology arms were being increased, it would be possible to add in additional tags to the repair template that could aid in selection. We turned to the Halotag as it offered a wide range of flexibility: there are a wide range of ligands commercially available that upon addition to the cell covalently bind to the Halotag. One way this tag can be used is during pull-down experiments, as it can be used to isolate the tagged protein of interest out of the cell lysate to aid in protein analysis experiments. Additionally, proteolysis targeting chimeras (PROTACs) developed specifically to target the Halotag have also become a useful tool, meaning that in further downstream experiments protein degradation could be explored (Buckley et al. 2015). The main use we saw in the generation of our clonal lines was the introduction of fluorescent ligands. If tagged CBP could be fluorescently labelled this would enable single cell sorting based on fluorescence, meaning that all of the generated clones should be clonal for the tag.

Therefore we chose to proceed with a plasmid based repair template, with 600 bp homology arms to insert a Flag-TEV-6xHis-Halotag onto CBP, but with a 3C protease

site between the 6xHis and the Halotag so that the Halotag could be cleaved away
and would not interfere with the CRAC and other downstream experiments.

4.6.3.3 Using the new editing strategy using plasmid based CRISPR/Cas9 reagents

After carrying out the CRISPR/Cas9 based tagging using the new strategy of plasmid
based reagents and FACS, the resulting polyclonal population was tested using PCR
and successful HDR editing was visualised. This was done by using PCR primers to
directly amplify within the tag sequence and provided the first evidence that successful
HDR editing was occurring as designed. We therefore proceeded to single cell sorting
to generate clonal lines (section 4.4.2).

To generate the clonal lines a fluorescent Halotag ligand called TMR was added
to the cells, and single cell sorting was carried out based upon TMR signal. After
TMR addition there was a significant shift in the population, so it appeared that
the tagged cells were being labelled correctly. However, WT cells with added TMR
were also tested to control for background TMR staining and there was also a large
shift in the fluorescent population visible. This means the level of cells that had
background staining were high. In hindsight the gates should have been made harsher
to discriminate further between background and true labelling, however at the time
we chose to proceed with the single cell sorting. Additionally the concentration of
TMR was high which likely lead to increased levels of background staining (section
4.4.3.1).

4.6.4 Validating tagged protein function

4.6.4.1 Clonal screening

Clones were grown up into sufficient sized populations that they could be screened
for the edit. As four 96 well plates were used in the single cell sorting process and a
high proportion of cells survived after sorting, this gave over 250 clones that required
screening for the CRISPR edit. There are many methods that could be used to ex-
amine whether the clones contain the tags, for example looking at protein expression
using western blotting or RT-qPCR. However, as such a high number of clones were

required to be screened a high throughput method was needed. Therefore DNA extraction and PCR analysis was chosen, as this can be done more feasibly on a larger scale.

After carrying out the screening of the clones, two heterozygous clones were identified (section 4.4.3.2). The editing efficiency of CBP was a lot lower than expected from our experimental design, despite utilising additional steps to enrich for cells containing tagged CBP. CRISPR/Cas9 mediated endogenous tagging of CBP's paralogue p300 with a GFP tag in mouse embryonic stem cells successfully yielded both two heterozygous and two homozygous cell lines (Ma et al. 2021). However, CRISPR/Cas9 endogenous tagging of CBP, particularly with larger tags such as the Halotag has not yet been described previously in the literature. Moreover, upon generation of a workflow to create base pair substitution mutations into Jurkat cells using CRISPR/Cas9, these small scale edits were only present in 2% of the clones screened (Borowicz et al. 2020). Taken together, CRISPR/Cas9 genome editing within Jurkat cells is extremely inefficient, coupled with the additional challenges of editing a large tag into a large protein such as CBP_{FTHH}, the efficiency for generating such a homozygous line is even lower. Therefore generation of our heterozygous line was no small feat and we proceeded with our lines.

4.6.4.2   Off-target editing

Examining off-target editing through PCR and sequencing at genomic loci with high homology to the designed gRNA recognition sequences is useful as it is a straightforward method that allows visualisation if any off-target editing is taking place. However, this method is limited as it does not look genome wide. There are a number of methods utilising different next generation sequencing techniques to examine off-target editing. For example Digenome-seq involves extraction of DNA or chromatin from the cell, and then to this cell-free DNA recombinant Cas9 and gRNA are added. Any successful Cas9 cleavage, be it on or off-target will result in an identical 5' or 3' end due to the nature of the Cas9 PAM requirement. Adaptors can then be ligated and these fragemnts can be sent for whole genome sequencing. Alignment to the reference genome would show where Cas9 is cleaving across the genome and would easily visualise if any off-target editing is occurring (D. Kim, Kang, and J.-S. Kim 2021).

However, Digenome-seq was noted to give very high levels of background, from this CIRCLE-seq was developed. In the CIRCLE-seq technique, genomic DNA is extracted from cells and sheared. The sheared fragments are then circularised and Cas9-gRNA RNP are added. If there is any cleavage activity by Cas9, be it on or off-target, it will break the DNA circle and allow adaptor ligation to occur and library preparation can be proceeded with. Therefore when sent for sequencing, this would identify any regions that contain regions of DNA where Cas9 is actively cleaving (Tsai et al. 2017).

4.6.5 Outlook for RNA binding experiments in our tagged line

Our aim in this chapter was to generate a cell line expressing endogenously tagged CBP, in order to characterise the RNA binding profile of CBP in the context of the CBP-TAL1 complex, through use of CRAC and TA-CRAC. Although ideally we would use a homozygous tagged CBP cell line to proceed with CRAC and generating further lines for TA-CRAC, it seemed we would be unable to generate this homozygous line. We were satisfied that the heterozygous tagged allele did retain function and it therefore would be good enough to use in our CRAC and TA-CRAC experiments.

In this chapter we have generated a cell line where CBP has been endogenously tagged using CRISPR/Cas9 and HDR. We have shown the steps taken in order to design and validate gRNA activity and also that the edit was being repaired into the cell using HDR. Strategies to achieve this editing have been adapted and changed to achieve the knock-in and we have generated a heterozygous tagged CBP line. Tagged CBP expression has been validated in terms of Western blot and RT-qPCR, showing that the tags are being expressed in the cell and are usable. Additionally, the tags are not impeding protein-protein interactions that are important in CBP-TAL1 complex formation, as shown in our Co-IP experiments.

Therefore it appeared we have established a tagged CBP line that can now be used in CRAC experiments. This cell line can now be taken and developed for TA-CRAC by tagging MYB.

Chapter 5

Generating a stable tandem tagged MYB cell line to investigate RNA binding

5.1 Introduction

In the former chapter we have begun to establish and validate methods to carry out endogenous tagging of CBP with affinity tags in Jurkat cells using CRISPR/Cas9 and homology directed repair (HDR). This involved generation of a tagged cell line containing CBP-Flag-TEV-6xHis-Halotag (CBP$_{FTHH}$), followed by validation steps to confirm tag expression. Although we did not test histone acetyltransferase (HAT) activity of CBP remained functional, our data suggested that the tags were not affecting protein-protein interactions within the CBP-TAL1 complex, and tagged CBP could bind at the *TAL1* enhancer.

5.1.1 TA-CRAC to investigate CBP-MYB interactions

The established CBP$_{FTHH}$ cell line could be used to carry out Crosslinking and Analysis of cDNAs (CRAC) (Figure 4.1B) to examine CBP-RNA binding interactions. However, we sought to develop this to investigate CBP-RNA binding when CBP is found within a specific regulatory protein complex. We planned to further the CBP$_{FTHH}$ cell line by tagging a second protein of interest in tandem, to carry out

Tandem Affinity CRAC (TA-CRAC, as described in section 4.1.2).

TA-CRAC will allow the examination of how RNA binding alters with changes in protein-protein interactions. We planned to investigate CBP and a protein binding partner that was found within a known protein complex: in our case CBP-TAL1. We wanted to examine how changing the protein composition of this regulatory protein complex would alter the RNA binding capacity of both CBP and the transcription factor binding partner, and determine if there was any implication on the species of RNA that were binding. As described in section 4.1.2.3, we chose to investigate RNA interactions between MYB and CBP, due to the role MYB plays in initiating enhancer formation (Sanda et al. 2012; Mansour, Abraham, et al. 2014, Figure 4.1).

5.1.2 Tagging MYB with affinity tags

To carry out TA-CRAC, MYB will require tandem affinity tags: 6xHis-TEV-HA/Strep (HTHAS) (Figure 4.1A). These tags will be introduced within the tagged CBP cell line, as TA-CRAC requires endogenous tags on both proteins to identify the RNA species that are binding only to MYB when it is found in complex with CBP.

As described in section 4.1.2.3, CBP-MYB TA-CRAC will produce an RNA binding map for MYB, but only for the MYB species that are found in complex to CBP (Figure 4.1C). This will provide greater understanding into the action of CBP within regulatory protein complexes, if any changes to RNA binding do occur as a result of CBP being in complex. Additionally, with a focus on eRNAs it may provide new insight as to the binding behaviour of eRNAs within regulatory protein complexes that function at enhancers.

This chapter will outline the tagging of MYB with a 6xHis-TEV-HA/Strep tag (MYB-HTHAS) within the CBP_{FTHH} cell line; a cell line we term MYB_{HTHAS}:CBP_{FTHH} to highlight that both MYB and CBP are tagged within this single cell line. We had observed such success with tagging CBP that we chose to follow the same methodologies used for generation of the CBP_{FTHH} cell line. After establishing the MYB_{HTHAS}:CBP_{FTHH} cell line, the same steps of validation were required to confirm tagged protein function and behaviour, prior to carrying out TA-CRAC.

5.2 Experimental design for tagging MYB

5.2.1 Planning the tagging of MYB

The planning of the CRISPR/Cas9 and HDR tagging of MYB was made easier by the fact we had been able to successfully tag CBP in our system and knew that our methodologies worked well. Into clone B of our CBP-Flag-TEV-6xHis-Halotag cell line we planned to insert 6xHis-TEV-HA/Strep tag, hereafter referred to as the HTHAS tag, onto the C-terminus of MYB (Figure 5.1A). We planned to base our strategy on that used in the tagging of CBP, with some small adjustments (Figure 5.1B).
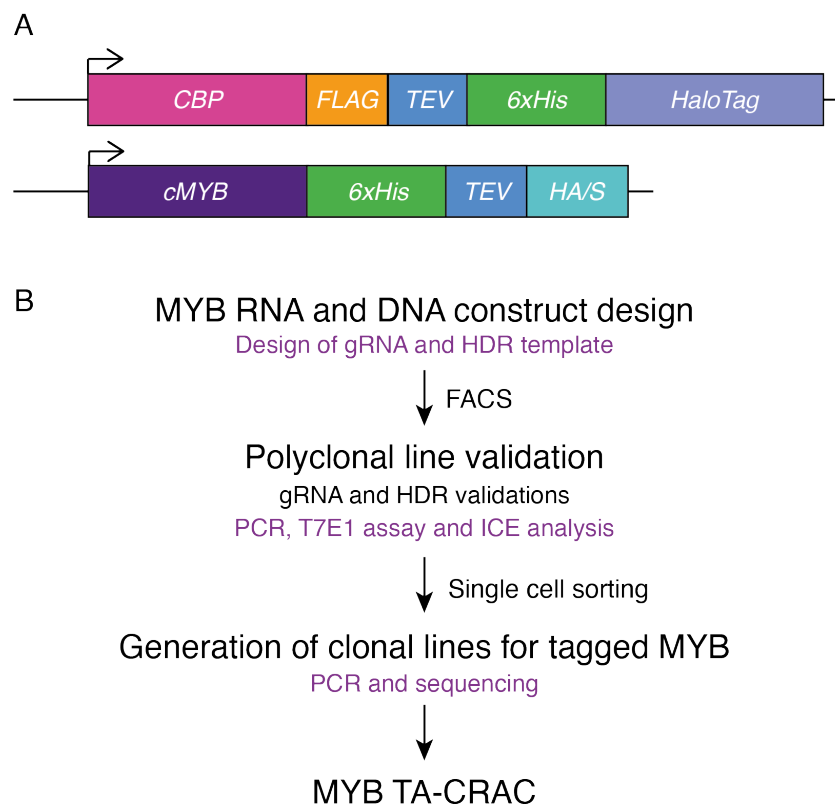
Figure 5.1: MYB tagging experimental plan
(A) Tags to be edited onto MYB in the tagged CBP line. (B) Experimental workflow of steps to achieve tagged MYB line.

For the first validation step we planned to transiently co-transfect both the Cas9-gRNA and repair template into the CBP_{FTHH} cells. If we used a GFP-Cas9 with a

self cleaving peptide, as we used previously in the tagging of CBP (Section 4.4.1), we could use Fluorescence Activated Cell Sorting (FACS) and then validate the gRNA and any potential editing using the T7E1 assay and ICE analysis as our first validation step. We had previously shown that using FACS to capture the GFP positive and therefore the Cas9 positive population was an effective way to enrich for edited cells, as low editing efficiency has been observed previously within Jurkat cells (Okamoto et al. 2019). This enrichment step should make the initial validations easier: as if the gRNA and the HDR template are working as designed, a higher proportion of the population should have editing successfully occurring.

Therefore when we validate the gRNA, using PCR we would be able to screen the polyclonal population for edits at the same time.

After validating the gRNA and any potential edits that are occurring within the cell, we could then proceed straight to single cell sorting, to generate and screen our clonal lines for tagged MYB$_{HTHAS}$.

5.2.2 MYB gRNA design

We firstly designed our gRNA to target the C-terminus of MYB, with chromosomal coordinates listed in Appendix 8.1. When previously tagging CBP, we used a plasmid encoding both the gRNA and a Cas9-GFP, separated by a self-cleaving T2A peptide. Through fluorescent activated cell sorting (FACS) to sort all GFP positive cells and therefore all cells expressing Cas9, we had observed a significant enrichment for cells edited with our designed tags onto CBP within our polyclonal population (Section 4.4.2). Therefore we chose to repeat this method when tagging MYB and used the GFP-Cas9 plasmid to encode our MYB gRNA (Figure 5.2A and B).
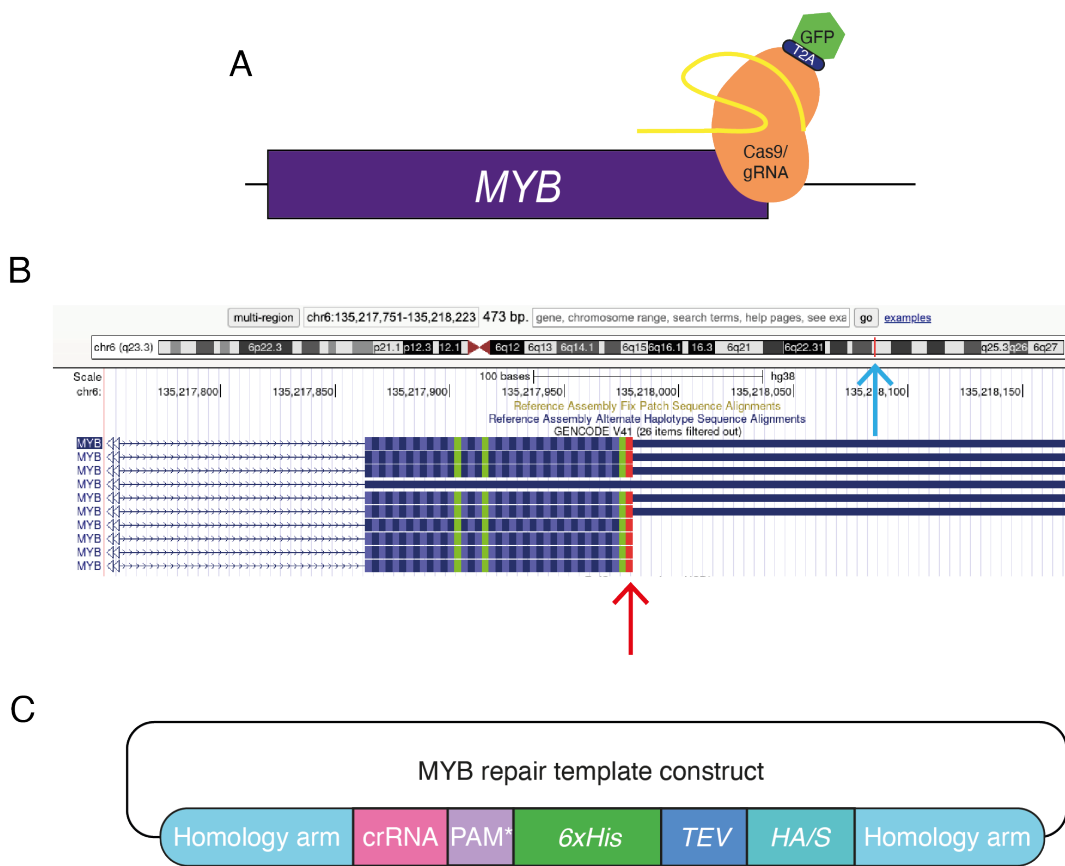
Figure 5.2: MYB gRNA and HDR template design

(A) MYB gRNA targets the C-terminus using a GFP-Cas9. (B) UCSC genome browser view to show the genomic locus of chr6:135,217,751. Light blue arrow indicates the location of the locus on the chromosome, with the red arrow indicating the specific locus at the C-terminus. Co-ordinates listed in Appendix 8.1 (C) Design of MYB HDR template.

5.2.3 MYB HDR template design

The HTHAS tag is a significantly smaller tag than the FTHH tag eventually used for CBP tagging, at 177 bp compared to 1005 bp (Section 4.4.2). To design an efficient repair template for such a large knock-in, the FTHH tag needed to be encoded via a plasmid, with homology arms of 600 bp. Although our HTHAS tag was a fraction of the size of that used previously, ssODN HDR templates were recommended for <100 bp tags and we therefore inferred that a strategy utilising an ssODN for the HTHAS

tag may not result in efficient knock-in of this tag (Okamoto et al. 2019).

Additionally, as we had established methods for achieving successful CRISPR/Cas9 editing using plasmid based HDR templates, we chose to proceed with using a plasmid as a HDR template for the HTHAS tag. This plasmid was also designed with 600 bp homology arms as we had observed this could induce knock-in edits successfully within Jurkat cells (Figure 5.2C).

5.3 Validating the generation of a polyclonal MYB-HTHAS line

As with tagging CBP, the gRNA designed to target MYB required validation to ensure that the gRNA sequence was correctly guiding the Cas9 to cleave at the correct site using the same methods as used in previous validations for tagging of CBP (Section 4.3.2). This included the T7E1 assay and analysis of Sanger sequencing traces from PCR products using ICE.

CBP$_{FTHH}$ Jurkat cells were co-transfected with the MYB gRNA-GFP-Cas9 and HTHAS-HDR template plasmids. 48 hours post-transfection cells were sorted using FACS to collect all of the cells that were GFP positive, enriching the population for edited cells (Figure 5.3A). This meant that subsequent gRNA validation was done on the enriched polyclonal population. DNA was then extracted and PCR was carried out to amplify the C-terminus of MYB (Figure 5.3B and C).

5.3.1 T7E1 assay

The T7E1 assay was carried out on the enriched, polyclonal population to show that Cas9 was actively cleaving in the cell. Figure 5.3B shows the expected PCR fragments for the T7E1 assay if Cas9 had edited the DNA at the desired site, with and without the designed tag. If Cas9 had cleaved at the MYB C-terminus, but erroneous mutations had been introduced due to non-homologous end joining (NHEJ) repair, this would cause the 650 bp PCR product to be cleaved into a 150 bp and a 500 bp fragment. The addition of the HTHAS tag onto MYB would increase the total PCR fragment size to 820 bp. T7E1 activity would then cause fragments of 500 bp and 320 bp.
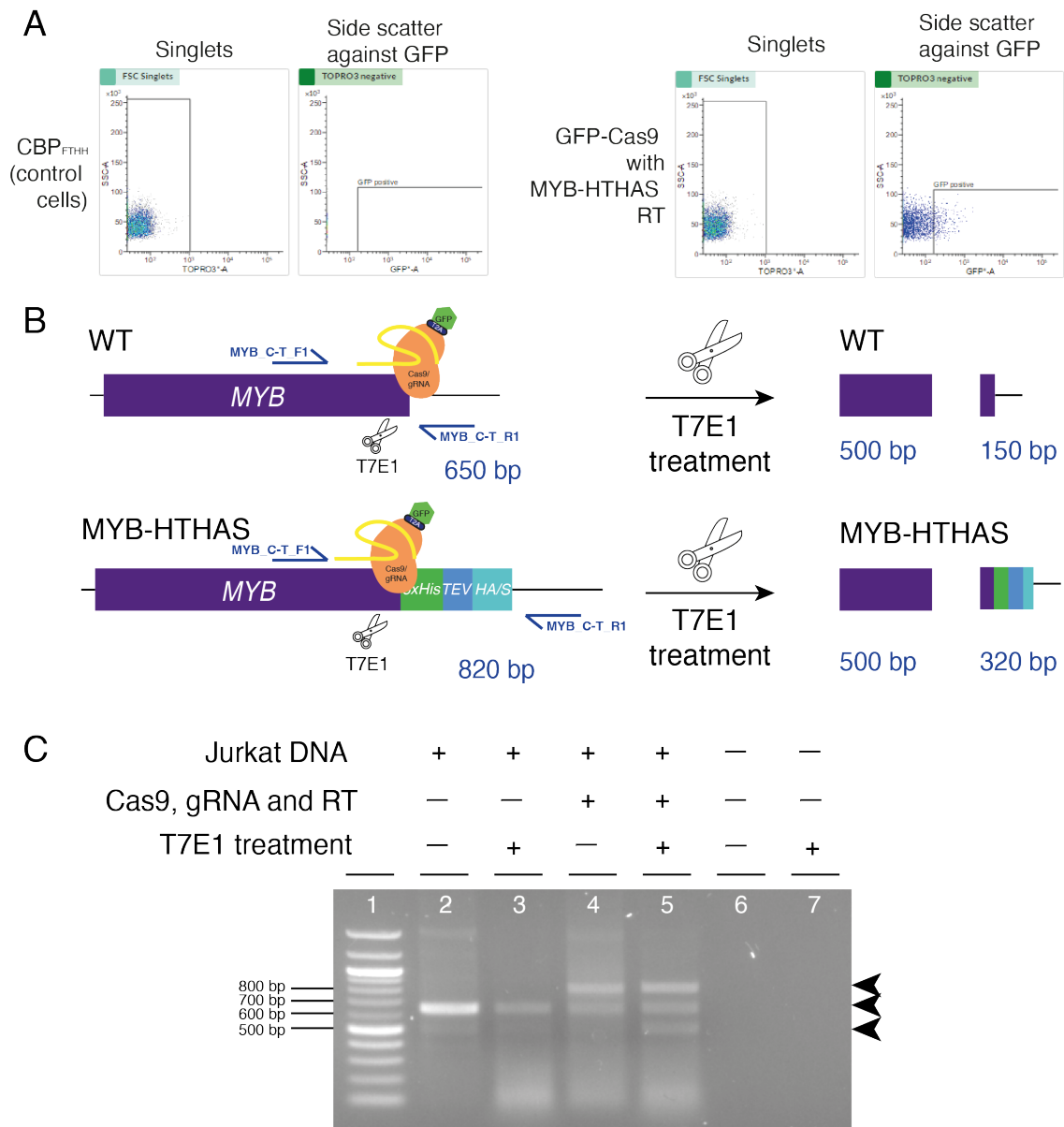
Figure 5.3: Validation of MYB gRNA and repair
(A) FACS data for generating tagged MYB polyclonal line. Left two panels show singlets in left panel and GFP signal in right panel for unedited cells. Right two panels show edited cells with GFP-Cas9 with the singlets in the left panel and the GFP signal in the right panel. (B) T7E1 assay with PCR primers (purple arrows) amplifying region of DNA cut by gRNA at the MYB C-terminus. Expected sizes of the T7E1 assay shown. (C) T7E1 assay for MYB gRNA. n=1.

When examining the T7E1 assay, there was a band in lanes 2, 3, 4 and 5 of 650 bp, indicative of the PCR fragment for WT sequence (Figure 5.3C). Lane 5 showed PCR fragments from edited cells that had been treated with T7E1. The band visible at 500 bp implied that Cas9 had generated double strand breaks (DSBs) at this site and these had subsequently been repaired by the cell. The 650bp representing the WT sequence was also present.

Of note, there was a band of 820 bp in both lanes 4 and 5, showing this band was present independent of T7E1 activity. This band was of the expected height if the HTHAS tag had been incorporated into the C-terminus of MYB. This was due to the addition of the repair template when carrying out the gRNA validation experiments. As this band was so prominent at this stage, it may imply that there was a reasonable proportion of successfully edited cells in our polyclonal population. In the editing process of CBP we were unable to visualise bands at the height of the inserted tag when looking at this stage so this indicated that the tag had been inserted at a higher efficiency than our editing for CBP.

Taken together, the T7E1 assay had indicated that not only was Cas9 actively cleaving within the Jurkat cells, but also there was preliminary evidence for our designed edit being inserted correctly.

5.3.2   ICE analysis

The PCR products generated for the T7E1 assay were sent for Sanger sequencing and analysed using ICE (Conant et al. 2022). Figure 5.3A showed a comparison of the trace from WT cells against the trace from cells edited with both GFP-Cas9 and HDR plasmids. As with our validation for the CBP gRNA (Figure 4.6), there was a distinct difference between the edited and unedited cells, with the unedited trace being very clean and there only being one clear sequence. However, in the edited sample there were clearly multiple different sequences within the polyclonal pool, as there were numerous peaks of each base at every Sanger coordinate. Therefore, this implied that Cas9 was correctly cleaving at the designed cut site and that errors were being repaired into the cell at the designed double strand break (DSB) site.
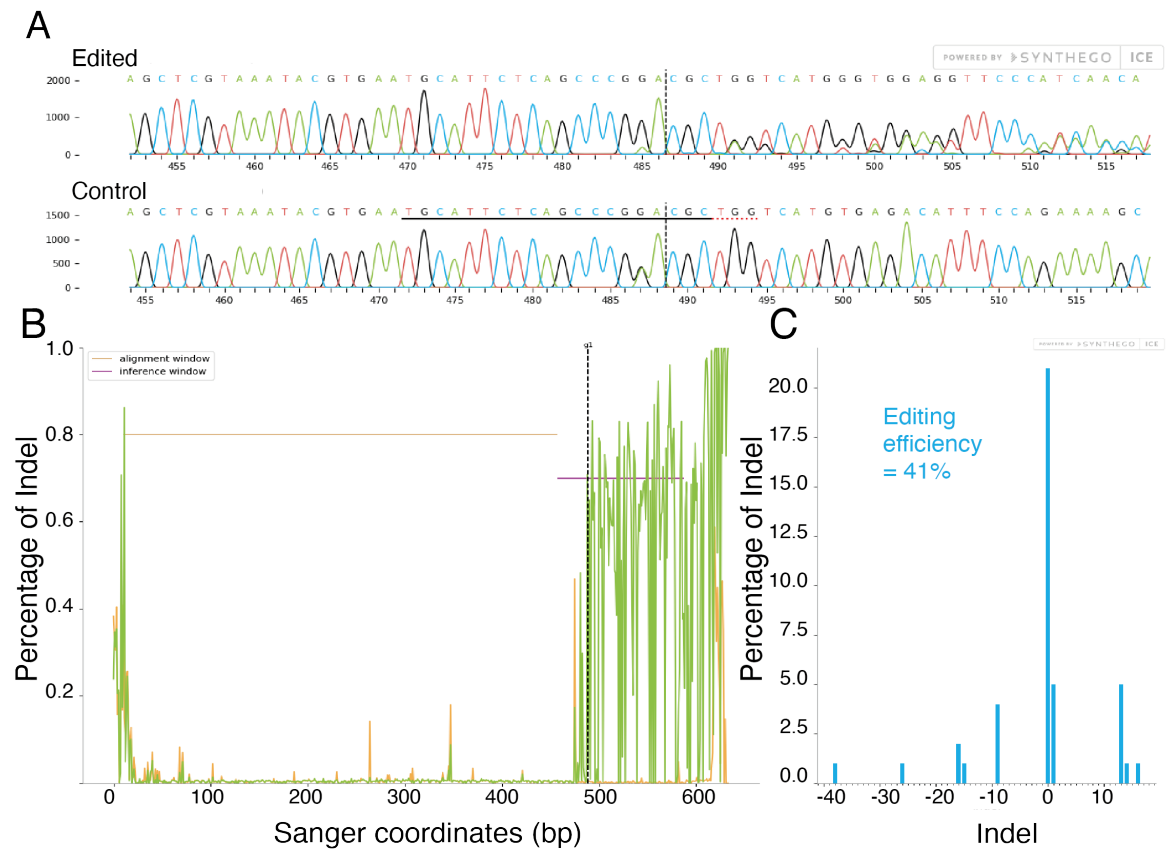
Figure 5.4: MYB gRNA Sanger sequencing analysis
ICE analysis of PCR products from Figure 5.3 showing Jurkat cells edited with both the MYB Cas9-gRNA plasmid and the MYB-HTHAS-RT plasmid (edited cells) (Conant et al. 2022) (A) ICE analysis of comparison of Sanger sequencing traces for Jurkat cells co-transfected with MYB tagging plasmids (upper) compared to WT (control) Jurkat cells (lower). (B) Divergence plot of how the sequences of DNA from edited and unedited cells vary. (C) Calculated distribution of indel size within the edited cell population. n=1.

Comparison of the discordance between the edited and unedited traces was high, showing that the new sequence within the edited cells was very different to that of the WT cells (Figure 5.3B). As the edited sequence differed greatly from the WT sequence this provided more evidence that editing was occurring at the designed site at the MYB C-terminus.

Interestingly, the ICE software was able to detect indels of large sizes present in

the polyclonal population. For example almost 1% of the edited population contained a deletion of 38 bp, and 5% contained an insertion of 13 bp. It was not possible to pick out the HTHAS tag, however it was clearly visible that a lot of CRISPR/Cas9 mediated editing was taking place within this population and in this case the ICE software identified there was an editing efficiency of 41% (Figure 5.3C).

Our ICE data provided evidence that Cas9 was cleaving at the designed DSB site at the MYB C-terminal region, and edits were being introduced at this site. Taken together with our T7E1 data this indicated that our designed gRNA worked efficiently to guide Cas9 to the target site for cleavage. Moreover, our T7E1 experiments implied that the HTHAS tag was present in the polyclonal population due to the band at 820 bp present in the edited population in lanes 4 and 5 (Figure 5.3C).

Therefore, we proceeded to single cell sorting of this polyclonal population to generate clonal lines. We hoped as there seemed some level of successfully edited cells within our polyclonal population that there would be a high enough proportion of correctly edited cells in the clones surviving single cell sorting. This would mean there would be fewer clones to screen to obtain a correctly tagged MYB_{HTHAS} cell line and raised hope of obtaining a homozygous tagged MYB line.

5.4 PCR genotyping of single cell sorted clones to search for tagged MYB

In our strategy for tagging MYB, the repair template used did not contain any tags that allowed for any method of selection, so the polyclonal population was single cell sorted into four 96 well plates. The clones were grown up and expanded into a suitably sized population for screening.

133 clones were DNA extracted and the C-terminal region was screened using PCR genotyping. It was expected that a WT clone would have one single band at 650 bp, a heterozygous clone would have two bands; one at 650 bp and a second at 820 bp. A homozygous clone would be expected to have only one band at 820 bp (Figure 5.5A). Within the first 42 clones screened, 7 heterozygous clones and 2 homozygous clones were identified (Figure 5.5B).

194
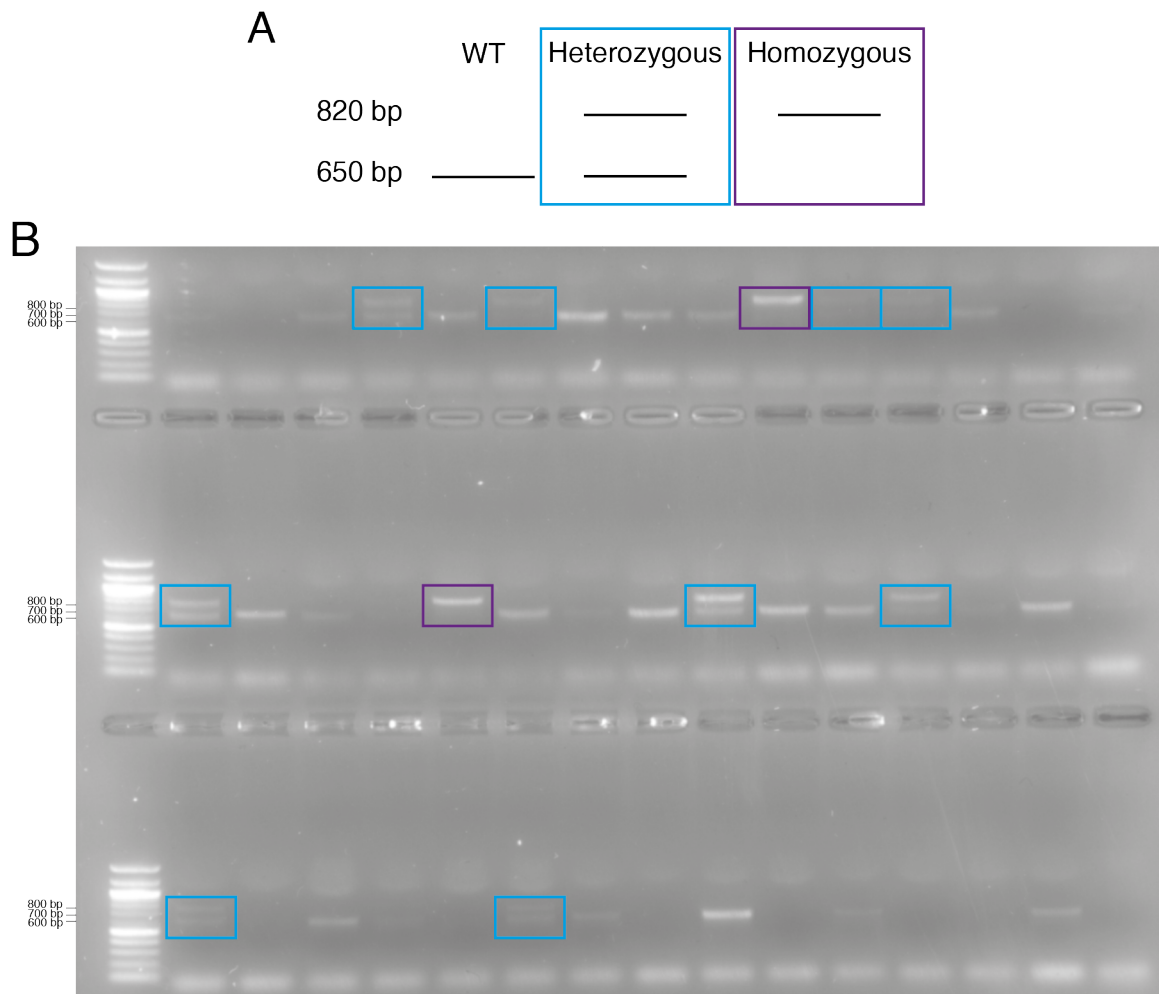
Figure 5.5: Generating clonal tagged MYB lines

(A) Expected sizes of PCR products when run on an agarose gel of the C-terminus of MYB, with WT DNA, heterozygous (blue box) or homozygous (purple box) tagged DNA. (B) Clones 1-42 were screened and heterozygous clones (blue boxes) and homozygous clones (purple boxes) were identified. (C) Homozygous identified clones were amplified using higher fidelity enzyme and ran on a 1% agarose gel. n=1.

From our clone screening it appeared that we had obtained two homozygously tagged MYB lines and we proceeded with these two lines for validation of tag function. These were clones 10 and 19, as shown in Figure 5.5, with clone 10 being denoted A and 19 being denoted B.

5.5 Validation of tagged MYB line

After identification of MYB_{HTHAS} tagged clones, these needed to be validated in the same ways as was done during the generation of the CBP_{FTHH} cell line (Section 4.5). It was important to check that the tag was being properly expressed and this was done using RT-qPCR and western blotting and protein-protein interactions needed to be examined using Co-IPs.

5.5.1 DNA sequence validation

DNA extractions were repeated for the homozygous clones and PCR was carried out using a higher fidelity polymerase to generate larger quantities of DNA. This was done using the same primers as used previously in initial validations for the T7E1 assay (Figure 5.3) and during clone screening (Figure 5.5). PCR amplification of WT sequence would produce a band of 650 bp, with the 170 bp HTHAS tag creating a band of 820 bp when the C-terminal region of MYB was amplified using the same primers.

The PCR products were first visualised on an agarose gel to confirm band size (Figure 5.6A). This showed both clones produced a single band at 820 bp, compared to a WT sample of 650 bp. Therefore, it implied we had obtained two homozygous lines.

The PCR products generated for Figure 5.6A were subsequently sent for Sanger sequencing and analysed to ensure there were no mutations or incorrect sequences that had been edited in (Figure 5.6B). The Sanger sequencing trace when visualised was perfectly clean, highlighting that the visible consensus sequence was homogenous, meaning that the same edit had been copied in precisely to both alleles. The sequence was the exact sequence for the HTHAS tag, providing strong evidence that the correct tag sequence had been correctly edited in. Additionally the silent substitution mutation introduced within the PAM sequence to prevent repetitive Cas9 cleavage was visible.

196
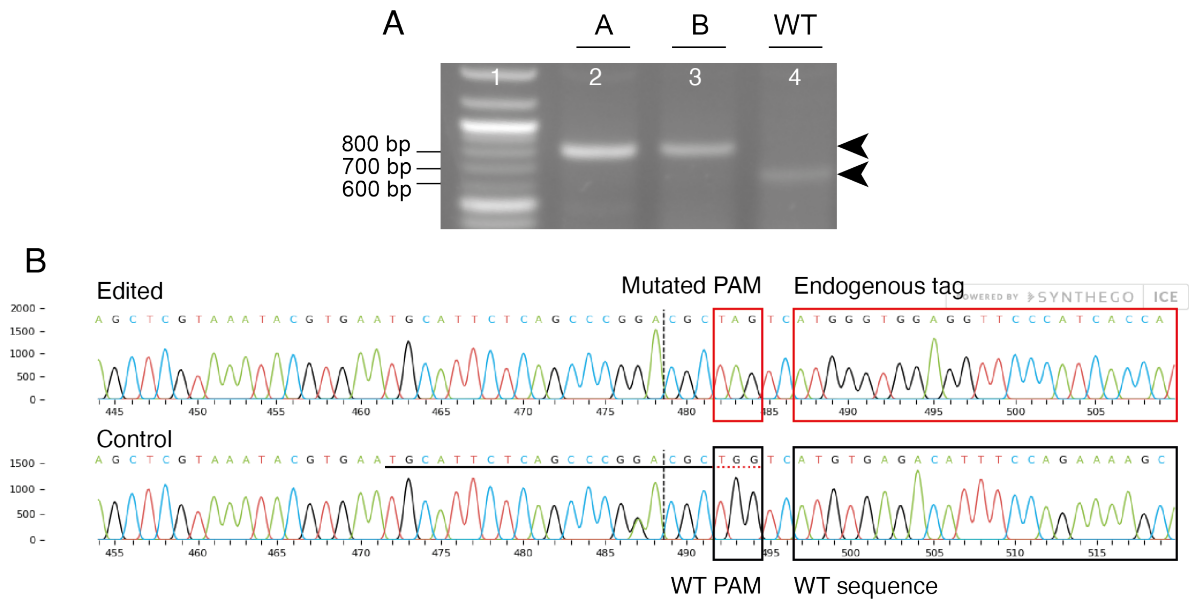
Figure 5.6: Tagged MYB DNA sequence validation
(A) PCR of previously identified homozygous tagged MYB clones to confirm homozygous status. (B) ICE analysis of WT DNA against tagged MYB in clone A, with mutated PAM and the inserted endogenous tag sequences highlighted. Clone B ICE analysis also carried out, data not shown. n=1 for each clone.

Therefore this indicated that the HTHAS tag had been inserted correctly into both MYB alleles and we had obtained two homozygous clones containing MYB-HTHAS within our CBP_{FTHH} cell line.

The final cell line generated is denoted from this point as MYB_{HTHAS}:CBP_{FTHH}, as it is important to emphasise that this cell line contains both tagged MYB and tagged CBP. In the interest of time, all further cell line validation was carried out on clone A.

5.5.2 Gene expression of the tagged clones

After showing that the inserted DNA sequence appeared to be as designed (Figure 5.6), it was important to show that the HTHAS tag was being expressed on an RNA level. The HTHAS tag was edited onto MYB within the CBP_{FTHH} tagged cell line so we additionally sought to confirm that the FTHH tag was still being expressed on CBP and that having both tags was not impeding activity at the *TAL1* enhancer, by

measuring expression of the *TAL1* target gene. As we had previously done RT-qPCR to confirm CBP tag expression (Figure 4.18), we chose to repeat this in our final MYB$_{HTHAS}$:CBP$_{FTHH}$ cell line as a confirmation of tag expression (Figure 5.7).

Clear expression of the HTHAS tag was confirmed in the final MYB$_{HTHAS}$:CBP$_{FTHH}$ cell line (Figure 5.7A). Expression normalised to *18S* showed that *MYB* levels were similar in WT and CBP$_{FTHH}$ cell lines, although *MYB* expression was reduced in MYB$_{HTHAS}$:CBP$_{FTHH}$ cell lines in comparison to WT (Figure 5.7A).

RT-qPCR of *CBP* and the FTHH tag was repeated to verify that gene expression levels of total *CBP* and tagged *CBP* in the final cell line remained similar to that of the CBP$_{FTHH}$ only line (Figure 5.7B). Expression levels for the FTHH tag were similar in both the CBP$_{FTHH}$ line and the MYB$_{HTHAS}$:CBP$_{FTHH}$ line, indicating that by tagging MYB the expression of tagged *CBP* was unaffected (Figure 5.7B). Furthermore, *CBP* expression showed similar expression patterns in WT, CBP$_{FTHH}$ and MYB$_{HTHAS}$:CBP$_{FTHH}$ cells (Figure 5.7B). This implied that the tagging of MYB was not affecting expression levels of *CBP*.

Gene expression of *TAL1* was also investigated using RT-qPCR (Figure 5.7C). Despite reduced *MYB* gene expression, there was no significant reduction in *TAL1* gene expression; the associated target gene of the *TAL1* enhancer. Additionally, there was no reduction in eRNA expression from the *TAL1* enhancer in the MYB$_{HTHAS}$:CBP$_{FTHH}$ cell line. This therefore suggested that although *MYB* expression was reduced in this cell line, in terms of RNA expression, there were no observable negative effects on *TAL1* enhancer activity (Figure 5.7C).

Together, our RT-qPCR data showed that MYB expression remained stable within the tagged line and suggested that by tagging MYB, this species was still active at the *TAL1* enhancer. Additionally CBP and tagged CBP expression was similar to that observed in the tagged CBP only line. This was the RT-qPCR pattern we expected for tagged MYB and tagged CBP that still retained function within the cell.

198
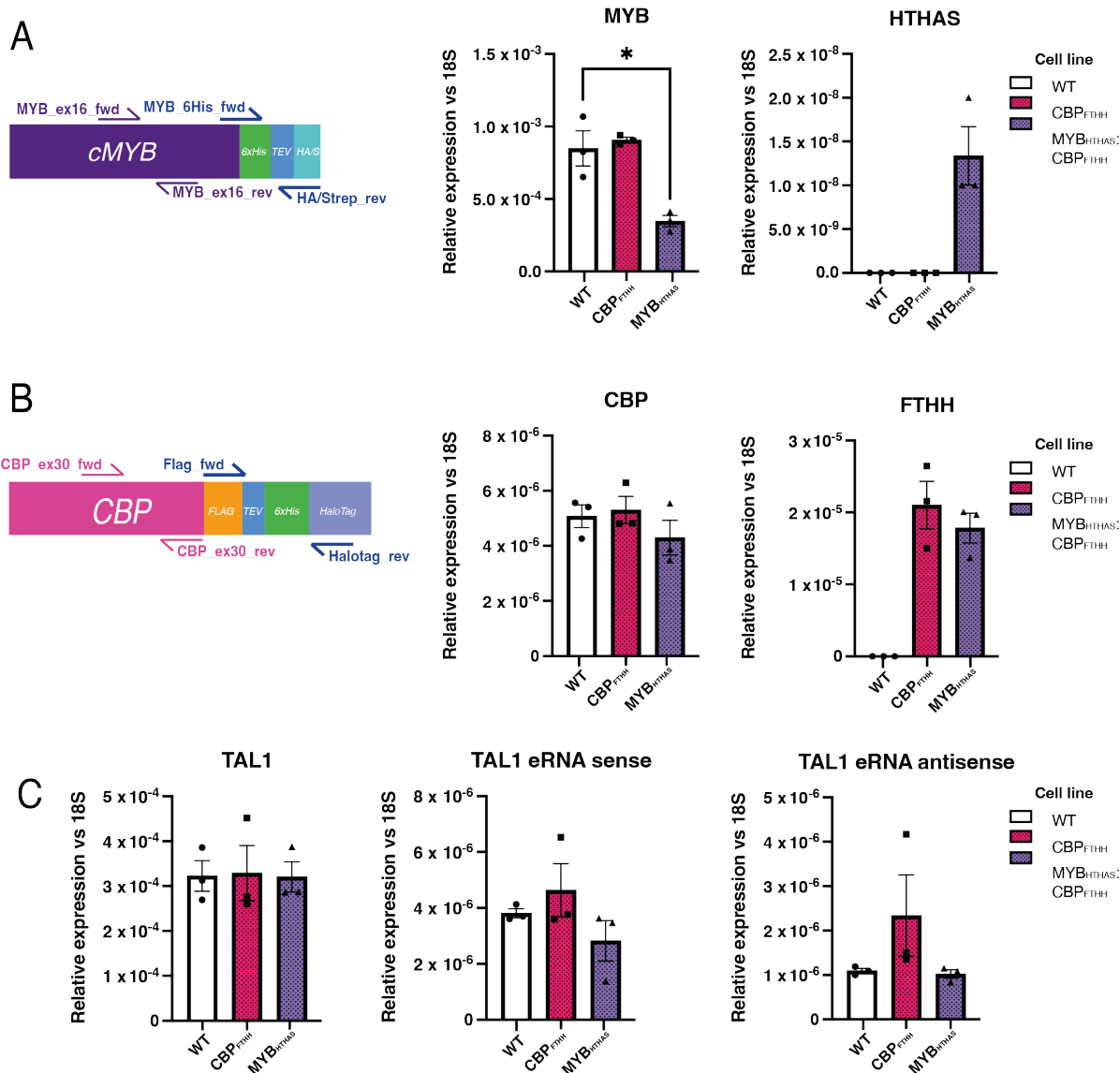
Figure 5.7: RT-qPCR confirmed expression of MYB-HTHAS tag
RT-qPCR of WT, CBP$_{FTHH}$ and MYB$_{HTHAS}$:CBP$_{FTHH}$ tagged lines. Gene expression normalised to *18S* expression. Error bars denote SEM, n=3. Data points from WT cells shown as circles, CBP$_{FTHH}$ cells as squares and MYB$_{HTHAS}$:CBP$_{FTHH}$ cells as triangles. (A) Expression of *MYB* and the HTHAS tag. (B) Expression of *CBP* and the FTHH tag (C) *TAL1* gene expression, *TAL1* enhancer eRNA sense and antisense expression. Across all, paired t-test was used: * = p<0.05. Where no significance is indicated, this represents no significance.

5.5.3 Western blotting

After analysing gene expression on an RNA level, we next turned to investigating protein expression. We carried out a western blot in our final MYB_{HTHAS}:CBP_{FTHH} cell line to examine expression of the MYB endogenous tags (Figure 5.8). For the western blot we used a control of WT Jurkat cells and we additionally used CBP_{FTHH} for comparison against the double tagged MYB_{HTHAS}:CBP_{FTHH} cell line.
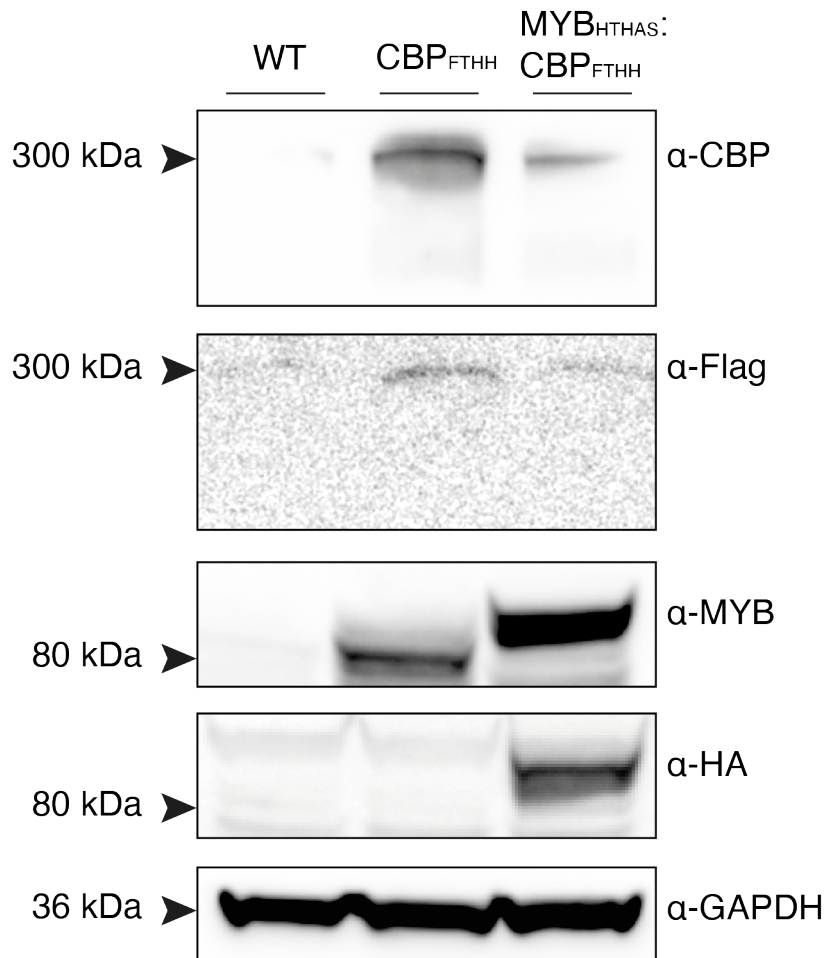
Figure 5.8: Tagged MYB is expressing a HTHAS tag
Western blot showing analysis of CBP_{FTHH} and MYB_{HTHAS}:CBP_{FTHH} tagged cell lines. n=1.

Full-length MYB appeared to be expressing a HA tag in our MYB_{HTHAS}:CBP_{FTHH} cell line using an antibody raised against the HA tag. A band at just higher than

80 kDa was visible, which was only visualised in the MYB_{HTHAS}:CBP_{FTHH} sample. This indicated this band was that of the HA tag on MYB.

Additionally by blotting against the MYB protein itself, there was a small shift in size between untagged MYB in the CBP_{FTHH} only line, compared to tagged MYB in the MYB_{HTHAS}:CBP_{FTHH} cell line. The HTHAS tag had a molecular weight of 5.5 kDa, which when added to the 80 kDa of full-length MYB generated an 85.5 kDa species. Therefore this small difference in size had likely caused the slight shift in band size that was observable in the anti-MYB blot. Additionally, there was no band of 80 kDa in the tagged MYB lane, all species were of the larger molecular weight running at 85.5 kDa, which provided further evidence for this cell line having both alleles tagged and therefore being homozygous.

The band for full-length MYB in the MYB_{HTHAS}:CBP_{FTHH} cell line appeared more intense than that of untagged MYB within the CBP_{FTHH} only cell line. This implied that the HTHAS tag may have had a stabilising role for the MYB protein, as presence of the tag caused more visible protein in our western blot. If this was the case, this could explain why in our RT-qPCR data, *MYB* gene expression was lower when tagged with the HTHAS tag, compared to in WT cells (Figure 5.7). Potentially, the HTHAS tag could have been stabilising MYB-HTHAS full-length protein, resulting in decreased degradation of MYB protein globally across the cell. Therefore to titrate MYB levels within the cell population, *MYB* gene expression may have been reduced.

To confirm that expression of the CBP-FTHH species was the same as in the CBP_{FTHH} only cell line, we also blotted for Flag and CBP. The Flag tag was observable upon CBP in both tagged CBP cell lines, at a height of 300 kDa, implying that full-length CBP was expressed with the additional protein tags. Full-length CBP was also visible at 300 kDa when blotting for the protein itself, confirming protein expression in the tagged lines, and further corroborating that the visualised 300 kDa species visualised when blotting for Flag was that of tagged CBP.

Unfortunately, despite the loading control of GAPDH confirming that the same amounts of protein lysate were loaded prior to running the protein gel, we were unable for this blot to visualise WT CBP and WT MYB expression in the WT Jurkat cells. Proteins containing large regions of intrinsic disorder, such as CBP, are prone to degradation and it is possible that in this case these proteins did degrade and were not running as full-length protein at the expected height. One possible explanation

was that the WT sample used was prepared at an earlier time and cell lysate was snap frozen and stored before use, compared to the tagged cell lines which used fresh cells. This method of snap-freezing can sometimes cause proteins that are prone to degradation to degrade further, and could be why in our control sample we could not visualise full-length CBP or MYB. However, in our previous RT-qPCR data, we were able to show that gene expression of both CBP and MYB was comparible in WT cells to both of our tagged cell lines, therefore we were not too concerned with the lack of full-length CBP and MYB in our WT samples. To draw firm conclusions, this would need to be repeated, ideally with a fresh sample of each cell line, however due to time constraints we were not able to repeat this.

In summary western blotting implied the edited tags were being expressed on both full-length CBP and MYB.

5.6 Validation of tagged MYB protein activity

Our validation experiments suggested that the MYB-HTHAS tag was being expressed in MYB_{HTHAS}:CBP_{FTHH} cells, the tagging of MYB did not appear to impede expression of *TAL1* more than in the parental CBP_{FTHH} cell line and had implied CBP-FTHH expression remained similar. The next step was to check that protein activity and behaviour of the tagged MYB species was not affected by the introduction of the tag. In the same way that we investigated that tagging of CBP with the FTHH tag did not appear to impede protein-protein interaction activities or chromatin binding activity (Section 4.5), we sought to confirm MYB-HTHAS was behaving in the same manner as MYB-WT. We also chose to carry out an RNA Immunoprecipitation (RIP) to confirm tagged MYB and tagged CBP in this cell line retained similar RNA binding activity to WT, indicating that we had produced cell lines that were ready to proceed with for carrying out TA-CRAC.

5.6.1 MYB Co-Immunoprecipitation

To examine whether tagging MYB with a HTHAS tag disrupted the ability of MYB to form protein-protein interactions, we carried out a Co-IP against the Strep tag introduced onto MYB to isolate the tagged species. We also carried this out in WT

cells, to control for non-specific binding (Figure 5.9).
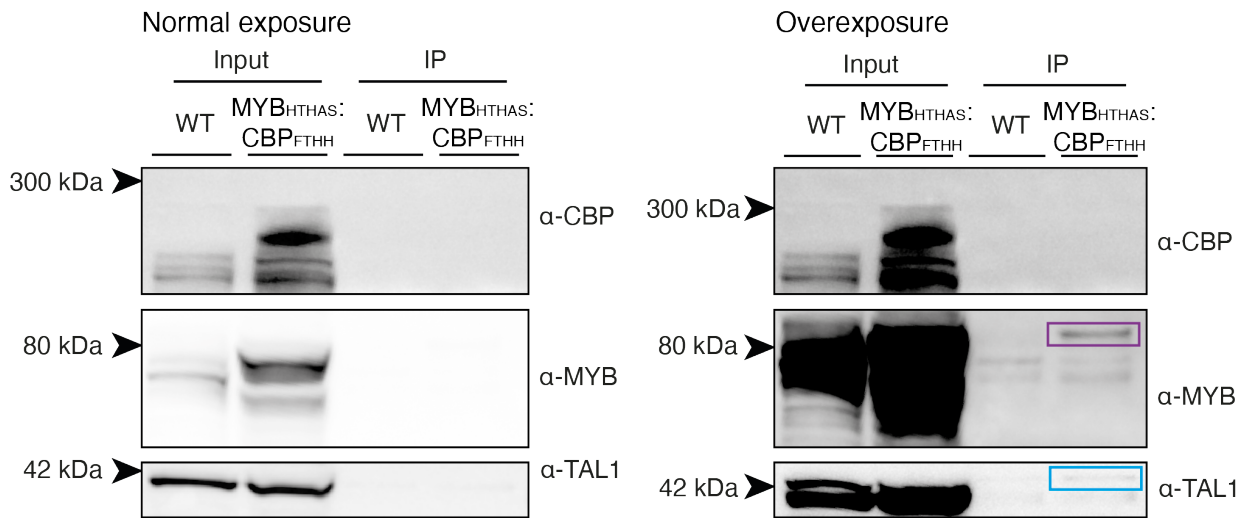
Figure 5.9: Streptavadin Co-Immunoprecipitation retains MYB-TAL1 interaction
Co-Immunoprecipitation using Streptavadin coated magnetic beads using WT Jurkats
and MYB$_{HTHAS}$:CBP$_{FTHH}$ cells. Left panels show normal levels of exposure. Right
panels show the same images as those on the left, however exposure has been greatly
increased. Bands for Co-Immunoprecipitated MYB is highlighted by a purple box.
Band for Co-Immunoprecipitated TAL1 is highlighted by blue box. Input shows 10%
of input to the Co-IP. n=1.

Co-IP against the Strep tag onto MYB at first did not appear to have resulted in
any successful IP (Figure 5.9, left panel). However, upon high overexposure of the
western blot from the Co-IP samples, a band at 85 kDa was visible in the tagged MYB
cell line sample when blotting for MYB (Figure 5.9, right panel). This suggested that
the Strep Co-IP was successful and was able to pull down full-length MYB-HTHAS
from the cell lysate, however it implied that this reaction was not efficient. This meant
that there was a much smaller proportion of protein in the IP lane than in the input,
which would explain why it was only able to be visualised upon high exposure of the
blot.

Blotting against TAL1 followed the same pattern as described for MYB. It was
not visible in the usual levels of exposure of the blot, however upon high overexposure
a band against TAL1 was visible (Figure 5.9, right panel). This suggested that pull-
down of MYB co-immunoprecipitated TAL1, implying that addition of the HTHAS

tag to MYB did not impede its ability to form protein-protein interactions with TAL1.

However, when blotting for CBP, even upon high overexposure of the western blot, no band for full-length CBP was visible. It is possible that as the level of MYB immunoprecipitated from the Strep pull-down was so low, this meant if there was any CBP binding to MYB, the amount of CBP interacting with the lower amounts of MYB visible was too low to be analysed in this way.

Overall, this implied that MYB and TAL1 protein-protein interactions were not impeded by the addition of the HTHAS tag onto MYB, however we were not able to confirm that CBP-MYB interactions were not disrupted. Previously, in Chapter 3, when we first examined protein-protein interactions within the CBP-TAL1 complex, we carried out a Co-IP for CBP and blotted for MYB (Figure 3.4). We should repeat this strategy of Co-IP here, co-immunoprecipitating for CBP and blotting for MYB to examine CBP-MYB interactions fully before we can draw conclusions as to the ability of these two proteins to interact upon the addition of affinity tags.

5.6.2   RIP-qPCR

When carrying out our initial investigation into the CBP-TAL1 complex as a whole, we carried out an RNA Immunoprecipitation followed by RT-qPCR to examine binding of each of the individual protein components of the CBP-TAL1 complex to the *TAL1* sense and antisense RNAs (Figure 3.7). This showed that each of the proteins within the complex did bind to both eRNAs. To confirm that RNA binding had not been disrupted by introduction of the FTHH tag onto CBP and the HTHAS tag onto MYB, we repeated the RIP-qPCR in our final MYB_{HTHAS}:CBP_{FTHH} cell line.

After we carried out the RIP-qPCR, binding of RNA to our tagged proteins was visible, particularly there was more enrichment of CBP binding to the *TAL1* eRNAs compared with both IgG samples. However, we noticed that there was significant binding of RNA to both IgG samples, and this binding was much higher compared to when we had previously done RIP-qPCR in WT cells. This meant that when we normalised to our high IgG levels in the MYB_{HTHAS}:CBP_{FTHH} cell line, it appeared as though MYB, Flag and HA were not binding to RNA. However, when comparing the raw Ct values with those of the WT RIP-qPCR done previously, the Ct values of CBP and MYB were similar (Figure 3.7). As we had only carried out one replicate of

204

this experiment due to time constraints, we were unable to draw any conclusions from this dataset. However, as this implied that there was contamination within the rabbit and mouse IgG samples providing potentially false positives for RNA binding, we also reanalysed the data to compare against WT IgG samples to identify any implications for RNA binding (Figure 5.10).
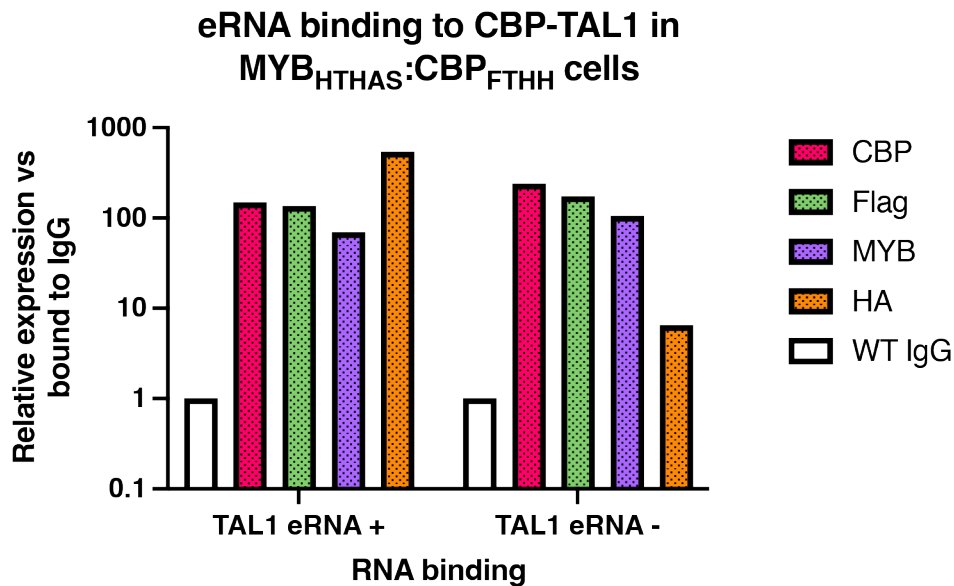
Figure 5.10: eRNA binding within the MYB_{HTHAS}:CBP_{FTHH} cell line
RNA immunoprecipitation followed by RT-qPCR shows binding to the eRNAs transcribed from the plus and minus strand of the *TAL1* enhancer. n=1. Data was normalised against the IgG data obtained from the WT RIP data.

Analysis of CBP, MYB, Flag and HA RIPs against RIPs from WT IgG data possibly indicated higher levels of eRNA binding than the WT IgG. This could have implied that tagging of CBP and MYB with the endogenous tags required for TA-CRAC did not abrogate RNA binding activity with the *TAL1* eRNAs. However, due to the nature of normalising against a different dataset and this data being from one replicate, we can only infer and cannot draw any conclusions.

This data suggested that CBP-TAL1 components retained the ability to bind RNA following tagging of MYB and CBP.

5.7 Discussion

In order to carry out TA-CRAC a single cell line containing two endogenously tagged proteins that are known to interact are required. In this chapter we have outlined the steps taken to generate a MYB_{HTHAS}:CBP_{FTHH} Jurkat cell line. This included design steps, based upon our successful tagging of CBP in Chapter 4 and carrying out the CRISPR/Cas9 mediated editing to introduce the HTHAS tag onto MYB. After generating our tagged cell line, we validated protein expression and further validated that tagged MYB and tagged CBP still retained function in terms of protein-protein interactions and RNA binding.

5.7.1 CBP CRAC

In the previous chapter we had successfully established methods to endogenously tag CBP with Flag-TEV-6xHis-Halotag in Jurkat cells. The Flag-TEV-6xHis tags inserted onto CBP could be utilised during CBP CRAC (Granneman, Kudla, et al. 2009; Thoms et al. 2015). CRAC involves double affinity purification steps, with the second step making use of the 6xHis tag on CBP. Due to the nature of the 6xHis not requiring any protein folding to bind to nickel beads, this interaction can be maintained during very strong denaturing wash steps. The strong denaturing wash in high salt will dissociate any interactions that are not between the nickel beads, the 6xHis tagged protein and any directly cross-linked RNA. Therefore, any artefacts or RNA that is bound indirectly to the protein of interest will not be extracted into this dataset. This therefore means that the RNA binding map produced for CBP will give only RNA bound directly to CBP in Jurkat cells.

5.7.1.1 CBP photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP)

RNA binding profiles for CBP have been established previously using different techniques, for example using photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP, D. A. Bose et al. 2017). PAR-CLIP involves labelling newly transcribed RNA within live cells using 4-thiouridine (4SU) to allow cross-linking at 365 nm (Hafner et al. 2010). Only 4SU labelled RNA will be

crosslinked to bound proteins at 365 nm and after immunoprecipitation for the protein of interest, crosslinked RNA can be labeled with radioactive γ^{32}P-ATP. These RNP complexes can be ran on an SDS gel and transferred to a membrane to confirm sizes, and upon band excision the crosslinked RNAs can be extracted. The resulting RNA species should therefore be those that are only binding to the protein of interest (Hafner et al. 2010).

5.7.1.2   Outlook for CBP-RNA binding experiments

If we were to carry out both CBP CRAC and CBP PAR-CLIP in Jurkat cells, the two techniques would complement each other well. As CRAC relies on the use of the affinity tags, PAR-CLIP could be used to control for artefact binding to the tags (Hafner et al. 2010). In addition CRAC is more specific and can isolate only RNA species bound directly to CBP, which would control for any indirect binding in PAR-CLIP (Granneman, Kudla, et al. 2009; Thoms et al. 2015).

The previous CBP PAR-CLIP data was carried out in Mouse Embryonic Fibroblasts (D. A. Bose et al. 2017), therefore both CRAC and PAR-CLIP for CBP in Jurkat cells could provide new insight as to the RNA binding behaviour of CBP within the context of the *TAL1* enhancer and the CBP-TAL1 regulatory protein complex, but also within a human cell line.

5.7.2   TA-CRAC to investigate CBP-MYB RNA binding profile

CBP CRAC would provide a wealth of information as to the RNA binding behaviour of CBP in Jurkat cells, however we sought to progress this a step further. We wanted to investigate how CBP binding affects the bound RNA population within a specific protein complex. In Jurkat cells, the CBP-TAL1 complex has been shown to be key for aberrant upregulation of the *TAL1* super-enhancer and driving transcription of oncogenic *TAL1* in this T-cell acute lymphoblastic leukemia cell line (Sanda et al. 2012; Mansour, Abraham, et al. 2014). Our previous investigations into the CBP-TAL1 complex have indicated that CBP may be part of this complex, due to its ability to Co-IP to each transcription factor within the complex (Section 3.2.2). Therefore, we sought to investigate the RNA binding activity within the CBP-TAL1 complex, to

determine if and how changing the protein composition within this regulatory protein complex would affect RNA binding to individual components.

TA-CRAC would allow us to gain an insight into the RNA binding behaviour of a specific protein, but only when this protein is found in complex with another. We can therefore apply this to investigating the transcription factors found within the CBP-TAL1 complex, by determining how the population of RNA bound to a certain transcription factor changes when in complex with CBP. In addition, TA-CRAC due to the nature of TA-CRAC utilising cross-linking, we can examine the exact section of an RNA's sequence that is bound to the protein of interest. This would therefore provide information as to the behaviour of the RNA species identified, as it may be possible to infer that the same RNA is binding to both proteins of interest examined using TA-CRAC at the same time, for example, if one protein was binding to a more upstream section of the RNA and the second protein was binding to a more downstream section.

5.7.2.1 RNA binding abilities of the CBP-TAL1 complex

When we examined RNA binding ability, each protein within the CBP-TAL1 complex not only exhibited RNA binding behaviour, but we were also able to show that each CBP-TAL1 complex protein bound to both the sense and antisense eRNA transcribed from the *TAL1* enhancer (Section 3.4). TA-CRAC is dependent upon two proteins that are known to interact, both with affinity tags and RNA binding capacity. In our case CBP was designed to be our first protein of interest, and we hoped to demonstrate that within a regulatory protein complex such as CBP-TAL1, our RNA binding maps produced by TA-CRAC may shed light on eRNA binding within regulatory protein complexes. Therefore, each transcription factor could have been suitable for investigation by TA-CRAC. As the CBP-TAL1 complex consists of six transcription factors: MYB, TAL1, GATA3, RUNX1, HEB and E2A, we needed to determine which transcription factor would be investigated first (Mansour, Abraham, et al. 2014).

5.7.2.2 TA-CRAC to investigate MYB

We chose to first investigate the RNA binding profile of MYB when found within the CBP-TAL1 complex. MYB knockdown in Jurkat cells has been shown to reduce *TAL1*, *RUNX1* and *GATA3* expression (Mansour, Abraham, et al. 2014). In addition, the *TAL1* enhancer element, where the CBP-TAL1 complex has been shown to bind and to influence activity of *TAL1* transcription, is formed around two MYB binding sites. This enhancer element is a *de novo* enhancer, meaning it is not found in other cell types or organisms and the existence of this regulatory element within this cell line is purely due to the introduction of the MYB binding sites. Moreover, this enhancer is monoallelic, where the WT allele does not contain this insertion mutation and therefore no enhancer is formed at this locus (Mansour, Abraham, et al. 2014). This heavily implies that MYB is an initiator of enhancer formation, and it has a very important role within this enhancer and this protein complex. Therefore, we chose to investigate the role of MYB within the CBP-TAL1 complex, and to determine if CBP binding to MYB does affect RNA binding, using TA-CRAC.

5.7.2.3 Prospects for using TA-CRAC to investigate TAL1

TAL1 was previously shown to be important for Jurkat cell survival, as upon TAL1 knockdown Jurkat cell survival was reduced (Sanda et al. 2012). In addition, investigating the role it plays within the CBP-TAL1 complex would shed light as to the action of a transcription factor found at its own enhancer, which could be very interesting. Moreover, in our WT RIP-qPCR experiment to investigate binding of each protein to the *TAL1* eRNAs, TAL1 showed the highest enrichment for binding to the sense eRNA, and showed the second highest binding to the antisense eRNA (Figure 3.7). Therefore TAL1-eRNA interactions appeared to be strong, and it would be interesting to determine if CBP had a role to play in determining this binding of RNA. Due to time constraints we were only able to generate one tagged cell line, through the tagging of MYB, however if we were to progress to a second transcription factor, we would have chosen TAL1 next.

5.7.2.4 Prospects for using TA-CRAC to investigate GATA3

Within our RIP experiment to determine global RNA binding ability to the transcription factors within CBP-TAL1, it appeared that GATA3 had the highest RNA binding ability due to the RNA visible in the GATA3 RIP lane being much more intense than for the other proteins (Figure 3.7). Therefore it would be interesting to gain an understanding as to the RNA binding profile of GATA3 within the cell line, and to determine if binding of CBP to GATA3 causes a change in the RNA species that are bound.

5.7.2.5 Prospects for using TA-CRAC to investigate RUNX1

A recent preprint suggested that many transcription factors contain RNA binding abilities, including RUNX1, which is also found within the CBP-TAL1 complex (Preprint, Oksuz et al. 2022). This corroborated our RIP data, where we also showed that RUNX1 had RNA binding abilities (Figure 3.7). Therefore TA-CRAC to examine binding of RNA to RUNX1 with CBP would be interesting, however there are stronger cases for other proteins within CBP-TAL1 to be investigated first.

5.7.3 Generating a cell line suitable for CBP-MYB TA-CRAC

In this chapter we described the establishment of the tandem tagged MYB_{HTHAS}:CBP_{FTHH} cell line, which will be used to carry out TA-CRAC.

During the generation of the CBP_{FTHH} cell line in Chapter 4, many optimisation steps were required to successfully yield our tagged CBP cell line. Despite addition of steps to enrich for successfully edited cells, the observed editing efficiency for correctly tagged CBP was low (Chapter 4). Therefore, upon designing the tagging strategy for tagging MYB, we had protocols and strategies in place that had higher likelihood of yielding successfully tagged MYB.

5.7.3.1 MYB

MYB is an 80 kDa protein, and is an important transcription factor, particularly within blood cell lineage. MYB has been shown to play an important role in the maintenance of progenitor cell states, as for differentiation to occur in these lineages

MYB expression must be reduced (Sheiness and Gardinier 1984; Sandberg et al. 2005). MYB has been implicated as oncogenic in many leukemias, including acute myeloid leukemia (AML) and chronic myeloid leukemia (CML) (Anfossi, Gewirtz, and Calabretta 1989). Also, MYB knockout has been shown to be embryonic lethal, with homozygous MYB knockout embryos observed as being deficient in full development of hematopoiesis, and thus did not survive the full gestation period (Mucenski et al. 1991). The ability of MYB not only to bind to DNA, but also to histone tails has implicated a role as a pioneer factor (Fuglerud, Lemma, et al. 2017).As previously discussed, binding of MYB to the *TAL1* enhancer is key for oncogenic *TAL1* transcription and Jurkat cell proliferation (Mansour, Abraham, et al. 2014). In K562 cells, through correlation of ChIP-seq, RNA-seq and ATAC-seq data, MYB has been argued in binding, either directly or indirectly, at over 12,000 genes (Lemma et al. 2021).

5.7.3.2 MYB-HTHAS

We sought to introduce a 6xHis-TEV-HA/Strep tag onto MYB. During the generation of the MYB$_{HTHAS}$:CBP$_{FTHH}$ cell line, many more positively tagged clones were identified than during the tagging of CBP. It was likely that this observed increase in efficiency was due to the smaller tag inserted, as insertion of larger tags is inherently less efficient. The HTHAS tag consisted of 170 bp, and was significantly smaller than the 1005 bp FTHH tag added onto the C-terminus of CBP (Chapter 4). Despite previous generation of CRISPR/Cas9 edited cell lines within Jurkat cells that suggested only 2% of clones contained designed edits (Borowicz et al. 2020), we were able to show that our added enrichment steps, such as use of a GFP tag to enrich for Cas9 editing, were successful in increasing the proportion of clones with our desired edits.

5.7.4 Further MYB$_{HTHAS}$:CBP$_{FTHH}$ validation experiments

Despite our promising results indicating that through tagging both MYB and CBP in tandem in one cell line, we had retained some function of both CBP and MYB to bind to the *TAL1* enhancer, there still remains future work to be carried out to confirm full protein function.

5.7.4.1 Confirmation of tagged CBP acetyltransferase activity

Firstly to confirm that CBP acetyltranserase activity, including to histones and other substrates, has not been impeded in the $\text{MYB}_{HTHAS}\text{:CBP}_{FTHH}$ cell line, a HAT assay. By extracting CBP-FTHH species, for example by doing a Flag IP, we could isolate the tagged CBP species, and then could use this in a HAT assay kit (for example Abcam). This would confirm that acetyltransferase activity by CBP is still functional, and could provide more information as to the interactions between CBP and MYB: as CBP acetylates MYB to activate its transcription factor activity (Dai et al. 1996; Fuglerud, Ledsaak, et al. 2018).

5.7.4.2 Establishing CBP-MYB protein-protein interactions in $\text{MYB}_{HTHAS}\text{:CBP}_{FTHH}$ cells

After tagging both CBP and MYB, we have not been able to confirm that MYB and CBP still interact. To achieve this, we should repeat the CBP Co-IP we carried out in section 3.2.2, and blot for not only MYB, but also for Flag and HA. If this was successful, this would confirm that the addition of the FTHH tag onto CBP and the HTHAS tag onto MYB have not disrupted protein-protein interactions. Furthermore, blotting for the remaining factors within the complex: TAL1, GATA3, RUNX1, HEB and E2A, would confirm that the CBP-TAL1 complex as a whole was still able to form in $\text{MYB}_{HTHAS}\text{:CBP}_{FTHH}$ cells.

5.7.4.3 Examining chromatin binding activity at the *TAL1* enhancer

If our hypothesis comprises that when CBP and MYB interact, these proteins are also able to bind to the *TAL1* eRNAs, we should confirm that through the introduction of the tags, chromatin binding is not impeded. ChIP-seq or ChIP-qPCR for MYB and CBP, alongside for Flag and HA would help to compare binding of the proteins against localisation of the tags and would confirm if the tags were affecting chromatin binding ability.

5.7.4.4 Outlook for MYB$_{HTHAS}$:CBP$_{FTHH}$ cells prior to TA-CRAC

As we have discussed there are still some experiments that are required to confirm characterisation of the CBP-TAL1 complex is still functional in terms of protein-protein interactions, HAT activity and activating behaviours at the *TAL1* enhancer. Upon confirming that the addition of the FTHH and HTHAS tags are not impeding protein functions, we will be able to progress to TA-CRAC preliminary experiments.

5.7.5 A plan to carry out TA-CRAC

In this chapter we have outlined the generation of a MYB$_{HTHAS}$:CBP$_{FTHH}$ cell line. We had begun to validate that through addition of the affinity tags to both CBP and MYB, protein expression was not abrogated, and both tagged CBP and MYB retained chromatin and RNA binding activity. Therefore, after exploring behaviour at the *TAL1* enhancer, and establishing functional protein-protein interactions (as detailed in section 5.7.4) we would then be able to progress to preliminary experiments prior to the TA-CRAC are now able to begin. Further preliminary experiments need to be carried out as proof of principles and as appropriate controls. It will be necessary to confirm that we can manipulate our tags and protease cleavage sites successfully in the way in which they are designed. For example, carrying out sequential pull-downs against the tags, will confirm that we can pulldown sufficient tagged MYB species from the first CBP pulldown. In addition, treating cells with TEV to access either the 6xHis tag on MYB or the Flag tag on CBP will show that the tags can be manipulated as designed in the TA-CRAC workflow.

After establishing the endogenous tags on CBP and MYB within this cell line can be manipulated as designed, we will then be able to use our MYB$_{HTHAS}$:CBP$_{FTHH}$ cell line to carry out TA-CRAC. This will identify the RNA population found binding to MYB, only when MYB is found within the CBP-TAL1 complex. This will shed light on the action of CBP and MYB within regulatory protein complexes. We hypothesise that as the CBP-TAL1 complex has key roles at enhancers in this cell line, any RNA binding maps generated may include eRNAs. This could provide new evidence for eRNA mechanisms and function.

Chapter 6

Dissecting the role of proto-oncogenic c-MYB at a model enhancer

6.1 Introduction

6.1.1 The *TAL1* enhancer is MYB dependent

As previously discussed in Chapters 3 and 5, the *TAL1* enhancer found 7.5 kb upstream of the *TAL1* transcriptional start site (TSS) is responsible for upregulation of high levels of oncogenic *TAL1* within Jurkat cells (Mansour, Abraham, et al. 2014). This enhancer element is formed due to the acquisition of two binding sites for the transcription factor MYB. Upon MYB binding, the remainder of the CBP-TAL1 protein complex is recruited, the chromatin region is marked for transcriptional activation through histone post-translational modifications (PTMs) and RNA polymerase II transcribes the *TAL1* enhancer into two defined enhancer RNAs (eRNAs) (Mansour, Abraham, et al. 2014; Lidschreiber et al. 2021).

CRISPR/Cas9 has been previously used to knockout one of the MYB binding sites within Jurkat cells (Mansour, Abraham, et al. 2014). Loss of MYB binding dramatically reduced H3K27ac levels at this enhancer element and also reduced *TAL1* gene expression (Mansour, Abraham, et al. 2014). Consequently, it is likely that this enhancer element is dependent upon MYB initiation.

We therefore sought to investigate if knocking down MYB within Jurkat cells would affect formation of the *TAL1* enhancer, the CBP-TAL1 complex and transcription of

the *TAL1* eRNAs (Figure 6.1).
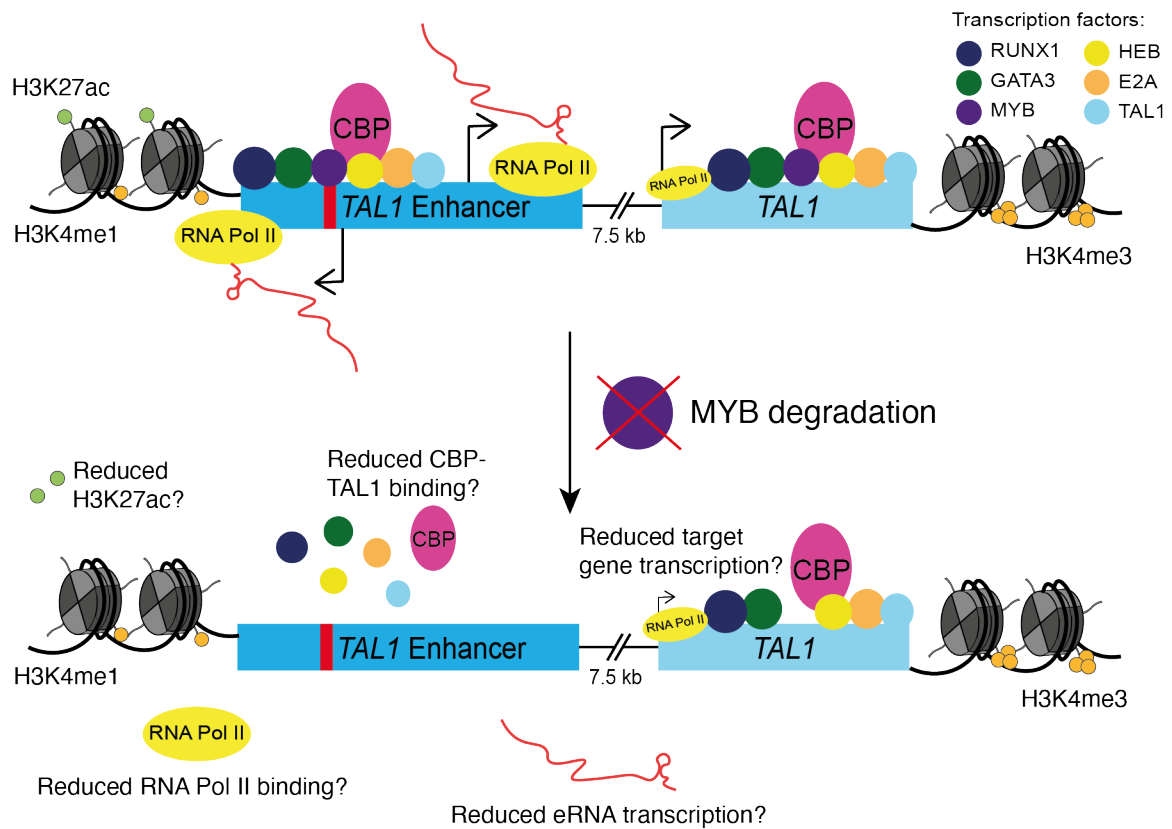
Figure 6.1: MYB degradation may affect *TAL1* enhancer activity
Upper panel shows active *TAL1* enhancer in WT Jurkat cells. Upon MYB degradation CBP-TAL1 complex binding at this locus may be reduced. This could result in reduced H3K27ac PTMs as less CBP would be bound. Consequently, recruitment of RNA polymerase II could be reduced and therefore fewer eRNAs would be transcribed. Consequently, target gene transcription may be reduced.

6.1.2 Knockdown and knockout systems

Loss of function experiments for a protein of interest are extremely popular methods used to decipher the function of a given protein. Proteins can be depleted at any stage of expression, be that due to changes in the gene coding sequence, targeting of the mRNA or targeting fully translated protein for degradation. There is therefore a wealth of choice in the methods available for inducing gene knockout or protein knockdowns within cells.

215

6.1.2.1   Mouse knockout models

A tool often used to examine the role of a protein, is to determine the effect of the loss of this protein within a whole organism. For example, mice that are homozygous null for a protein of interest can determine how critical a particular protein is to development (as reviewed by Capecchi 2001).

Homozygous null *MYB* mice do not survive past day 15 of foetal development, due to deficient heamatopoeisis (Mucenski et al. 1991). When loss of a protein results in inviable pups, this is termed embryonic lethality, and can highlight particular proteins that are critical for normal mammalian development.

6.1.2.2   CRISPR/Cas9 mediated gene knockout

Not only can CRISPR/Cas9 be used for genome editing and knock-in insertions, but gRNAs can target Cas9 cleavage to specific exonic regions (J.-P. Zhang et al. 2016). After Cas9 cleavage causes double strand breaks (DSBs), the error-prone DNA repair mechanism non-homologous end joining (NHEJ) can be utilised by the cell, introducing many erroneous mutations. Due to these insertions and deletions (indels), this can often result in a frame-shift of the coding sequence, meaning that the open reading frame for the protein coding sequence no longer contains the correct nucleotide sequence for the full-length protein.

Therefore, if gRNAs were to be designed to target early within the protein coding sequence, this can result in only truncated forms of proteins being translated. Consequently, no full-length protein would be expressed in the cell and this can therefore knock-out expression full-length protein (J.-P. Zhang et al. 2016).

6.1.2.3   Knockdown of mRNA by RNAi

Transcribed mRNA from the target gene of interest can be targeted in many ways to prevent translation of full-length protein product.

RNA interference (RNAi) using small interfering RNAs (siRNAs) are designed as double stranded RNA sequences, that target an mRNA for degradation (Elbashir et al. 2001). Upon entry to the cell, the Dicer enzyme cleaves the RNA to form a complete siRNA (Bernstein et al. 2001). The siRNA is subsequently recruited by the RNA induced silencing complex (RISC) and becomes single stranded (Matranga et al.

216

2005). After mRNA export from the nucleus to the cytoplasm, the siRNA can bind to any mRNA species with homology, resulting in mRNA cleavage and final mRNA degradation by the cell, blocking translation (J. Liu et al. 2004).

Short-hairpin RNAs (shRNAs) prevent mRNA translation in the same way as siRNAs, but siRNAs and shRNAs differ in their delivery into the cell (Paddison et al. 2002). siRNAs are required to enter the cell as double stranded RNA, however shRNAs can be transcribed within the cell itself. Therefore plasmids containing shRNAs can be transfected into cells for transient knockdowns, or lentiviral based methods can allow stable integration of a shRNA sequence within a cell line, causing stable knockout of the protein of interest (Paddison et al. 2002).

6.1.2.4   Knockdown of mRNA by CRISPR/Cas13

CRISPR/Cas9 is an RNA guided DNA nuclease, that targets double-stranded DNA within the genome for cleavage (Jinek et al. 2012). However, there are many other classes of Cas enzymes that utilise the CRISPR system, and one such example is CRISPR/Cas13. CRISPR/Cas13 is an RNA guided RNA nuclease that can be programmed to target RNA, including mRNAs, for degradation, therefore blocking translation and preventing protein expression (Abudayyeh, Gootenberg, Konermann, et al. 2016; Abudayyeh, Gootenberg, Essletzbichler, et al. 2017). CRISPR/Cas13 based knockdowns have been shown to reduce full-length protein expression to equal levels as comparative shRNAs, however off-target effects were observed to be much lower in the Cas13 samples as compared to the shRNA treated samples (Abudayyeh, Gootenberg, Essletzbichler, et al. 2017).

Another advantage of CRISPR/Cas13 systems, is that by developing cell lines with stably expressed Cas13, gRNA libraries encompassing many different RNA targets can be added and used for screening, in a similar way to screens done using Cas9 (Wessels et al. 2020).

6.1.2.5   dTAG

Proteins of interest can be endogenously tagged with degron tags and upon treatment with a small chemical, this can induce degradation of the target protein. The dTAG system was designed to harness CRISPR/Cas9 genome editing to insert an

FKBP12^{F36V} tag onto a protein of interest (Nabet et al. 2018). Upon the introduction of dTAG ligands, this tag has been shown to induce target protein degradation nearly completely (Nabet et al. 2018).

6.1.2.6 The Halotag can be utilised as a degron system

As previously discussed in Chapter 4, the Halotag is a 33 kDa protein tag that can be added endogenously to a protein of interest, with a large variety of ligands available for various functions (Los, Encell, et al. 2008). In addition to fluorescent ligands which we have utilised previously in Section 4.4.3.1 for single cell sorting of tagged cells, ligands that target the Halotag for protein degradation are also available (Buckley et al. 2015).

Proteolysis targeting chimeras (PROTACs) are small molecules that target a specific protein of interest for E3 ligase ubiquitination, causing subsequent degradation by the proteasome (Sakamoto et al. 2001). PROTACs have been designed to specifically label the Halotag (HaloPROTACs), which upon addition have been shown to cause Halotagged proteins to be degraded (Buckley et al. 2015).

6.1.3 Investigations into MYB knockdown in Jurkat cells

As MYB is a key proto-oncogene in many leukemias, not just within the Jurkat cell line, many groups have begun to investigate knockout or knockdown of MYB in various leukemia systems, as well as T-ALL and Jurkat cells.

6.1.3.1 RNAi targeting MYB in Jurkats

siRNAs have been used in T cell acute lymphoblastic leukemia (T-ALL) patient derived cell lines and other leukemic lines to target *MYB* for degradation (Lahortiga et al. 2007). This indicated that *MYB* degradation decreased proliferation rate of these cell lines (Lahortiga et al. 2007).

MYB knockdown using shRNAs was examined in Jurkat cells and it was found that many of the same genes that were downregulated after *MYB* knockdown were also downregulated after *TAL1* shRNA knockdown (Sanda et al. 2012). This implied further that the CBP-TAL1 complex acts together to influence target gene transcriptional activity (Sanda et al. 2012).

To investigate activity of the *TAL1* enhancer, a 400 bp region of the *TAL1* enhancer, incorporating the MYB binding site, was cloned into an enhancer luciferase reporter vector (Mansour, Abraham, et al. 2014). Upon transfection into Jurkat cells, if this was a viable enhancer, this would activate levels of luciferase transcription. Repetition of this enhancer assay coupled with *MYB* knockdown using siRNAs, showed that the *TAL1* enhancer sequenced exhibited significantly less transcriptional and enhancer activity upon *MYB* knockdown (Mansour, Abraham, et al. 2014). In addition, upon *MYB* knockdown, 27% of genes associated with a super-enhancer in Jurkat cells exhibited decreased gene expression (Mansour, Abraham, et al. 2014).

6.1.3.2 MYB small molecule inhibitors

A small molecule inhibitor for MYB known as mexicanin-1 has been developed, which when added to K-562 leukemia cells halted cell proliferation and decreased MYB target gene transcription (Bujnicki et al. 2012). Further small molecules to inhibit MYB have been developed, however these have been designed to disrupt the MYB-CBP interaction (Joy et al. 2021). CBP is a coactivator of MYB (Dai et al. 1996), where MYB has been shown to bind to the KIX domain of CBP (Parker et al. 1999). Competitive inhibitors designed to mimic MYB and bind to the KIX domain, disrupting CBP-MYB interactions have been developed to decrease MYB activity (Joy et al. 2021).

6.1.3.3 Knockdown of MYB in Jurkat cells to treat T-ALL *in vivo*

A long established anti-parasitic drug called Mebendazole has been shown to additionally induce MYB degradation, and has been used in Acute Myeloid Leukemia (AML) mouse models to halt disease progression (Walf-Vorderwülbecke et al. 2018).

To examine T-ALL progression, Jurkat cells were transplanted into mice and treated with Mebendazole as a preventative and curative measure (Smith et al. 2023). As a result, disease progression slowed (Smith et al. 2023), highlighting the clinical relevance of MYB knockdown in this cell line, and the potential for MYB as a T-ALL target.

6.1.3.4 MYB dTAG knockdown system

In a recent preprint developing a method to combine ChIP with Mass Spectrometry to analyse protein complexes, a cell line was established in Jurkat cells with MYB-dTAG (Nabet et al. 2018; Yong et al. 2023). The cell line was generated to examine the effect on protein-protein interactions after MYB knockdown. ChIP for MYB was carried out after MYB knockdown, and the subsequent samples were analysed by mass spectometry. Upon knockdown of MYB, interactions with MYB and the CBP-TAL1 complex did not appear to be diminished: TAL1, HEB, RUNX1, E2A, GATA3 and the paralogue of CBP; p300, were all still apparent, compared to a ChIP for IgG after MYB knockdown (Yong et al. 2023). ChIP for MYB in untreated cells compared to dTAG treated did show a reduction in MYB, however the remainder of the CBP-TAL1 complex was still detected, with none of these proteins being above the significance threshold for enrichment in either sample, showing the CBP-TAL1 complex was being pulled down by MYB to similar extents with and without dTAG treatment (Yong et al. 2023). Both ChIP and particularly mass spectrometry require large amounts of sample to be successful. Therefore, if after knockdown of MYB, a ChIP for MYB still resulted in a pulldown sample large enough to be able to carry out mass spectrometry, this implies that the knockdown of MYB in this cell line may not have been efficient enough to observe an effect (Yong et al. 2023).

For our own investigations we sought to also investigate the formation of the CBP-TAL1 complex, however we wanted to take a different approach, to specifically examine the recruitment of the complex at the *TAL1* enhancer.

6.1.4 Targeting MYB for degradation in Jurkat cells

Establishment and activity of the *TAL1* enhancer within Jurkat cells is likely to be MYB dependent (Mansour, Abraham, et al. 2014), and through degradation of MYB in this cell line we sought to uncouple and decipher action of the CBP-TAL1 complex, the *TAL1* enhancer and the *TAL1* eRNAs.

As we have discussed, many methods are available for loss of function studies and many strategies have begun to be used to decipher the role of MYB in Jurkat cells. However, full MYB knockdown or knockout to examine how formation of the CBP-TAL1 complex and *TAL1* enhancer along with transcription of the *TAL1* eRNAs is

affected has yet to be investigated.

In this chapter we will outline trialling different methods to both knockout and knockdown MYB to examine the effect upon Jurkat cells.

6.2 CRISPR/Cas9 mediated knockout of c-MYB

To firstly examine the role played by MYB in Jurkat cells, we sought to determine the effect of knocking out *MYB* using CRISPR/Cas9.

We chose to use this method of knocking out *MYB* within Jurkats first, as we had previously shown that we could successfully use the CRISPR/Cas9 system within this cell line (Chapters 4 and 5). We had shown that we were able to design effective guides that could target a specific recognition sequence, and although we had previously carried out knock-in experiments, not knock-out, these established techniques were readily available and could be easily adapted for this application.

6.2.1 gRNA design and construct generation

Three gRNAs targeting early exonic regions of *MYB* were designed to maximise potential for knockout of the gene. gRNAs were designed as described in Section 2.4.1.1 and are listed in Table 2.5. Using the UCSC genome browser CRISPR targets tool, there were no high scoring gRNAs that could be designed to cleave within exon 1 of MYB. Therefore we designed three gRNAs targeting across exon 2 (Figure 6.2, Appendix 8.1 for coordinates).

Each of the three gRNAs were cloned into the pX458 vector (as described in Section 2.2.6.3), as use of this GFP tagged Cas9 vector had previously yielded great success with correct cleavage of the designed target site, used previously for knock-in experiments (Sections 4.4.2 and 5.5).

Figure 6.2: Design of a CRISPR/Cas9 mediated knockout of MYB
(A) UCSC genome browser showing exon 2 of *MYB* gene. Blue arrow indicates location over the chromosome, red arrows indicate the three designed gRNAs (Table 2.5 for sequence, Appendix Table 8.1 for gene coordinates). (B) CRISPR/Cas9 mediated knockout of genes through introduction of DSBs within coding regions. NHEJ repair introduces errors and mutations that prevent protein transcription and thus results in knockout of the protein.

6.2.2 CRISPR/Cas9 targeting of MYB

Each gRNA-pX458 vector was transfected into Jurkat cells using the Neon electroporation system (Section 2.3.5.1). Transfected cell populations were incubated for 96 hours to allow sufficient time for the knock-out to take place. This included time for

several steps to occur: Firstly, for Cas9 to be expressed and form an RNP complex with the gRNA. Then for this RNP complex to target and cleave at the correct DNA sequence. Next to allow the double strand break (DSB) site to be repaired and consequently introduce erroneous mutations causing knockout of the *MYB* gene. Finally, this would allow sufficient time for downstream affects in the cell as a result of *MYB* knockout to have taken place.
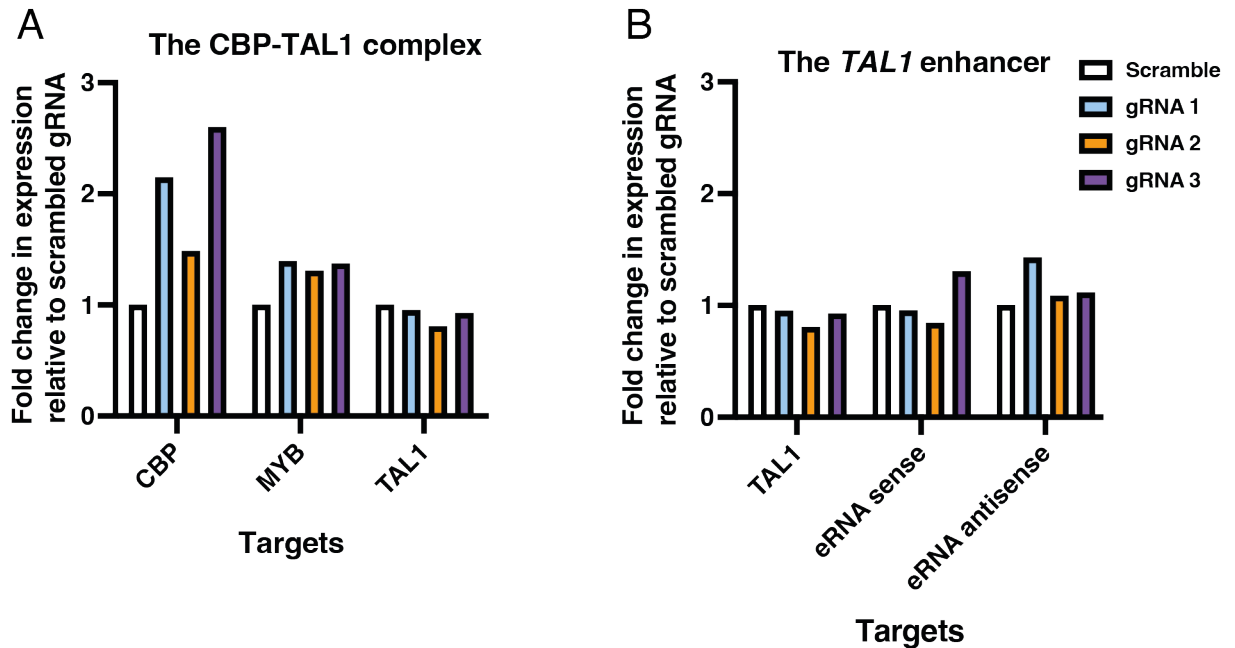
Figure 6.3: Expression of CBP-TAL1 complex factors and *TAL1* eRNAs after CRISPR/Cas9 knockout of *MYB*
Three gRNAs designed to target Cas9 to *MYB* exon 2 altered gene expression in Jurkat cells. RNA was extracted after 96 hours post-transfection, and RT-qPCR of cDNA carried out for *CBP*, *MYB*, *TAL1* and the *TAL1* sense and antisense eRNAs. Expression data was normalised to both *18S* and expressed as a fold change to the scrambled gRNA. n=1. (A) Fold change in gene expression of the CBP-TAL1 complex: *CBP*, *MYB* and *TAL1* after treatment with each *MYB* targeting gRNA. (B) Fold change in expression at the *TAL1* enhancer: *TAL1*, *TAL1* sense eRNA and *TAL1* antisense eRNA after treatment with each *MYB* targeting gRNA.

After 96 hours RNA was extracted and Reverse Transcriptase quantitative PCR (RT-qPCR) was carried out. We hypothesised that CRISPR/Cas9 mediated knockout of *MYB* may not only caused a reduction in *MYB* gene expression, but in addition

may reduce expression of the CBP-TAL1 complex due to the positive autoregulatory loop observed for this protein complex within Jurkat cells (Sanda et al. 2012). Consequently, we may observe reduced activity at the *TAL1* enhancer, including a reduction in the levels of eRNAs transcribed from the *TAL1* enhancer alongside a reduction in *TAL1* target gene expression.

This did not appear to be the case when examining gene expression of the CBP-TAL1 complex after our CRISPR/Cas9 mediated knockout of *MYB* (Figure 6.3A). Upon addition of gRNAs to target *MYB* for knock-out, there may have been a slight increase in *MYB* gene expression across the three targeting gRNAs compared to samples treated with the scrambled gRNA, but these were very small changes. In addition *CBP* expression may have increased with each targeting gRNA sample, with two of the three gRNAs observing over a two-fold increase in expression in comparison to the scrambled gRNA. There were no significant changes to *TAL1* expression after targeting *MYB* for knockout, and there were no distinct changes observable to the expression of the eRNAs transcribed from the *TAL1* enhancer. However, as each gRNA was only targeted once, more repeats need to be carried out before conclusions can be drawn.

A potential explanation for why our observed *MYB* knock-out did not show the same gene expression pattern as we hypothesised could lie with the fact that we had harvested samples for RNA extraction after 96 hours. Due to the drastic increase in *CBP* expression observed, a conceivable explanation was that this time point was too late. It could be possible that downstream effects of *MYB* knock-out had taken place in the cell, and further to this the cells had then compensated for this loss by activating higher levels of gene expression across the cell. An increase in *CBP* expression could stimulate higher levels of transcriptional activity throughout the genome, due to the increase in histone acetyltransferase activity exhibited by the more numerous CBP population, and the increased activating chromatin environments.

6.2.2.1 Outlook for CRISPR/Cas9 knockout of *MYB*

As we had shown previously during the validation of our gRNAs for targeting the CBP C-terminus to tag CBP, Jurkat cells exhibited low efficiency with CRISPR/Cas9 mediated editing (section 4.3.4). In order to observe any editing at the DSB site in

Chapter 4, we had to enrich the cells for Cas9 editing through the GFP tag on the Cas9 by Fluorescence Activated Cell Sorting (FACS, section 4.4.2).

During our *MYB* knockout experiments we had not carried out FACS to enrich for Cas9, therefore the proportion of transfected cells and edited cells was likely to have been low.

In addition, although carrying out FACS to generate polyclonal enriched cell lines for each gRNA knockout would be useful, to gain the most insight into *MYB* loss of function in this manner, it would be best to generate clonal cell lines for each knockout. This would allow us to truly observe the effect of *MYB* knockout throughout the population and would allow us to be certain that the effect we would observe would be due to the knockout of *MYB*.

6.3   Generating a MYB$_{Halotag}$ Jurkat cell line

6.3.0.1   Considerations for a *MYB* knockout Jurkat cell line

As previously discussed, one potential issue of developing the CRISPR/Cas9 mediated knockout of *MYB* further is the necessity of generating clonal lines. In mice, *MYB* has been shown to be embryonic lethal, meaning that embryos that were homozygous null for *MYB* did not survive the full gestation period (Mucenski et al. 1991). Of course, mouse foetal development is a different system to the one used in our study, but this still provided evidence as to the importance of MYB in heamatopoeisis.

As Jurkat cells seem to be dependent upon *MYB* expression through survival being tied to the *TAL1* enhancer (Sanda et al. 2012; Mansour, Abraham, et al. 2014), there is the possibility that a Jurkat cell line with stable complete knockout of *MYB* would not be viable.

6.3.0.2   Developing a MYB knockdown system in Jurkat cells

As a complete knockout of *MYB* within Jurkats may not be a viable option, we turned to inducible degradation systems. This would allow us under normal conditions to keep the cell population healthy through expression of full-length MYB, with degradation of MYB being induced at chosen times.

225

This also would provide scope for titratable knockdowns of MYB: as we may be able to achieve a knockdown to such a level where some full-length MYB is still expressed to allow cell survival, but sufficient MYB would be knocked down to observe a measurable effect at the *TAL1* enhancer and throughout the genome. In addition, by pulsing with the required reagents for knockdown, we could not only knockdown MYB, but experiments could be done well after the knockdown, to measure the effect on the cell of *MYB* gene expression re-establishing MYB levels within the cell. This could lead to a new avenue of investigations, as it could potentially allow us to measure on a temporal level how the CBP-TAL1 complex is acting at the *TAL1* enhancer and how this enhancer is initiated.

6.3.0.3   Choice of a Halotag based system for MYB knockdown

As previously discussed, there are many inducible degradation systems available, in particular those that utilise the addition of a tag inserted by CRISPR/Cas9, such as the dTAG system (Nabet et al. 2018). We had previously shown success with tagging CBP with a Halotag within Jurkat cells and had shown this tag was well tolerated in this cell line (Section 4.4.3). Additionally, we had shown that in Jurkat cells the Halotag could be successfully labelled using ligands that were commercially available, in our case a fluorescent ligand named TMR (Promega) (Section 4.4.3.1). An alternative ligand to label the Halotag are HaloPROTAC ligands, which are small molecules designed specifically to target the Halotag for degradation by E3 Ubiquitin ligase (Buckley et al. 2015). This has been shown to efficiently degrade Halotagged protein in other cell lines (J.-H. Lee et al. 2021).

Therefore, we had many of the tools in place to be able to generate a $\text{MYB}_{Halotag}$ cell line. Upon tagging MYB with a Halotag, we could then use HaloPROTACs to degrade MYB across Jurkat cells, to examine the effect of loss of MYB within the CBP-TAL1 complex and at the *TAL1* enhancer (Figure 6.4).

226
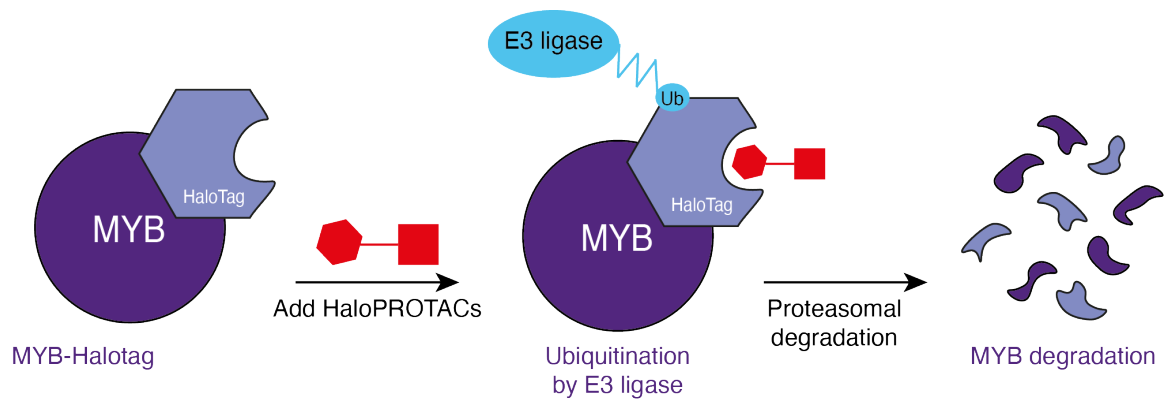
Figure 6.4: HaloPROTACs target MYB-Halotag for degradation

MYB tagged with a Halotag is subject to ubiquitination by E3 ligase upon labelling by HaloPROTACs. The Proteasome recognises the protein ubiquitination and targets MYB-Halotag for degradation.

6.3.1 Experimental design
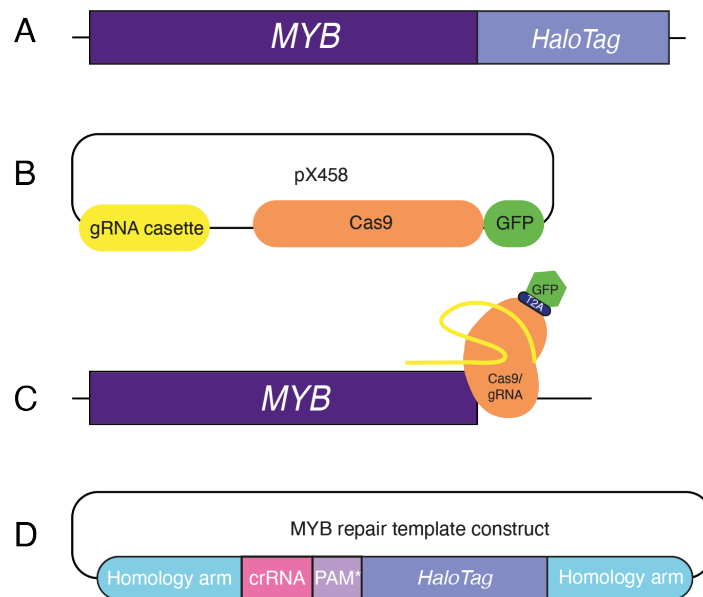
Figure 6.5: MYB-Halotag construct design

(A) A Halotag is to be inserted onto the C-terminus of MYB in Jurkat cells. (B) The pX458 plasmid containing the MYB C-terminal gRNA was used, encoding a self-cleaving GFP-Cas9, separated by a T2A cleavage site. (C) The gRNA from MYB-pX458 targets the C-terminus of MYB. (D) MYB-Halotag repair template design.

As we had successfully been able to generate homozygous tagged MYB with a 6xHis-TEV-HA/Strep tag (MYB$_{HTHAS}$) in the previous chapter (section 5.1), we chose to use the same tagging approach to tag MYB with a Halotag. The same gRNA to target MYB C-terminus was used within the pX458 vector (MYB-pX458), and the Halotag sequence was cloned into the MYB homology arms plasmid (MYB-Halotag-RT, section 2.2.6.4) for the HDR template (Figure 6.5).

We chose to insert the Halotag into wild type (WT) Jurkats, so the only tag present would be the Halotag on MYB.

6.3.2   Generating a polyclonal MYB$_{Halotag}$ cell line

To begin to generate the endogenously tagged MYB$_{Halotag}$ cell line, firstly the MYB-pX458 and MYB-Halotag-RT plasmids were co-transfected into WT Jurkat cells using the Neon electroporation system (Section 2.3.5.2). After 48 hours transfected cells were taken for FACS in order to collect a population of cells that were expressing GFP-Cas9. This collected population of cells would be enriched for Cas9 mediated gene editing and the number of correctly edited cells would therefore be at a higher proportion (Figure 6.6A).

After FACS, DNA was extracted from the polyclonal population and PCR was carried out to identify if the Halotag was visible within the DNA sequence. PCR of WT sequence should have given rise to a fragment of 650 bp, whereas insertion of the Halotag should have resulted in this sequence being 900 bp larger, giving a total fragment of 1550 bp (Figure 6.6B).

Lane 2 showed PCR of DNA from cells transfected with both the pX458-MYB and MYB-Halotag-RT plasmids. There were two bands visible, one at 650 bp showing the WT sequence, and a larger band running at approximately 1550 bp indicating the presence of Halotag sequence.
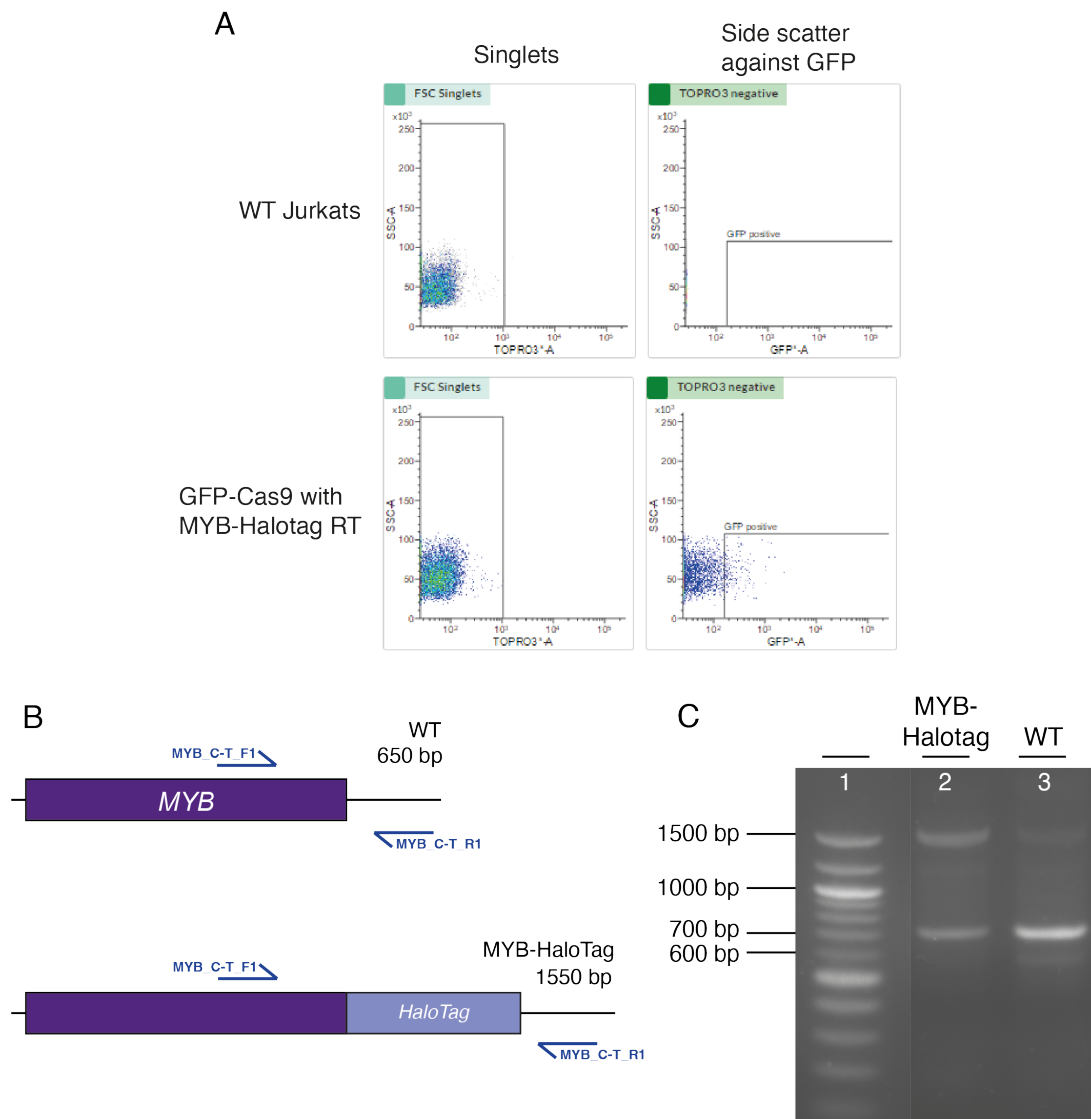
Figure 6.6: Generating a polyclonal MYB-Halotag population
(A) FACS data for generating MYB-Halotag polyclonal line. Upper panels show WT Jurkats with the singlets in left panel and GFP signal in right panel. Lower panels show edited cells with GFP-Cas9 and MYB-Halotag with the singlets in the left panel and the GFP signal in the right panel. (B) PCR for C-terminal region of MYB, with the WT sequence and the MYB-Halotag sequence. Purple arrows show binding of the PCR primers to indicate the difference in sizes with the addition of the Halotag. WT MYB gives rise to 650 bp fragment, MYB-HaloTag results in a 1550 bp DNA fragment (C) PCR of polyclonal MYB-Halotag DNA. n=1.

Lane 3 showed DNA from untransfected WT cells as a control. There was a strong band at 650 bp, showing the WT sequence had been successfully amplified. Additionally there were further bands that were very faint, but present in the PCR product from the WT cells. There was a faint band at 500 bp, and there was also a faint band at 1550 bp, which was of similar height to the band used in the adjoining lane to imply successful Halotag insertion. As this band was so faint it was likely due to overspill of the edited cells into the unedited cells sample, however we could not conclude this definitively.

Overall, this implied that there had been some successful insertion of the Halotag onto MYB, and we could progress this polyclonal line by carrying out single cell sorting.

6.3.3 Generating a clonal MYB$_{Halotag}$ cell line

Our previous analysis of the polyclonal MYB$_{Halotag}$ cell line indicated that there may have been some proportion of the population that successfully contained the Halotag (Figure 6.6C). Therefore we chose to proceed with single cell sorting of the population to obtain and screen for clonal lines. During the previous generation of the CBP-Flag-TEV-6xHis-Halotag (CBP$_{FTHH}$) cell line, cells were labelled with a fluorescent Halotag ligand prior to single cell sorting to aid in selection of positive clones (Section 4.4.3.1). However, as we did not observe significant enrichment for edited clones with this method, we proceeded with single cell sorting the cell population without any sorting based upon fluorescence.

After single cell sorting, clones were expanded until they reached a sufficiently large population for screening by PCR. 120 clones were picked and screened using PCR with a lower fidelity enzyme to enable large scale screening and products were ran on an agarose gel. From this, eight clones of interest were identified and DNA extraction was repeated. PCR was then carried out using a higher fidelity enzyme to more clearly genotype the clones (Figure 6.7) and the same primers were used as those for the screening of the polyclonal population (Figure 6.6).

As shown previously in Figure 6.6B, WT cells would give rise to a single band of 650 bp. If the Halotag was present this would result in a band at 1550 bp. If only a single band was present at 1550 bp this would indicate a clone that was homozygous

for MYB-Halotag, however if both a band at 650 bp and at 1550 bp was present this would indicate a heterozygous clone.

Of the eight clones of interest that were further screened, it appeared that six clones contained some level of Halotag insertion (Figure 6.7A). This was shown by clones 68, 71, 115, 112, 114 and 120 having bands at 1550 bp which was the correct size for Halotag insertion. In this case this 1550 bp band was not present in the WT control, providing further evidence that this was the Halotag sequence. There were multiple bands of varying sizes in each clone, however it is of note that PCR of DNA from clone 112 did not appear to produce a band at 650 bp, indicating this may have been a homozygous clone.
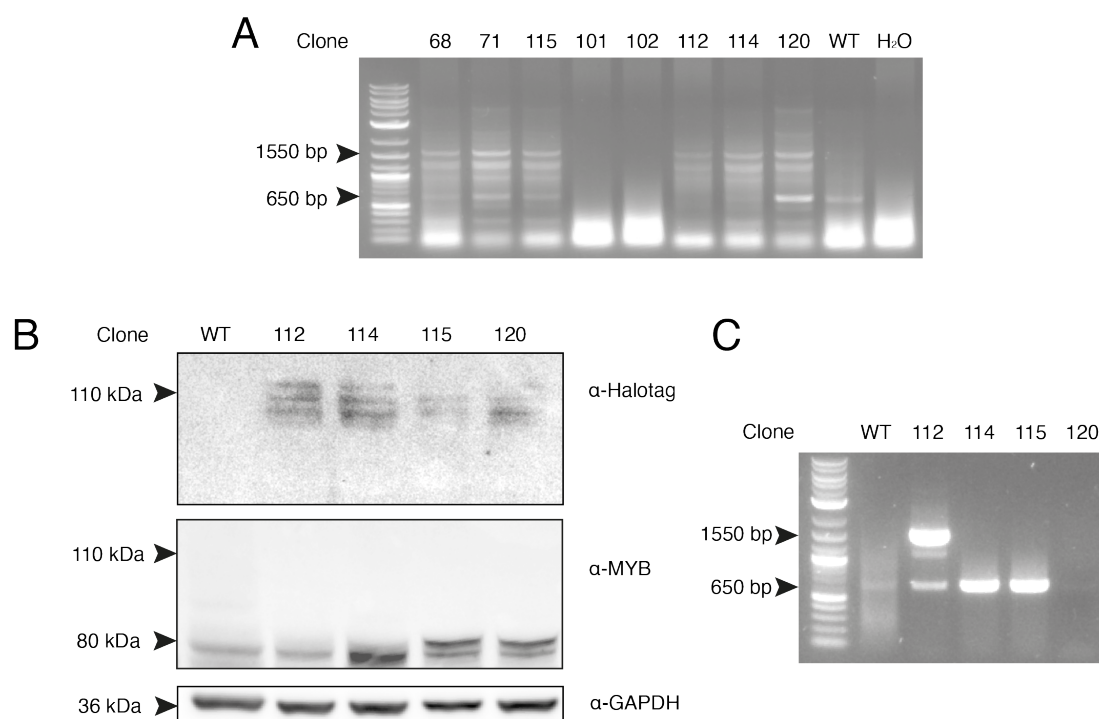
Figure 6.7: Clonal screening of MYB-Halotag clones
(A) PCR of MYB-Halotag clones ran on an agarose gel. (B) Western blot of WT Jurkats and MYB-Halotag clones. (C) Further PCR screening of MYB-Halotag clones. n=1.

6.3.3.1 Western blotting for MYB$_{Halotag}$ clones

After this second stage of clonal screening, we sought to determine whether the clones of interest were expressing Halotag protein. Some clones did not survive cell expansion, therefore clones 112, 114, 115 and 120 were used for western blot analysis (Figure 6.7B).

MYB has a molecular weight of 80 kDa and the Halotag is also of substantial size, adding 33 kDa to the size of the full-length protein. Therefore, it would be expected that MYB-Halotag should be visible at around 110 kDa.

When probing whole cell lysate from each of the clones for the Halotag, bands were visible at 110 kDa. Additionally these bands appeared in a triple banding pattern which we have observed previously as a very distinctive marker for MYB in our previous blots from Jurkat lysate (Figures 3.3 and 3.4). Therefore, this indicated that the Halotag and in particular MYB-Halotag was being expressed in each of the clones.

Blotting for MYB showed bands visible at 80 kDa, which is what would be expected of full-length untagged or WT MYB. If MYB had been successfully tagged, we would have expected to visualise a band at 110 kDa when blotting for MYB also. We were unable to visualise this, potentially due to the larger MYB-Halotag species being more unstable and may be prone to degradation during cell lysis steps.

6.3.3.2 Confirmation of tag insertion through PCR

As our western blots indicated MYB-Halotag expression, but did not confirm it, we chose to PCR screen these clones a further time. PCR of clones appeared to differ when done previously (Figures 6.7A and C). On this occasion, clones 114, 115 and 120 only had a single band at 650 bp, showing the same banding pattern as the WT DNA, indicating they were WT and there was no Halotag present. However, clone 112 had a very distinct banding pattern of 1550bp and 650 bp, implying this clone was heterozygous.

One possible explanation for this is that the cells may be selecting amongst the population to lose the tag, as clone 112 initially seemed to be homozygous and 114, 115 and 120 did appear to have the tag.

We therefore proceeded with clone 112 as a heterozygous clone, however we regularly repeated genotyping of this line to ensure the tag was still being expressed.

6.4 Titrated knockdown of MYB in Jurkat cells

As we had established the MYB$_{Halotag}$ cell line, we sought to use this in knockdown of MYB within Jurkat cells. As we had generated a heterozygous tagged line, this would mean that upon HaloPROTACs addition, the induced knockdown would not be complete across the cell. However, due to the embryonic lethality of *MYB* knockout (Mucenski et al. 1991) and the Jurkat cell dependence upon MYB (Sanda et al. 2012; Mansour, Abraham, et al. 2014), it is possible that complete knockdown of MYB would not be viable. Therefore we chose to proceed with our heterozygous cell line, hoping to diminish MYB levels to a suitable level to see an effect on the cell, whilst still retaining sufficent levels of MYB to still retain cell viability.

6.4.1 Optimising HaloPROTACs conditions

We firstly labelled MYB-Halotag cells with HaloPROTACs according to the manufacturer's instructions for CRISPR-edited Halotag cell lines, with DMSO added to MYB$_{Halotag}$ Jurkats for untreated samples. We added HaloPROTACs to a final concentration of 300 nM and left cells to incubate for 18 hours, then harvested samples for RT-qPCR to analyse changes in gene expression patterns (Figure 6.8).

As a result of MYB degradation, we hypothesised we would observe a downregulation of expression of *TAL1* and both eRNAs transcribed from the *TAL1* enhancer. If MYB were to be degraded, this would mean fewer species were available to bind and initiate the *TAL1* enhancer. Thus, *TAL1* enhancer activity would be decreased, and the output of *TAL1* gene expression as well as eRNA transcription would be reduced. Although the MYB$_{Halotag}$ line was heterozygous for the tag, meaning that not all MYB species found within this cell line could be degraded through HaloPROTACs addition, we wanted to test whether the monoallelic tag was sufficient to observe downstream consequences of MYB degradation.
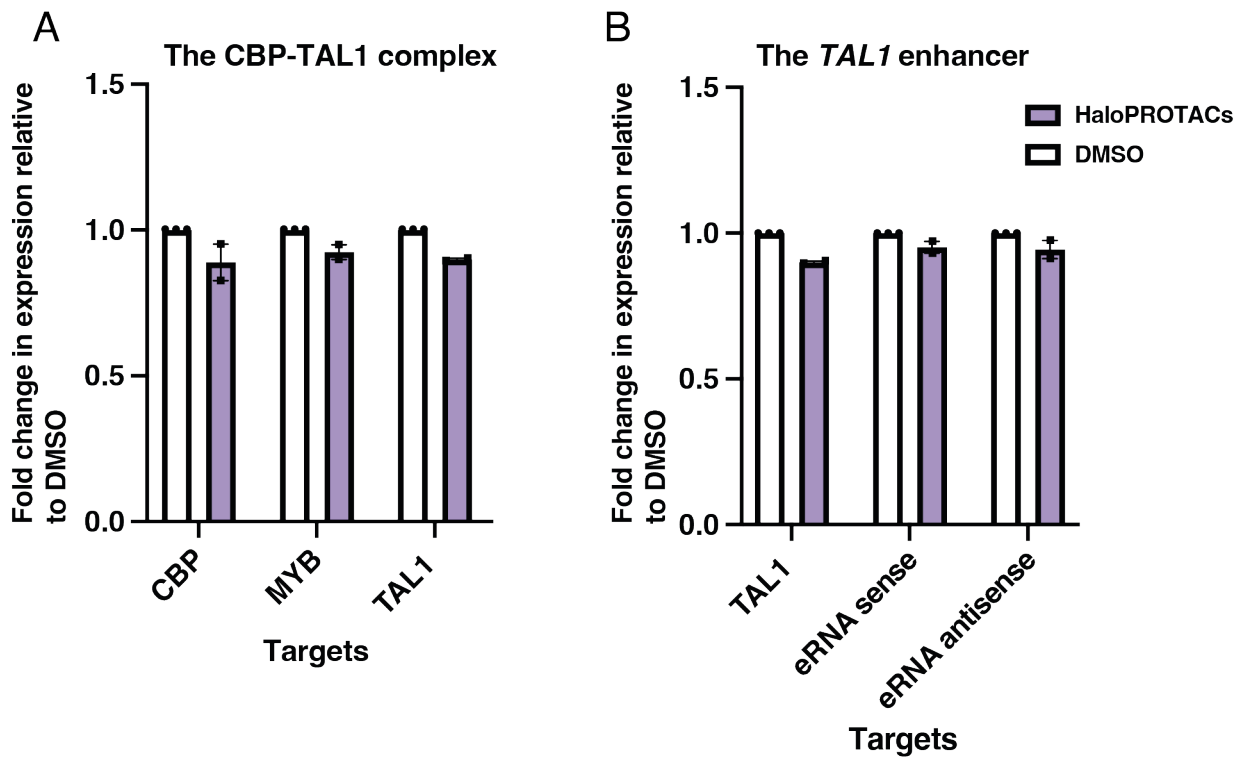
Figure 6.8: HaloPROTACs knockdown of MYB$_{Halotag}$

RT-qPCR of MYB$_{Halotag}$ cells treated with HaloPROTACs (purple), or with DMSO (white). Expression is normalised to *18S*, but expressed as a fold-change relative to DMSO treated cells. Error bars denote standard error mean (SEM), DMSO n=3, HaloPROTACs n=2. (A) Fold change in expression of components of the CBP-TAL1 complex: *CBP, MYB, TAL1*. (B) Fold change in expression of components of the *TAL1* enhancer: *TAL1, TAL1* sense eRNA and *TAL1* antisense eRNA.

After 18 hours of HaloPROTACs treatment, there may have been a very slight decrease in gene expression of the CBP-TAL1 complex components (Figure 6.8A), however more replicates would be required to draw conclusions as to the extent of the knockdown's effect. *CBP, MYB* and *TAL1* may have appeared to have a small decrease in expression compared to in untreated cells. This could suggest that MYB was being degraded and as a consequence of the autoregulatory loop within Jurkat cells, *CBP* and *TAL1* gene expression were being downregulated.

At the *TAL1* enhancer, both the *TAL1* sense eRNA and *TAL1* antisense RNA also may have shown a small decrease in expression compared to untreated cells. This could have implied some level of downregulation of eRNA transcription as a consequence

of MYB knockdown, albeit a small one. This may have suggested that a decrease in MYB was reducing activity at the *TAL1* enhancer, however this difference was not statistically significant.

This data provided a first glimpse that knockdown of MYB, utilising HaloPROTACs and MYB-Halotag could at least weakly affect gene expression in Jurkat cells. However, as such a small decrease in expression was observed, it suggested that conditions needed optimising, as we had potentially managed to knockdown some MYB within the cell, but not enough to observe significant changes to gene expression and enhancer activity.

6.4.2   Increasing HaloPROTACs knockdown time

As HaloPROTACs treatment indicated there may be some small level of MYB knockdown having an effect on the cell (Figure 6.8), we sought to increase the time and concentration of HaloPROTACs treatment. We hoped that increasing both the time and concentration of HaloPROTACs treatment would enable us to infer why only small changes in gene expression were observed previously. It could have been due to the HaloPROTACs treatment not being at a high enough concentration, or the treatment not being done for sufficient length of time. Or, it could have been due to the cell line itself, with not enough MYB species containing the Halotag to observe a knockdown and its consequences.

Therefore, HaloPROTACs concentration was increased from 300 nM to a final concentration of 1 μM, and left to incubate for 24 and 48 hours. After 24 hours, samples were taken for western blot analysis and after 48 hours samples were taken for both western blotting and RNA extraction for RT-qPCR.

Firstly to examine levels of MYB protein and to confirm the protein was being degraded, samples were analysed by western blotting (Figure 6.9).
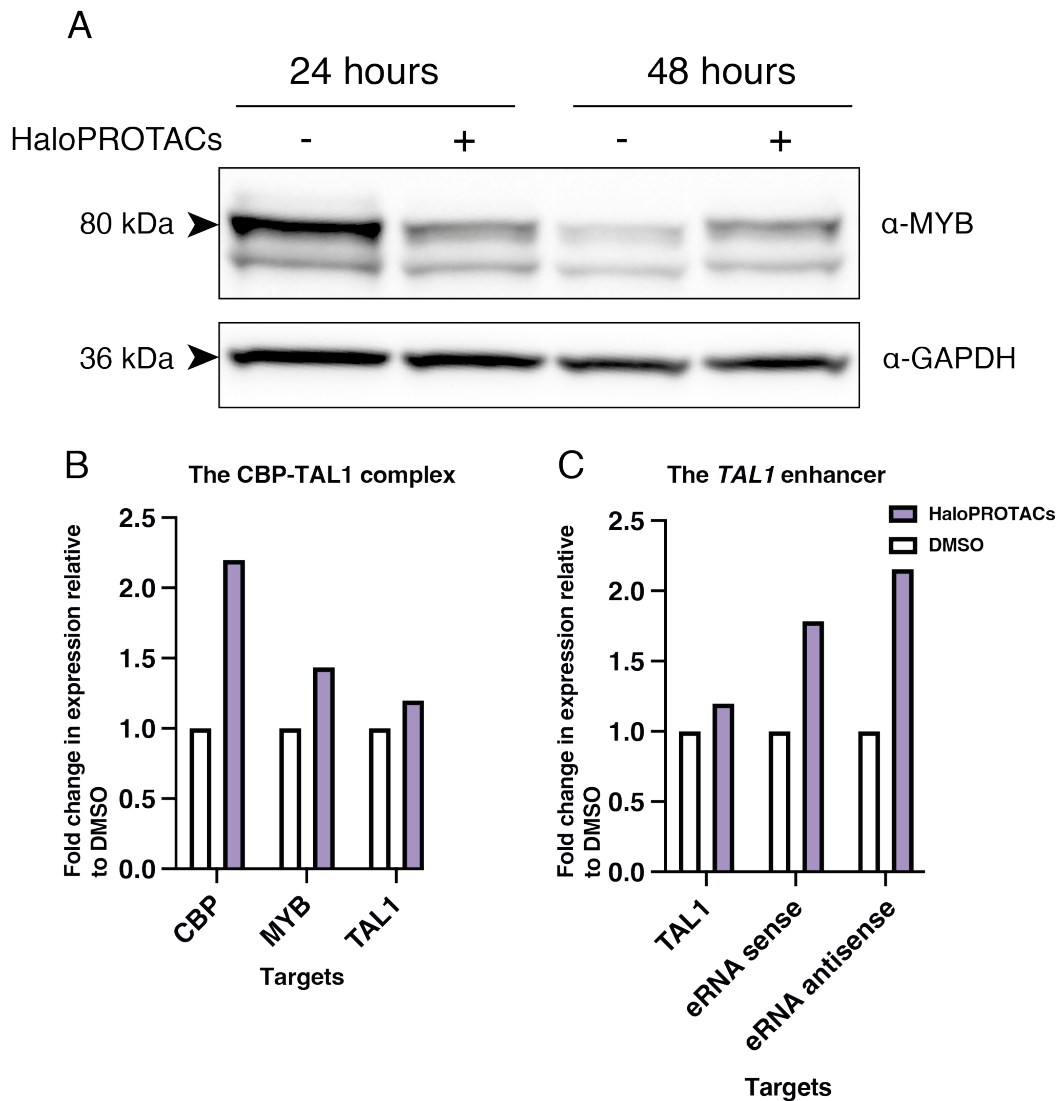
Figure 6.9: Addition of increased concentration of HaloPROTACs to MYB$_{Halotag}$ cells (A) Western Blot showing treatment of HaloPROTACs after 24 and 48 hours in MYB-Halotag cells. Probing against MYB (upper) and GAPDH as a loading control (lower). (B and C) MYB-Halotag cells treated with HaloPROTACs for 48 hours were taken for RT-qPCR analysis. Expression is normalised to *18S*, but expressed as a fold-change relative to DMSO treated cells. (B) Fold change in expression of components of the CBP-TAL1 complex: *CBP*, *MYB*, *TAL1*. (C) Fold change in expression of components of the *TAL1* enhancer: *TAL1*, *TAL1* sense eRNA and *TAL1* antisense eRNA. of *CBP*, *MYB*, *TAL1*, *TAL1* sense RNA and *TAL1* antisense RNA expression. n=1.

6.4.2.1 Examining protein levels of MYB knockdown

Upon examining protein levels by western blotting, after 24 hours of HaloPRO-TACs treatment, full-length MYB levels appeared to be reduced compared to samples treated with DMSO (untreated) (Figure 6.9A). Full-length MYB at 80 kDa was observed with an intense band in untreated samples, and a band of comparible size was observed from cells treated with HaloPROTACs for 24 hours, however the intensity of this band was reduced. This therefore implied that with a higher concentration of HaloPROTACs and a longer treatment time, some degradation of MYB-Halotag full-length protein was observable.

However, MYB protein levels showed a different pattern after treatment for 48 hours at the same HaloPROTACs concentration. Expression of full-length MYB appeared very low in the untreated samples after 48 hours, observed by a very faint band at 80 kDa. Comparatively, a more intense band was visible at 80 kDa in the treated samples compared to the untreated. Therefore, this implied there was higher expression of full-length MYB in $MYB_{Halotag}$ cells treated with HaloPROTACs for 48 hours, compared to the control.

When comparing the band intensities across the two time points, full length MYB protein levels appeared to be similar after both 24 and 48 hours of HaloPROTACs treatment, however it was possible that there were slightly fainter bands for GAPDH after 48 hours, compared to after 24 hours. Therefore, there were two possible explanations, depending upon comparison between the two timepoints or within a single timepoint. One possible explanation for the MYB protein expression observed was that HaloPROTACs treatment induced MYB degradation to similar extents upon treatment for both 24 and 48 hours, when comparing between the two timepoints. Or, when examining treated vs untreated within each timepoint, after 24 hours Halo-PROTACs treatment induced MYB knockdown, but after 48 hours MYB protein expression was re-established. This could be because there were no longer sufficient HaloPROTACs molecules available in the cell, due to the half-life of the HaloPRO-TACs molecules or the ligands being used up for degradation, meaning that an increase in MYB gene expression was favoured by the cell to maintain cell viability.

To be able to draw further conclusions from this western blot analysis, this would need to be repeated.

6.4.2.2 Gene expression may have been altered after HaloPROTACs treatment

Upon treatment with HaloPROTACs for 48 hours, samples were also taken for RNA extraction, and were used for RT-qPCR to analyse gene expression within MYB$_{Halotag}$ Jurkat cells (Figure 6.9B).

Treatment with HaloPROTACs of MYB$_{Halotag}$ cells appeared to increase gene expression across the targets examined. *CBP* expression appeared to be higher compared to the control, untreated cells. In addition, expression of the sense *TAL1* eRNA appeared to increase, along with expression of the antisense *TAL1* eRNA, which was increased even further compared to control, untreated cells. *TAL1* gene expression was slightly increased and there was also an increase in gene expression of *MYB* visible.

Due to time constraints, this experiment was only done with one replicate, so more need to be carried out to draw true conclusions. However, this analysis could have suggested that after 48 hours of HaloPROTACs treatment, *MYB* gene expression was being reestablished, therefore increasing the amount of protein that could initiate the *TAL1* enhancer, increasing activity of this enhancer element and thus giving rise to the upregulation of the *TAL1* eRNAs observed. This would also explain the increase in *TAL1* expression; the target gene of the *TAL1* enhancer. It was possible that in order to achieve this increase in *MYB* gene expression, *CBP* expression was also being increased, to drive transcriptional activation through the gene expression environment.

However, to draw any conclusions from this, further replicates are required. Coupled with our western blot analysis at both 24 and 48 hours (Figure 6.9A), samples for RT-qPCR should be taken at various time points, including 24 and 48 hours, as this would help to gain further insight as to the gene expression environment at this stage of MYB knockdown.

6.5   Discussion

6.5.1   Generation of CRISPR/Cas9 mediated *MYB* knockout cell lines

In previous chapters, we had shown great success with targeting designed gRNAs to the correct target site and we were successfully able to induce indel mutations at these loci (Chapters 4 and 5). Therefore, we had established methods to try and generate *MYB* knockouts using CRISPR/Cas9.

6.5.1.1   Outlook for generation of *MYB* knockout cell lines

Transfection of the gRNAs targeting *MYB* for knockout into Jurkat cells resulted in polyclonal cells (Figure 6.3). These were analysed using RT-qPCR and there had been no form of enrichment carried out for editing within these cells. We have shown previously that Jurkat cells are very difficult to transfect, with very low numbers of positively edited cells identified when carrying out previous CRISPR/Cas9 mediated genome editing, despite addition of steps to enrich for editing (sections 4.3.4 and 4.4.3.2). Therefore, within a population of Jurkat cells where no steps for enrichment of editing had taken place, the proportion of edited cells would be very low. If we were to generate clonal cell lines, this would ensure that each cell within the population was not expressing full-length MYB protein due to *MYB* being knocked out. This would enable us to gain greater insight as to the scale of the effect of *MYB* knockout, as all cells within the population would not contain MYB and would have the gene expression pattern of a cell lacking MYB.

Another method that could be used for generation of a CRISPR/Cas9 *MYB* knockout cell line, is through use of lentiviral integration into the genome. Cas9 could be stably integrated into the Jurkat genome, and upon selection for a clonal cell line this would result in all cells within this population expressing Cas9. Then, gRNAs targeting *MYB* for knockout could be added, either transiently to measure initial loss of MYB, or additionally integrated into the genome to measure stable *MYB* knockout within the cell line. Although this method is usually carried out for large scale CRISPR knockout screens, this could provide flexibility in how we measure *MYB* knockout in our own experiments (Koike-Yusa et al. 2014; Shalem et al. 2014).

If we were to pursue generating both CRISPR/Cas9 *MYB* knockout cell lines alongside our MYB-Halotag cell lines, this would enable us to compare stable versus transient loss of MYB within the Jurkat cell environment.

6.5.1.2 *MYB* knockout cell lines could further uncover MYB action within CBP-TAL1

After generation of these clonal lines, RT-qPCR should be repeated, similar to that done initially (Figure 6.3), however this should be done in triplicate.

To add to the RT-qPCR data showing the gene expression pattern after *MYB* knockout, western blotting could provide insight as to the protein levels within the cell. Western blotting for MYB would confirm whether full-length MYB protein was being expressed as a result of *MYB* knockout. In addition, western blotting for TAL1 would provide insight as to the activity of the *TAL1* enhancer and would complement *TAL1* RT-qPCR data.

It would also be interesting to carry out a CBP Co-IP in the *MYB* knockout lines, to investigate formation of the CBP-TAL1 complex (section 3.2.2). CBP is able to interact with each component of the CBP-TAL1 complex (sections 3.2.2 and 3.3), however should the CBP-TAL1 complex not be able to form upon *MYB* knockout, this would indicate that MYB initiation is critical to the formation of the complex. This would provide evidence to the hypothesis that it is MYB binding that recruits CBP to the complex, leading to the subsequent recruitment of the remaining proteins within the CBP-TAL1 complex (Mansour, Abraham, et al. 2014). Although, a CBP Co-IP would capture all CBP complexes throughout the cell, not just the CBP-TAL1 complex forming only at the *TAL1* enhancer.

6.5.2 Developing a MYB$_{Halotag}$ stable cell line for MYB knockdown experiments

6.5.2.1 Planning a MYB$_{Halotag}$ cell line using CRISPR/Cas9 and HDR

As we had established CRISPR/Cas9 and HDR methods that we have used previously to successfully tag CBP with a Halotag in Jurkat cells (Chapter 4), we were also well set up to generate cell lines with endogenously tagged MYB-Halotag. We had

previously designed and validated gRNAs that could not only target the MYB C-terminus, but yielded successful tagging at this locus (section 5.3). In addition we had the components of a HDR template, as we had a vector containing homology arms that successfully tagged the MYB C-terminus (section 5.4), and we had a Halotag sequence in the CBP_{FTHH} HDR vector that had yielded successfully tagged CBP with a Halotag (section 4.4.3.2). Therefore upon cloning of the Halotag into the MYB homology arms vector, we had all of the CRISPR/Cas9 reagents to carry out tagging of MYB with a Halotag.

6.5.2.2 Generating an endogenously tagged $MYB_{Halotag}$ cell line

We showed that we were able to successfully generate a heterozygous cell line, expressing full-length MYB-Halotag (section 6.3). There was some level of difficulty with the generation of this cell line because upon the first rounds of validation, through PCR and western blotting analysis we were able to show that four clones were expressing full-length MYB-Halotag. However, upon repetition of this screening using PCR, only one of the clones appeared to be heterozygous for the Halotag, where it had previously appeared homozygous. This implied that the addition of the Halotag to MYB was not favourable to the cell, and there was potential that all four of our clones had lost at least one copy of the Halotag from MYB.

It would have been helpful during the clonal generation process to take samples for RNA extraction and RT-qPCR analysis to examine expression of the Halotag in these cell lines. This could have provided further evidence on an mRNA expression level as to whether the Halotag was being expressed on MYB in these cell lines. This would be similar to the clonal validation RT-qPCR steps carried out for CBP and MYB tagging upon generating our cell lines for TA-CRAC previously (Figures 4.18 and 5.7), however due to time constraints we were unable to do this for $MYB_{Halotag}$ clones.

If we had time, we could have made the clonal cell lines again. We could use the polyclonal population, of which we had frozen stocks, and repeat the single cell sorting process. When we initially single cell sorted this polyclonal cell line, we did not use any fluorescent Halotag ligands to enrich for edited clines. We chose not to do this at the time as when we had generated our CBP_{FTHH} cell line, which had a Halotag,

we had labelled with TMR but we had observed large levels of background staining, giving rise to high levels of false positives. If we were to do this again, we could label with a lower concentration of TMR and label for less time, to reduce background staining. In addition, other fluorescent Halotag ligands are available, such as Janelia Fluor (Promega), which was developed to have higher fluorescence and give reduced background staining than TMR (Grimm et al. 2015). If we were to single cell sort our $MYB_{Halotag}$ polyclonal cells, when the cell population had been labelled with low concentrations of Janelia Fluor for low amounts of time, with very stringent gating, we could enrich for successfully Halotagged MYB and increase the number of clones obtained with the Halotag.

If we were to repeat this, we would aim for homozygous $MYB_{Halotag}$ cell line. The aim of this cell line was to determine how the cell copes with loss of MYB, so we therefore need to completely degrade MYB across the cell through each MYB allele having a Halotag. Although in this round of screening we did not obtain a homozygous clone, we have previously shown that tagging of MYB using these methods to generate a homozygous cell line is possible (section 5.5). In addition, a Jurkat MYB-dTAG cell line has been described in the literature and generated, with one homozygous being identified during clonal generation, thus showing that addition of larger tags onto MYB is possible (Yong et al. 2023). Therefore, with our added enrichment steps at the single cell sorting stage, this should be a valid strategy for achieving homozygous $MYB_{Halotag}$.

6.5.2.3   HaloPROTACs degradation of MYB-Halotag

With our generated heterozygous $MYB_{Halotag}$ cell line, we were still able to show some level of MYB degradation and consequent effects on gene expression within the cells.

Due to time constraints, we were not able to fully optimise the best conditions for MYB knockdown using the HaloPROTACs. From treating $MYB_{Halotag}$ cells with HaloPROTACs for 18, 24 and 48 hours, it appeared that 24 hours of treatment caused the greatest reduction in protein levels. However, as we did not do comparable experiments between time conditions, concentrations and outputs, it was difficult to get a complete picture for the optimal conditions.

To gain a full picture of the best conditions for MYB-Halotag knockdown using

242

HaloPROTACs, we would suggest carrying out a timecourse for the two different concentrations of HaloPROTACs used previously (300 nM, section 6.4.1 and 1 μM, section 6.4.2). This would include a range of times including 18, 24 and 48 hours, but an additional timepoint of 36 hours could be useful as we observed differences in our previous investigations into 24 and 48 hours (Figure 6.9). If we were to take samples at each timepoint for both concentrations, samples could be taken for both RT-qPCR and western blotting. This would allow us to gain a deeper insight into the gene expression environment within these cells and to analyse the levels of protein degradation observed. If we could blot for both MYB and Halotag at every timepoint it would enable us to confirm that Halotag and MYB were being degraded together. Overall, this experiment would then provide insight as to the best conditions for HaloPROTACs mediated knockdown of MYB-Halotag in Jurkat cells.

6.5.3 Perspectives for future HaloPROTACs induced degradation of MYB-Halotag experiments

After optimising knockdown conditions using the HaloPROTACs, there are many downstream experiments that can be carried out to examine the downstream consequences of loss of MYB within Jurkat cells.

A novel way to use the MYB-Halotag cell line would be to titrate in the concentration of HaloPROTACs used, so that not only could we observe the effect on the cell upon the loss of MYB, but as MYB degradation wears off due to the degradation of the HaloPROTACs molecules themselves, we could examine the effect reinitiation of MYB would have within Jurkat cells. If MYB degradation did abrogate *TAL1* enhancer function as we hypothesise, then through our timecourse validations we should be able to identify timepoints for when the enhancer is completely switched off, and then we may be able to identify timepoints for when the enhancer element is reinitiated. At these timepoints we could carry out various techniques that we have previously optimised within the lab for this system, used previously to examine the CBP-TAL1 complex and the *TAL1* enhancer in WT Jurkat cells.

Carrying out ChIP at these various timepoints may allow us to examine the temporal assembly of the CBP-TAL1 complex at the *TAL1* enhancer, for example we may be able to capture the order in which the proteins assemble at this locus. In

243

addition, we may be able to examine when the initiation of histone marks begins, and if this is reflective of the proteins bound, alongside eRNA transcription at this locus. To complement this, TT-seq at these various timepoints, would allow us to examine if upon MYB reinitiation there was a genome wide effect on eRNA transcription. RIP-qPCRs could also provide us with information as to the RNA binding activity of the CBP-TAL1 complex. For example, if the factors are only binding to the *TAL1* eRNAs when bound at the *TAL1* enhancer, or if there are eRNA-protein interactions away from this enhancer locus.

We could also combine our MYB degradation system with the addition of CBP inhibitors, such as A-485, which we have previously shown to have potency in targeting CBP and downstream gene expression. Inhibition of both proteins together could provide further information as to the targets of the CBP-TAL1 complex, and the action of both proteins in this system.

6.5.3.1 Dissecting the *TAL1* enhancer as a model for enhancer assembly

In essence, this will allow us to truly deconstruct and observe the rebuilding of the *TAL1* enhancer, which if we were to use as a model, could provide a wealth of information about enhancer elements as a whole. There have been various proposed mechanisms for how transcriptional machinery, including transcription factors and coactivators assemble at an enhancer to activate and upregulate enhancer function.

The "enhanceosome" model suggests a more exact and defined manner for enhancer assembly (Thanos and Maniatis 1995). It has been suggested that defined groups of transcription factors and coactivators assemble in particular orientations, with particular direct protein binding partners that bind in a specific order at the DNA level to cause enhancer activity (Guturu et al. 2013).

The "billboard" model for enhancer function, is a development of the enhanceosome model, but with a slight increase in flexibility (Arnosti and Kulkarni 2005). This billboard model suggests that enhancers evolutionarily conserve the number and type of transcription factor binding sites, however the orientation, the order and the binding motifs themselves may differ. There is a role suggested for indirect cooperativity of transcription factors, where two or more factors bind at the same site to influence enhancer activation, however there is no direct binding between the two proteins at

244

this locus. (Arnosti and Kulkarni 2005)

An additional model, known as the transcription factor collective model, is a lot less rigid than the previously suggested enhanceosome and billboard models (Junion et al. 2012). Through direct and indirect cooperativity between transcription factors, binding to transcription factor motifs, or being recruited by other cofactors, enhancer function can be initiated. As a result, transcription factor motif sequences, number and order are not well evolutionarily conserved (Junion et al. 2012).

Through use of the *TAL1* enhancer as a model, due to our ability to be able to degrade MYB and therefore manipulate the *TAL1* enhancer activity, through turning it off and reinitiating activity, we may be able to shed further light as to the action of transcription factors and coactivators at enhancer elements.

6.5.4   Dissecting the role of MYB at a model enhancer

In this chapter we have developed a Jurkat cell line containing MYB endogenously tagged with a Halotag. The aim of the work in this chapter was to create a model system for examination of regulatory complex function found at enhancers, and we have done this through generation of a degradation system for the initiating factor of the *TAL1* enhancer element. We have discussed future work for this cell line, in which it may be possible to dissect this model enhancer and gain deeper insight into the formation and function of regulatory portein complexes found at enhancers.

Chapter 7

Discussion

In this research project we have investigated the *TAL1* enhancer found within Jurkat cells as a model for enhancer function. This enhancer offers a unique opportunity for deconstruction of enhancer activity, due to it forming *de novo* within the Jurkat cell line (Mansour, Abraham, et al. 2014). This enhancer is formed surrounding an insertion for two MYB binding sites, where upon MYB binding a regulatory protein complex we have referred to as the CBP-TAL1 complex is able to form (Sanda et al. 2012; Mansour, Abraham, et al. 2014). Whilst the 3D genomic environment of the *TAL1* locus is the same in Jurkat cells compared to other human cell lines, such as HEK293T cells (Hnisz, Weintraub, et al. 2016), this enhancer exists only within this cell line. This enhancer has measurable features that can be used to characterise the activity of this regulatory element: for example target gene transcription of *TAL1* not only can be itself measured, but due to the cells within this line developing a dependency upon this enhancer and *TAL1* expression, cell viability and growth can be measured as a downstream effect of *TAL1* enhancer activity too (Mansour, Abraham, et al. 2014). In addition, the bidirectional nascent transcription of this enhancer element, producing both a sense and antisense enhancer RNA (eRNA) can be measured, to reflect the action and transcriptional activity of RNA Pol II at this locus (Danko et al. 2015; Lidschreiber et al. 2021).

We have begun to generate a toolkit using the *TAL1* enhancer as a model to examine different aspects of enhancer formation and function, with particular regard to the eRNAs transcribed from this regulatory element and the interplay with the

regulatory protein complex found at this enhancer.

7.1 The CBP-TAL1 protein complex acts at the *TAL1* enhancer

7.1.1 CBP may act with the TAL1 complex

At every locus of action for the TAL1 complex, where the TAL1 complex has been shown to bind, investigators have established that the histone acetyltransferase and transcriptional coactivator CBP is also present. CBP has been shown to bind not only at the *TAL1* enhancer, but in fact shares many of the same binding sites as MYB and the remainder of the TAL1 complex throughout Jurkat cells (Mansour, Abraham, et al. 2014). In addition, CBP has been shown to localise at the *ARID5B* locus and the *miR-223* locus with identical peaks to the TAL1 complex (Leong et al. 2017; Mansour, Sanda, et al. 2013). However, the co-localisation of CBP alongside this complex has been explained due to the role of CBP as a coactivator and as a marker of enhancers, and evidence for whether CBP is an integral part of this complex was lacking.

CBP is known to be a key marker of enhancer activity and has been implicated in binding at nearly all enhancers throughout the genome (Holmqvist and Mannervik 2013), so a counter argument could be that CBP is simply co-localising to carry out its role as a HAT at enhancers. However, it is worth considering that not only is HAT activity key for CBP, but it also plays a key role in scaffolding proteins, particularly with regard to scaffolding proteins at enhancers (Holmqvist and Mannervik 2013). In addition, CBP is known to bind to over 400 proteins, including TAL1 (Bedford et al. 2010) and has long-been established as a coactivator of MYB (Dai et al. 1996). Therefore, we believed there was a precedence to investigate whether CBP was binding together within the TAL1 complex, or whether CBP was simply co-localising. To do this we carried out a co-immunoprecipitation for CBP and suggested that CBP was also part of the TAL1 complex, due to CBP appearing to Co-IP each factor within the TAL1 complex (Section 3.2.2). This also further built on the established data surrounding the TAL1 complex, as to date the only protein-protein interactions that have been investigated were between MYB and TAL1 (Mansour, Abraham, et al.

2014). The remainder of the complex was assumed as interacting due to the high levels of co-localisation (Mansour, Abraham, et al. 2014). We therefore suggested that CBP should be included as a key part of this protein complex, as it could be possible that CBP is acting as a scaffold to hold together the complex at not only the *TAL1* enhancer, but at many other enhancers throughout the Jurkat genome (Section 3.2.2). We henceforth referred to this regulatory complex as CBP-TAL1.

7.1.1.1 Other methods of examining formation of the CBP-TAL1 complex

Co-IPs are not the only way to examine protein-protein interactions within cells. Cross-linking mass spectrometry allows proteins that interact within the cell environment to be cross-linked together, and after degradation of any non-cross-linked protein, mass-spectrometry can detect peptides of interacting proteins (reviewed by O'Reilly and Rappsilber 2018). This could therefore be used to analyse formation of a protein complex, such as the CBP-TAL1 complex. An alternative technique is offered through hybridisation-proximity (HyPro) labelling, which allows capture of proteins and RNAs that associate within a 20 nm radius of a particular protein or RNA sequence of interest (Yap, Chung, and Makeyev 2022). If this is followed by mass spectrometry this could allow examination of proteins interacting closely together within the space of the cell and could confirm that the CBP-TAL1 complex is forming in close proximity (Yap, Chung, and Makeyev 2022).

7.1.1.2 CBP inhibitors will further investigate the extent of the role of CBP within CBP-TAL1

After determining that CBP plays a key scaffolding role in the formation of the CBP-TAL1 complex, we sought to investigate an additional role of CBP: its acetyltransferase activity (Section 3.3). We inhibited CBP acetyltransferase activity, through addition of small molecule inhibitors and observed decreases in MYB expression, alongside marked decreases in expression of the *TAL1* eRNAs (Figure 3.5). This highlighted that CBP acetyltransferase activity was crucial for transcriptional activation at this regulatory element.

In addition as we have shown that in our system we can use the CBP HAT inhibitor A-485 to disrupt expression of the CBP-TAL1 complex components and activity at the

TAL1 enhancer, it would be interesting to carry out further experiments after CBP inhibition. Following HAT inhibition, repetition of CBP Co-IPs and additionally carrying out Co-IPs for other factors, such as MYB, would allow investigation of the requirement for CBP HAT activity for formation of the CBP-TAL1 complex. For example, without CBP HAT activity we could determine whether protein-protein interactions between the transcription factors are still able to occur. Or whether the impaired HAT activity may affect the ability of CBP to function as a scaffold. This could be the case as it has been shown that for MYB to function as a pioneer factor, it needs to be acetylated by CBP (Dai et al. 1996; Fuglerud, Ledsaak, et al. 2018). Furthermore, following A-485 treatment with ChIP would allow us to observe how HAT activity affects recruitment of the CBP-TAL1 complex to the *TAL1* enhancer.

As we have observed that upon treatment of A-485, expression of the *TAL1* eRNAs was reduced (Section 3.3), we could expand this to analyse eRNA expression on a more global scale, using TT-seq (Schwalb et al. 2016).

7.1.2 The CBP-TAL1 complex may bind to RNA

An emerging mode of action for transcription factors is to exhibit RNA binding behaviour, to further enable their ability to activate transcription. There are thought to be nearly 1600 transcription factor encoding genes, producing almost 3000 transcription factors throughout the human genome (ENCODE Project Consortium 2012; Wingender, Schoeps, and Dönitz 2013). Some human transcription factors were identified as having interactions with RNA many years ago, but these appeared to be specific cases and were not necessarily applied to transcription factors as a whole (Cassiday and Maher 2002). Recently, with the increased understanding of transcription occuring in the non-coding genome, examples of transcription factors binding to such non-coding RNAs have begun to emerge. For example, the transcription factor Yin-Yang 1 (YY1) was observed binding to eRNAs at enhancers, and in particular it was noted that binding of the eRNA retained the transcription factor at the enhancer, a behaviour termed "transcription factor trapping" (Sigova et al. 2015). An additional investigation showed that the transcription factor SOX2 directly bound to lncRNAs (Holmes et al. 2020). In addition, computational investigations into transcription factors that exhibit both DNA and RNA binding behaviours showed that the proportion

of classical DNA binding proteins that had RNA binding properties was high (Hudson and Ortlund 2014). Furthermore, chromatin associated proteins have been shown to be common RNA binding proteins (Hendrickson et al. 2016).

In our own investigations, we were able to indicate preliminary data that each of the transcription factors within the CBP-TAL1 complex, alongside CBP, may have some RNA binding activity within Jurkat cells, and may also bind to the *TAL1* eRNAs through native RIPs and qPCR (Section 3.4).

To complement our own data implying transcription factors binding to RNA, a recent preprint analysed transcription factor-RNA interactions on a more global scale (Oksuz et al. 2022). In K562 leukemia cells, 4SU labelling combined with UV crosslinking and mass spectrometry was used in RNA-binding region identification (RBR-ID, Oksuz et al. 2022). This showed that almost half of the transcription factors within this cell line bound to RNA, and of the transcription factors identified as binding to RNA; TAL1, MYB, RUNX1, HEB and E2A were all identified (Oksuz et al. 2022). To date, this is the only additional evidence for TAL1, MYB, HEB and E2A to exhibit RNA binding properties.

Taken together, this suggests an important role for RNA generally, but eRNAs in particular, in co-ordinating the function of the CBP-TAL1 complex. However, what this role is still remains to be clearly determined. Apart from studying the interactions with the CBP-TAL1 complex and RNA more globally (described later in more detail) further experiments could be used to investigate the requirement of the eRNAs within this protein complex. This could be done through utilising knockdown of eRNAs, for example using RNAi or antisense oligonucleotides (ASOs). However, this can be challenging, as due to the nature of eRNAs being short-lived, once they have been degraded through RNAi mechanisms, they can rapidly be transcribed to re-establish transcript levels. If we were to couple this with use in our system, in which Jurkat cells are very difficult to transfect and exhibit low transfection efficiency, RNAi does not appear to be a suitable method for knockdown of the eRNAs in these cells. In additional projects in our lab, we have preliminary data suggesting that use of an RNA guided RNA nuclease, such as the CRISPR/Cas13 system, can be an effective method of RNA degradation (Wessels et al. 2020). If we could integrate Cas13 stably into a Jurkat cell line, we could be able to achieve stable and prolonged knockdown of RNAs and this could then be applied to the *TAL1* eRNAs. This would then allow

us to follow up *TAL1* eRNA knockdown with our established Co-IPs and RT-qPCRs alongside ChIP-seq or ChIP-qPCR for the CBP-TAL1 complex components, allowing us to observe how CBP-TAL1 complex expression, formation and behaviour is affected through loss of the *TAL1* eRNAs.

7.2 Developing a system to study CBP-TAL1:RNA interactions

As we had implied that RNA, including the *TAL1* eRNAs, may be binding to each factor within the CBP-TAL1 complex (Section 3.4), we sought to develop a system whereby we would be able to study how protein components affect RNA binding. We chose to do this through tandem affinity crosslinking and analysis of cDNAs (TA-CRAC), where the RNA interactions of a protein of interest could be determined, only in the context of being bound to an additional protein of interest (Granneman, Kudla, et al. 2009; Thoms et al. 2015). TA-CRAC uses different combinations of affinity tags and therefore we generated a tandem tagged cell line, where CBP was tagged with a Flag-TEV-6xHis-Halotag (FTHH) and MYB was tagged with a 6xHis-TEV-HA/Strep (HTHAS) tag (Chapters 4 and 5). This final cell line was denoted MYB_{HTHAS}:CBP_{FTHH}.

7.2.1 Validation of our tagged MYB_{HTHAS}:CBP_{FTHH} cell line

We had begun to validate in our tagged cell line that the affinity tags were not negatively affecting protein function, in terms of protein-protein interactions and RNA binding. However, due to time constraints, we did not manage to fully validate everything.

7.2.1.1 The affinity tags could potentially have a stablising role

In our final MYB_{HTHAS}:CBP_{FTHH} cell line, using RT-qPCR it appeared that *MYB* gene expression was decreased compared to in WT cells (Section 5.5, Figure 5.7), however in our western blotting levels for full length protein appeared more pronounced (Figure 5.8). Our combined results could have suggested that higher levels of tagged

MYB in particular were available in the cell, and potentially as a consequence gene expression of *MYB* had been reduced. Carrying out ChIP-seq or ChIP-qPCR for MYB and the HA tag would provide further information as to whether there were changes to the chromatin binding behaviour of tagged MYB. It is interesting to speculate whether the tag has a compensatory effect on *TAL1* enhancer activity, and cell proliferation. RT-qPCR of TAL1 and the TAL1 eRNAs showed no significant differences in expression across the three cell lines, indicating that the addition of tags to MYB are not affecting enhancer activity (Figure 5.7). In addition, we have prepared TT-seq libraries to examine in greater detail the levels of eRNA expression in our MYB_{HTHAS}:CBP_{FTHH} cell line.

7.2.1.2 Limitations of current data

As our data are not perfect, for example we do not appear to have a good enough WT control in our western blotting for comparisons (Figure 5.8), this set of data should be replicated to gain a better understanding of what we are observing in this cell line. In addition TT-seq carried out globally throughout Jurkat cells could provide further information as to if the higher levels of protein expression of MYB are affecting the activity of the *TAL1* enhancer. We have not yet carried out a repeat of our CBP Co-IP carried out in section 3.2.2, in our final MYB_{HTHAS}:CBP_{FTHH} cell line to confirm that the CBP-TAL1 complex is still able to form with the addition of the affinity tags.

In addition, confirming that tagged CBP was still active as a HAT would confirm that another role of CBP was unaffected by the introduction of affinity tags.

7.2.2 TA-CRAC to investigate MYB-CBP:RNA interactions

7.2.2.1 Tandem tagged cell lines will allow more physiologically relevant analysis

It could be argued that as we had spent a large portion of time and resources on generating our MYB_{HTHAS}:CBP_{FTHH} cell line to carry out TA-CRAC, could we not simply transfect in constructs containing the tagged proteins to be able to carry out TA-CRAC? Although this would save us time, it would not allow us to gain a true representation as to the action of the CBP-TAL1 complex physiologically. With tran-

sient transfection, often the genes being expressed are under the control of strong promoters, meaning that the proteins expressed from the plasmids would likely be overexpressed compared to physiological levels. Therefore, it would be very difficult to be certain what was a true characteristic of these proteins at specific loci and with specific RNAs, or whether due to the overabundance of these proteins there was a higher chance of binding to RNA species with decreased affinity. We note that in our MYB_{HTHAS}:CBP_{FTHH} cell line we did not see upregulation of MYB gene expression that could be attritubuted to the tag (Figure 5.7), although we do observe a stronger band for full-length MYB protein in western blotting (Figure 5.8). We do not observe increased $TAL1$ expression in our MYB_{HTHAS}:CBP_{FTHH} cell line, indicating that $TAL1$ enhancer activity is not increased. We have prepared TT-seq libraries in WT Jurkats and in MYB_{HTHAS}:CBP_{FTHH} cells, and this data will allow us to examine $TAL1$ enhancer activity through any changes to the expression of the $TAL1$ eRNAs in our tagged cell line.

7.2.2.2 TA-CRAC can identify nucleotide resolution of RNA-protein interactions

Once this cell line has been fully validated, we will be able to begin to carry out the TA-CRAC analysis. This will provide global RNA binding maps to show how MYB is binding to RNA, when bound to CBP. As we have shown that individually both MYB and CBP can bind to the $TAL1$ eRNAs (Section 3.4), it will be interesting if the TA-CRAC analysis provides more detail as to this interaction. Not only can TA-CRAC produce information on which species of RNA are binding to a particular protein, but we can also gain insight into which exact RNA nucleotides are binding to specific sequences within the protein (Granneman, Kudla, et al. 2009; Thoms et al. 2015).

This would allow us to gain an idea as to what domain or region of MYB is binding to RNA and we can also determine whether there are any recurrent structural features of the regions of RNA bound. For example, after identification of bound RNAs, these nucleotide sequences can be fed into secondary structure prediction softwares such as RNA structure (Reuter and Mathews 2010) and more recent tools that use machine learning such as MXFold2 (Sato, Akiyama, and Sakakibara 2021). We can also use

experimental approaches to determine structures of RNA species of interest. For example, selective 2'-hydroxyl acylation analysed by primer extension coupled with mutational profiling (SHAPE-Map) can be done within cells to obtain RNA secondary structure conformations found within the cell environment (Smola, Rice, et al. 2015; Smola and Weeks 2018). Upon determining the secondary structure of an RNA, we could then begin to understand which structures are binding to the protein of interest, in our case MYB, and we could get a more visual idea of how these proteins are interacting.

For example, if we are able to determine how an RNA is binding with MYB-CBP through TA-CRAC, if the sequences bound from the RNA are different and spatially far apart in the secondary structure predictions of the RNA, it could point to a scaffolding role for the RNA within the CBP-TAL1 complex. lncRNAs have previously been implicated in scaffolding protein complexes, so it would be interesting if we could gain specific examples of this within the CBP-TAL1 complex (Ribeiro et al. 2018).

7.3 Deconstructing *TAL1* enhancer function through MYB degradation

We have started generating a system to enable titrated knockdown of MYB, particularly through use of the Halotag and HaloPROTACs system (Los and K. Wood 2007; Buckley et al. 2015). Currently we have developed a heterozygous tagged $MYB_{Halotag}$ cell line, which although could be used to begin to investigate MYB depletion in Jurkat cells, it is likely that a homozygous $MYB_{Halotag}$ cell line is required as this will be the only way to achieve complete degradation. Therefore, to take this line of investigation further, it would be advisable to remake this cell line, to obtain a homozygous tagged MYB. Upon generation of this cell line, there then exists a wealth of experiments that can be carried out to determine the action of MYB at the *TAL1* enhancer and within the CBP-TAL1 complex.

In order to determine the action of MYB within the CBP-TAL1 complex, depletion of MYB could be followed by CBP and MYB Co-IPs to examine the formation of the CBP-TAL1 complex without MYB. In addition, if ChIP was carried out not only at the

TAL1 enhancer, where recruitment of the remainder of the CBP-TAL1 complex should be MYB dependent, but genome-wide using ChIP-seq, this could provide information as to whether MYB does play any role in initiating the complex at other loci, or whether it is just at the *TAL1* enhancer in particular.

7.3.1 The Halotag offers more than just depletion

Degradation tags and inducible systems are very popular, with a wide variety of such systems available (section 6.1.2). However, an advantage offered by the Halotag is that it is not just a degradation tag, there are many further ligands available that we could utilise. For example, through addition of fluorescent Halotag ligands such as TMR and Janelia Fluor; a brighter ligand (Grimm et al. 2015), we could observe the dynamic behaviour and chromatin association of MYB within Jurkat cells through live cell imaging.

7.3.2 Investigating MYB depletion at the *TAL1* enhancer

Through depleting MYB we should be able to markedly reduce *TAL1* enhancer function, potentially to abrogate enhancer function entirely. There is possibility then that this enhancer element will revert to a more transcriptionally silent and closed chromatin form, for example with repressive chromatin modifications deposited by factors such as Polycomb group (PcG) proteins (as reviewed by Di Croce and Helin 2013). Chromatin silencing can be targeted at enhancers through use of the CRISPR interference (CRISPRi) system, where a Kruppel associated box (KRAB) domain which represses transcription through recruitment of transcriptionally repressive machinery, is fused to a catalytically dead Cas9, which can guide the KRAB to a specific site (Gilbert et al. 2014). CRISPRi has been used to induce transcriptional repression at targeted enhancers, for example in K562 leukemia cells, near complete loss of H3K27ac and activating TF binding was observed (J. Huang et al. 2018). In addition, when the *TAL1* enhancer was first identified, the MYB binding site was deleted in Jurkat cells to observe the effect on the cells (Mansour, Abraham, et al. 2014). This caused a marked reduction in H3K27ac levels and MYB binding, indicating that this decreased the activating transcriptional activity of this region (Mansour, Abraham, et al. 2014).

Therefore, it is likely that loss of MYB would have a similar effect on the chromatin environment.

Through the titration of the HaloPROTACs, there is the possibility that should we establish the correct timings, we can carry out many of our experiments to capture this enhancer at the stage when it is more transcriptionally silent, and progress through to transcriptional activation. For example, carrying out ChIP at various timepoints could enable us to determine temporally how the enhancer is initiated, for example if MYB binding preceeds deposition of activating histone modifications such as H3K27ac. In addition, it could provide further evidence that CBP is functioning as a scaffold, as it may highlight that CBP is required to bind after being recruited by MYB, and then enables sequestering of the remainder of the CBP-TAL1 complex. It could also provide information as to when RNA Pol II is recruited, and potentially develop our understanding of which factors recruit RNA Pol II.

If we were able to also examine the *TAL1* promoter, the target gene of the *TAL1* enhancer, we may also be able to establish at what point the enhancer-promoter interactions are established, and whether particular factors are required for this to occur.

Carrying out RT-qPCR for the eRNAs transcribed at the *TAL1* enhancer, or TT-seq, could provide information as to when the eRNAs begin to be transcribed, and coupled with our potential ChIP-seq data could provide insight as to at what stage eRNA transcription occurs and what factors are required to bind at enhancers for transcription to be able to occur.

7.3.2.1 MYB depletion could be coupled with CBP HAT inhibition

If we were to not only deplete MYB across Jurkat cells, but in addition inhibit the acetyltransferase activity through addition of CBP HAT inhibitor A-485 (Section 3.3), this could provide further evidence as to the MYB-CBP interactions. MYB has been shown to be a pioneer factor, and it has been shown that acetylation of MYB by CBP is essential for pioneer factor activity (Fuglerud, Lemma, et al. 2017; Fuglerud, Ledsaak, et al. 2018). Therefore, if we were to inhibit this acetylation of MYB by CBP, it could inhibit formation of the enhancer and therefore provide further evidence as to the reliance of not only MYB, but also CBP on this enhancer element.

7.4   Summary

Our investigations have begun to deconstruct further the mode of action of the *TAL1* enhancer in Jurkat cells, which can be used as a model for enhancer function. We have built upon previous knowledge to suggest that CBP plays a role within the regulatory protein complex found at this enhancer, and have begun to generate a system to study how specific combinations of proteins found within the CBP-TAL1 complex are binding to RNA, with the hope this will shed light as to the action of eRNAs. MYB depletion experiments should allow further examination of the formation of enhancers and the role of MYB within this leukemia cell line.

Overall, our work within this research project has allowed us to generate a toolkit to be able to deconstruct and uncover how this model enhancer is regulated, how it functions and how regulatory protein complexes and eRNAs work at this regulatory element. We hope that with our future work we will be able to gain further understanding as to the role eRNAs play within regulatory protein complexes, and we also hope that our work will uncover further evidence as to how enhancers form and function.

Bibliography

Abudayyeh, Omar O, Jonathan S Gootenberg, Patrick Essletzbichler, et al. (Oct. 2017). "RNA targeting with CRISPR-Cas13". en. In: *Nature* 550.7675, pp. 280–284.

Abudayyeh, Omar O, Jonathan S Gootenberg, Silvana Konermann, et al. (Aug. 2016). "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector". en. In: *Science* 353.6299, aaf5573.

Aifantis, Iannis, Elizabeth Raetz, and Silvia Buonamici (May 2008). "Molecular pathogenesis of T-cell leukaemia and lymphoma". en. In: *Nat. Rev. Immunol.* 8.5, pp. 380–390.

Allfrey, V G, R Faulkner, and A E Mirsky (May 1964). "Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 51.5, pp. 786–794.

Anderson, L and J Seilhamer (1997). "A comparison of selected mRNA and protein abundances in human liver". en. In: *Electrophoresis* 18.3-4, pp. 533–537.

Anfossi, G, A M Gewirtz, and B Calabretta (May 1989). "An oligomer complementary to c-myb-encoded mRNA inhibits proliferation of human myeloid leukemia cell lines". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 86.9, pp. 3379–3383.

Arnold, Cosmas D et al. (Mar. 2013). "Genome-wide quantitative enhancer activity maps identified by STARR-seq". en. In: *Science* 339.6123, pp. 1074–1077.

Arnosti, David N and Meghana M Kulkarni (Apr. 2005). "Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards?" en. In: *J. Cell. Biochem.* 94.5, pp. 890–898.

Baba, Atsushi et al. (June 2011). "PKA-dependent regulation of the histone lysine demethylase complex PHF2-ARID5B". en. In: *Nat. Cell Biol.* 13.6, pp. 668–675.

Balsalobre, Aurelio and Jacques Drouin (July 2022). "Pioneer factors as master regulators of the epigenome and cell fate". en. In: *Nat. Rev. Mol. Cell Biol.* 23.7, pp. 449–464.

Banani, Salman F et al. (July 2016). "Compositional Control of Phase-Separated Cellular Bodies". en. In: *Cell* 166.3, pp. 651–663.

Banerji, J, S Rusconi, and W Schaffner (Dec. 1981). "Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences". en. In: *Cell* 27.2 Pt 1, pp. 299–308.

Bannister, A J and T Kouzarides (Dec. 1996). "The CBP co-activator is a histone acetyltransferase". In: *Nature* 384.6610, pp. 641–643.

Bannister, Andrew J and Tony Kouzarides (Mar. 2011). "Regulation of chromatin by histone modifications". en. In: *Cell Res.* 21.3, pp. 381–395.

Bedford, David C et al. (Jan. 2010). "Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases". en. In: *Epigenetics* 5.1, pp. 9–15.

Begley, C G et al. (Mar. 1989). "Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 86.6, pp. 2031–2035.

Belton, Jon-Matthew et al. (Nov. 2012). "Hi-C: a comprehensive technique to capture the conformation of genomes". en. In: *Methods* 58.3, pp. 268–276.

Bernard, O et al. (Jan. 1990). "Two distinct mechanisms for the SCL gene activation in the t(1;14) translocation of T-cell leukemias". en. In: *Genes Chromosomes Cancer* 1.3, pp. 194–208.

Bernstein, E et al. (Jan. 2001). "Role for a bidentate ribonuclease in the initiation step of RNA interference". en. In: *Nature* 409.6818, pp. 363–366.

Bhatia, Shipra et al. (Nov. 2021). "Quantitative spatial and temporal assessment of regulatory element activity in zebrafish". en. In: *Elife* 10.

Biedenkapp, H et al. (Oct. 1988). "Viral myb oncogene encodes a sequence-specific DNA-binding activity". en. In: *Nature* 335.6193, pp. 835–837.

Black, Joshua C, Capucine Van Rechem, and Johnathan R Whetstine (Nov. 2012). "Histone lysine methylation dynamics: establishment, regulation, and biological impact". en. In: *Mol. Cell* 48.4, pp. 491–507.

Bonn, Stefan et al. (Jan. 2012). "Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development". en. In: *Nat. Genet.* 44.2, pp. 148–156.

Borowicz, Paweł et al. (Apr. 2020). "A simple and efficient workflow for generation of knock-in mutations in Jurkat T cells using CRISPR/Cas9". en. In: *Scand. J. Immunol.* 91.4, e12862.

Bose, Daniel A et al. (Jan. 2017). "RNA Binding to CBP Stimulates Histone Acetylation and Transcription". en. In: *Cell* 168.1-2, 135–149.e22.

Bowers, Erin M et al. (May 2010). "Virtual ligand screening of the p300/CBP histone acetyltransferase: identification of a selective small molecule inhibitor". en. In: *Chem. Biol.* 17.5, pp. 471–482.

Brangwynne, Clifford P et al. (June 2009). "Germline P granules are liquid droplets that localize by controlled dissolution/condensation". en. In: *Science* 324.5935, pp. 1729–1732.

Brenner, S, F Jacob, and M Meselson (May 1961). "An unstable intermediate carrying information from genes to ribosomes for protein synthesis". en. In: *Nature* 190, pp. 576–581.

Brinkman, Eva K, Tao Chen, et al. (Dec. 2014). "Easy quantitative assessment of genome editing by sequence trace decomposition". en. In: *Nucleic Acids Res.* 42.22, e168.

Brinkman, Eva K, Arne N Kousholt, et al. (June 2018). "Easy quantification of template-directed CRISPR/Cas9 editing". en. In: *Nucleic Acids Res.* 46.10, e58.

Buckley, Dennis L et al. (Aug. 2015). "HaloPROTACS: Use of Small Molecule PROTACs to Induce Degradation of HaloTag Fusion Proteins". en. In: *ACS Chem. Biol.* 10.8, pp. 1831–1837.

Buecker, Christa and Joanna Wysocka (June 2012). "Enhancers as information integration hubs in development: lessons from genomics". en. In: *Trends Genet.* 28.6, pp. 276–284.

Buenrostro, Jason D et al. (Jan. 2015). "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide". en. In: *Curr. Protoc. Mol. Biol.* 109, pp. 21.29.1–21.29.9.

Bujnicki, T et al. (Apr. 2012). "Inhibition of Myb-dependent gene expression by the sesquiterpene lactone mexicanin-I". en. In: *Leukemia* 26.4, pp. 615–622.

Calo, Eliezer and Joanna Wysocka (Mar. 2013). "Modification of enhancer chromatin: what, how, and why?" en. In: *Mol. Cell* 49.5, pp. 825–837.

Capecchi, M R (Oct. 2001). "Generating mice with targeted mutations". en. In: *Nat. Med.* 7.10, pp. 1086–1090.

Cassiday, Laura A and L James Maher 3rd (Oct. 2002). "Having it both ways: transcription factors that bind DNA and RNA". en. In: *Nucleic Acids Res.* 30.19, pp. 4118–4126.

Castello, Alfredo et al. (Aug. 2016). "Comprehensive Identification of RNA-Binding Domains in Human Cells". en. In: *Mol. Cell* 63.4, pp. 696–710.

Celadova, Petra (Mar. 2022). "Saturating mutagenesis screen to dissect enhancer RNA function". PhD thesis. University of Sheffield.

Chan, H M and N B La Thangue (July 2001). "p300/CBP proteins: HATs for transcriptional bridges and scaffolds". en. In: *J. Cell Sci.* 114.Pt 13, pp. 2363–2373.

Chatterjee, Pranam, Noah Jakimo, and Joseph M Jacobson (Oct. 2018). "Minimal PAM specificity of a highly similar SpCas9 ortholog". en. In: *Sci Adv* 4.10, eaau0766.

Chatterjee, Pranam, Noah Jakimo, Jooyoung Lee, et al. (Oct. 2020). "An engineered ScCas9 with broad PAM range and high specificity and activity". en. In: *Nat. Biotechnol.* 38.10, pp. 1154–1158.

Chen, Q et al. (Feb. 1990). "The tal gene undergoes chromosome translocation in T cell leukemia and potentially encodes a helix-loop-helix protein". en. In: *EMBO J.* 9.2, pp. 415–424.

Conant, David et al. (Feb. 2022). "Inference of CRISPR Edits from Sanger Trace Data". en. In: *CRISPR J* 5.1, pp. 123–130.

Condorelli, G L et al. (Nov. 1996). "T-Cell-directed TAL-1 Expression Induces T-Cell Malignancies in Transgenic Mice1". en. In: *Cancer Res.* 56.22, pp. 5113–5119.

Core, Leighton J, Joshua J Waterfall, and John T Lis (Dec. 2008). "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters". en. In: *Science* 322.5909, pp. 1845–1848.

Creyghton, Menno P et al. (Dec. 2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 107.50, pp. 21931–21936.

Crick, F (Aug. 1970). "Central dogma of molecular biology". en. In: *Nature* 227.5258, pp. 561–563.

Cruz-Molina, Sara et al. (May 2017). "PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation". en. In: *Cell Stem Cell* 20.5, 689–705.e9.

Dai, P et al. (Mar. 1996). "CBP as a transcriptional coactivator of c-Myb". en. In: *Genes Dev.* 10.5, pp. 528–540.

Danko, Charles G et al. (May 2015). "Identification of active transcriptional regulatory elements from GRO-seq data". en. In: *Nat. Methods* 12.5, pp. 433–438.

De Braekeleer, Etienne et al. (Jan. 2011). "RUNX1 translocations and fusion genes in malignant hemopathies". en. In: *Future Oncol.* 7.1, pp. 77–91.

De Santa, Francesca et al. (May 2010). "A large fraction of extragenic RNA pol II transcription sites overlap enhancers". en. In: *PLoS Biol.* 8.5, e1000384.

Decker, Carolyn J et al. (May 2022). "RNA is required for the integrity of multiple nuclear and cytoplasmic membrane-less RNP granules". en. In: *EMBO J.* 41.9, e110137.

Dekker, Job et al. (Feb. 2002). "Capturing chromosome conformation". en. In: *Science* 295.5558, pp. 1306–1311.

Di Croce, Luciano and Kristian Helin (Oct. 2013). "Transcriptional regulation by Polycomb group proteins". en. In: *Nat. Struct. Mol. Biol.* 20.10, pp. 1147–1155.

Dion, Michael F et al. (Apr. 2005). "Genomic characterization reveals a simple histone H4 acetylation code". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 102.15, pp. 5501–5506.

Dixon, Jesse R et al. (Apr. 2012). "Topological domains in mammalian genomes identified by analysis of chromatin interactions". en. In: *Nature* 485.7398, pp. 376–380.

Doench, John G et al. (Feb. 2016). "Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9". en. In: *Nat. Biotechnol.* 34.2, pp. 184–191.

Dostie, Josée et al. (Oct. 2006). "Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements". en. In: *Genome Res.* 16.10, pp. 1299–1309.

Dowen, Jill M et al. (Oct. 2014). "Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes". en. In: *Cell* 159.2, pp. 374–387.

Elbashir, S M et al. (May 2001). "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells". en. In: *Nature* 411.6836, pp. 494–498.

ENCODE Project Consortium (Sept. 2012). "An integrated encyclopedia of DNA elements in the human genome". en. In: *Nature* 489.7414, pp. 57–74.

Engel, I et al. (2001). "Early thymocyte development is regulated by modulation of E2A protein activity". In: *J. Exp. Med.* 194.6, pp. 733–745.

Ewing, Adam D and Haig H Kazazian Jr (June 2011). "Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans". en. In: *Genome Res.* 21.6, pp. 985–990.

Fuglerud, Bettina M, Marit Ledsaak, et al. (June 2018). "The pioneer factor activity of c-Myb involves recruitment of p300 and induction of histone acetylation followed by acetylation-induced chromatin dissociation". en. In: *Epigenetics Chromatin* 11.1, p. 35.

Fuglerud, Bettina M, Roza B Lemma, et al. (July 2017). "A c-Myb mutant causes deregulated differentiation due to impaired histone binding and abrogated pioneer factor function". en. In: *Nucleic Acids Res.* 45.13, pp. 7681–7696.

Furlong, Eileen E M and Michael Levine (Sept. 2018). "Developmental enhancers and chromosome topology". en. In: *Science* 361.6409, pp. 1341–1345.

García-González, Estela et al. (Aug. 2016). "Chromatin remodeling effects on enhancer activity". en. In: *Cell. Mol. Life Sci.* 73.15, pp. 2897–2910.

Ghoneim, Mohamed, Harrison A Fuchs, and Catherine A Musselman (July 2021). "Histone Tail Conformations: A Fuzzy Affair with DNA". en. In: *Trends Biochem. Sci.* 46.7, pp. 564–578.

Giaimo, Benedetto Daniele et al. (June 2019). "The histone variant H2A.Z in gene regulation". en. In: *Epigenetics Chromatin* 12.1, p. 37.

Gilbert, Luke A et al. (Oct. 2014). "Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation". en. In: *Cell* 159.3, pp. 647–661.

Golub, T R et al. (May 1995). "Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 92.11, pp. 4917–4921.

Granneman, Sander, Grzegorz Kudla, et al. (June 2009). "Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 106.24, pp. 9613–9618.

Granneman, Sander, Elisabeth Petfalski, and David Tollervey (Aug. 2011). "A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease". en. In: *EMBO J.* 30.19, pp. 4006–4019.

Greenberg, Maxim V C and Deborah Bourc'his (Oct. 2019). "The diverse roles of DNA methylation in mammalian development and disease". en. In: *Nat. Rev. Mol. Cell Biol.* 20.10, pp. 590–607.

Grimm, Jonathan B et al. (Mar. 2015). "A general method to improve fluorophores for live-cell and single-molecule microscopy". en. In: *Nat. Methods* 12.3, 244–50, 3 p following 250.

Gross, D S and W T Garrard (1988). "Nuclease hypersensitive sites in chromatin". en. In: *Annu. Rev. Biochem.* 57, pp. 159–197.

Gu, W and R G Roeder (Aug. 1997). "Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain". en. In: *Cell* 90.4, pp. 595–606.

Gurumurthy, Aishwarya et al. (Feb. 2021). "Super-enhancer mediated regulation of adult β-globin gene expression: the role of eRNA and Integrator". en. In: *Nucleic Acids Res.* 49.3, pp. 1383–1396.

Guturu, Harendra et al. (Dec. 2013). "Structure-aided prediction of mammalian transcription factor complexes in conserved non-coding elements". en. In: *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368.1632, p. 20130029.

Haeussler, Maximilian et al. (July 2016). "Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR". en. In: *Genome Biol.* 17.1, p. 148.

Hafner, Markus et al. (Apr. 2010). "Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP". en. In: *Cell* 141.1, pp. 129–141.

Hall, Michael A et al. (Feb. 2009). "High-resolution dynamic mapping of histone-DNA interactions in a nucleosome". en. In: *Nat. Struct. Mol. Biol.* 16.2, pp. 124–129.

Hannah, Rita R, Martha L Jennens-Clough, and Keith V Wood (1998). "Rapid Luciferase Reporter Assay Systems for High Throughput Studies[No title]". In: *Promega Notes*.

Harrison, Laura J and Daniel Bose (Aug. 2022). "Enhancer RNAs step forward: new insights into enhancer function". en. In: *Development* 149.16.

Harrow, Jennifer et al. (Sept. 2012). "GENCODE: the reference human genome annotation for The ENCODE Project". en. In: *Genome Res.* 22.9, pp. 1760–1774.

He, Chongsheng et al. (Oct. 2016). "High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells". en. In: *Mol. Cell* 64.2, pp. 416–430.

Heintzman, Nathaniel D et al. (Mar. 2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome". en. In: *Nat. Genet.* 39.3, pp. 311–318.

Heinz, Sven et al. (Mar. 2015). "The selection and function of cell type-specific enhancers". In: *Nat. Rev. Mol. Cell Biol.* 16.3, pp. 144–154.

Hendrickson, David et al. (Feb. 2016). "Widespread RNA binding by chromatin-associated proteins". en. In: *Genome Biol.* 17, p. 28.

Henikoff, Steven et al. (Nov. 2020). "Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation". en. In: *Elife* 9.

Henninger, Jonathan E et al. (Jan. 2021). "RNA-Mediated Feedback Control of Transcriptional Condensates". en. In: *Cell* 184.1, 207–225.e24.

Henry, Ryan A, Yin-Ming Kuo, and Andrew J Andrews (Aug. 2013). "Differences in specificity and selectivity between CBP and p300 acetylation of histone H3 and H3/H4". en. In: *Biochemistry* 52.34, pp. 5746–5759.

Hnisz, Denes, Brian J Abraham, et al. (Nov. 2013). "Super-enhancers in the control of cell identity and disease". en. In: *Cell* 155.4, pp. 934–947.

Hnisz, Denes, Krishna Shrinivas, et al. (Mar. 2017). "A Phase Separation Model for Transcriptional Control". en. In: *Cell* 169.1, pp. 13–23.

Hnisz, Denes, Abraham S Weintraub, et al. (Mar. 2016). "Activation of proto-oncogenes by disruption of chromosome neighborhoods". In: *Science* 351.6280, pp. 1454–1458.

Hochuli, E et al. (Nov. 1988). "Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent". en. In: *Biotechnology* 6.11, pp. 1321–1325.

Hoelzer, Dieter et al. (June 2002). "Outcome of adult patients with T-lymphoblastic lymphoma treated according to protocols for acute lymphoblastic leukemia". en. In: *Blood* 99.12, pp. 4379–4385.

Holmes, Zachariah E et al. (Apr. 2020). "The Sox2 transcription factor binds RNA". en. In: *Nat. Commun.* 11.1, p. 1805.

Holmqvist, Per-Henrik and Mattias Mannervik (Jan. 2013). "Genomic occupancy of the transcriptional co-activators p300 and CBP". In: *Transcription* 4.1, pp. 18–23.

Hsieh, Chen-Lin et al. (May 2014). "Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 111.20, pp. 7319–7324.

Hsu, H L, J T Cheng, et al. (June 1991). "Enhancer-binding activity of the tal-1 oncoprotein in association with the E47/E12 helix-loop-helix proteins". en. In: *Mol. Cell. Biol.* 11.6, pp. 3037–3042.

Hsu, H L, I Wadman, and R Baer (Apr. 1994). "Formation of in vivo complexes between the TAL1 and E2A polypeptides of leukemic T cells". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 91.8, pp. 3181–3185.

Huang, Jialiang et al. (Mar. 2018). "Dissecting super-enhancer hierarchy based on chromatin interactions". en. In: *Nat. Commun.* 9.1, p. 943.

Hudson, William H and Eric A Ortlund (Nov. 2014). "The structure, function and evolution of proteins that bind DNA and RNA". en. In: *Nat. Rev. Mol. Cell Biol.* 15.11, pp. 749–760.

Hyun, Kwangbeom et al. (Apr. 2017). "Writing, erasing and reading histone lysine methylations". en. In: *Exp. Mol. Med.* 49.4, e324.

Jacob, F and J Monod (June 1961). "Genetic regulatory mechanisms in the synthesis of proteins". en. In: *J. Mol. Biol.* 3, pp. 318–356.

Jin, Chunyuan and Gary Felsenfeld (June 2007). "Nucleosome stability mediated by histone variants H3.3 and H2A.Z". en. In: *Genes Dev.* 21.12, pp. 1519–1529.

Jin, Qihuang et al. (Jan. 2011). "Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation". en. In: *EMBO J.* 30.2, pp. 249–262.

Jinek, Martin et al. (Aug. 2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity". en. In: *Science* 337.6096, pp. 816–821.

Johnson, David S et al. (June 2007). "Genome-wide mapping of in vivo protein-DNA interactions". en. In: *Science* 316.5830, pp. 1497–1502.

Joshi, Preeti et al. (2013). "The functional interactome landscape of the human histone deacetylase family". en. In: *Mol. Syst. Biol.* 9, p. 672.

Joy, Stephen T et al. (Sept. 2021). "A dual-site inhibitor of CBP/p300 KIX is a selective and effective modulator of myb". en. In: *J. Am. Chem. Soc.* jacs.1c04432.

Junion, Guillaume et al. (Feb. 2012). "A transcription factor collective defines cardiac cell fate and reflects lineage history". en. In: *Cell* 148.3, pp. 473–486.

Kagey, Michael H et al. (Sept. 2010). "Mediator and cohesin connect gene expression and chromatin architecture". en. In: *Nature* 467.7314, pp. 430–435.

Kaikkonen, Minna U et al. (Aug. 2013). "Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription". en. In: *Mol. Cell* 51.3, pp. 310–325.

Kalkhoven, Eric (Sept. 2004). "CBP and p300: HATs for different occasions". en. In: *Biochem. Pharmacol.* 68.6, pp. 1145–1155.

Karrman, Kristina and Bertil Johansson (Feb. 2017). "Pediatric T-cell acute lymphoblastic leukemia". en. In: *Genes Chromosomes Cancer* 56.2, pp. 89–116.

Katsumura, Koichi R, Emery H Bresnick, and GATA Factor Mechanisms Group (Apr. 2017). "The GATA factor revolution in hematology". en. In: *Blood* 129.15, pp. 2092–2102.

Kim, Daesik, Beum-Chang Kang, and Jin-Soo Kim (Feb. 2021). "Identifying genome-wide off-target sites of CRISPR RNA-guided nucleases and deaminases with Digenome-seq". en. In: *Nat. Protoc.* 16.2, pp. 1170–1192.

Kim, Tae Hoon and Bing Ren (2006). "Genome-wide analysis of protein-DNA interactions". en. In: *Annu. Rev. Genomics Hum. Genet.* 7, pp. 81–102.

Kim, Tae-Kyung et al. (May 2010). "Widespread transcription at neuronal activity-regulated enhancers". en. In: *Nature* 465.7295, pp. 182–187.

Kleinstiver, Benjamin P, Vikram Pattanayak, et al. (Jan. 2016). "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects". en. In: *Nature* 529.7587, pp. 490–495.

Kleinstiver, Benjamin P, Michelle S Prew, et al. (July 2015). "Engineered CRISPR-Cas9 nucleases with altered PAM specificities". en. In: *Nature* 523.7561, pp. 481–485.

Klempnauer, K H and J M Bishop (Apr. 1984). "Neoplastic transformation by E26 leukemia virus is mediated by a single protein containing domains of gag and myb genes". en. In: *J. Virol.* 50.1, pp. 280–283.

Ko, L J et al. (May 1991). "Murine and human T-lymphocyte GATA-3 factors mediate transcription through a cis-regulatory element within the human T-cell receptor delta gene enhancer". en. In: *Mol. Cell. Biol.* 11.5, pp. 2778–2784.

Koike-Yusa, Hiroko et al. (Mar. 2014). "Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library". en. In: *Nat. Biotechnol.* 32.3, pp. 267–273.

Kung, A L et al. (Feb. 2000). "Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP". en. In: *Genes Dev.* 14.3, pp. 272–277.

Lahortiga, Idoya et al. (May 2007). "Duplication of the MYB oncogene in T cell acute lymphoblastic leukemia". en. In: *Nat. Genet.* 39.5, pp. 593–595.

Lai, William K M and B Franklin Pugh (Sept. 2017). "Understanding nucleosome dynamics and their links to gene expression and DNA replication". en. In: *Nat. Rev. Mol. Cell Biol.* 18.9, pp. 548–562.

Lambert, Samuel A et al. (Feb. 2018). "The Human Transcription Factors". en. In: *Cell* 172.4, pp. 650–665.

Lasko, Loren M et al. (Oct. 2017). "Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours". en. In: *Nature* 550.7674, pp. 128–132.

Lee, Joo-Hyung et al. (Aug. 2021). "Enhancer RNA m6A methylation facilitates transcriptional condensate formation and gene activation". en. In: *Mol. Cell.*

Lemma, Roza B et al. (Apr. 2021). "Chromatin occupancy and target genes of the haematopoietic master transcription factor MYB". en. In: *Sci. Rep.* 11.1, p. 9008.

Leong, Wei Zhong et al. (Dec. 2017). "ARID5B as a critical downstream target of the TAL1 complex that activates the oncogenic transcriptional program and promotes T-cell leukemogenesis". en. In: *Genes Dev.* 31.23-24, pp. 2343–2360.

Lettice, Laura A et al. (July 2003). "A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly". en. In: *Hum. Mol. Genet.* 12.14, pp. 1725–1735.

Li, Guoliang et al. (Feb. 2010). "ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing". en. In: *Genome Biol.* 11.2, R22.

Li, Wenbo, Dimple Notani, and Michael G Rosenfeld (Apr. 2016). "Enhancers as non-coding RNA transcription units: recent insights and future perspectives". en. In: *Nat. Rev. Genet.* 17.4, pp. 207–223.

Liang, Xiquan et al. (Jan. 2017). "Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA". en. In: *J. Biotechnol.* 241, pp. 136–146.

Lidschreiber, Katja et al. (Jan. 2021). "Transcriptionally active enhancers in human cancer cells". en. In: *Mol. Syst. Biol.* 17.1, e9873.

Liu, Jidong et al. (Sept. 2004). "Argonaute2 is the catalytic engine of mammalian RNAi". en. In: *Science* 305.5689, pp. 1437–1441.

Liu, Mingjun et al. (Oct. 2021). "H3K4 di-methylation governs smooth muscle lineage identity and promotes vascular homeostasis by restraining plasticity". en. In: *Dev. Cell* 56.19, 2765–2782.e10.

Livingstone, Mark et al. (May 2010). "Mechanisms governing the control of mRNA translation". en. In: *Phys. Biol.* 7.2, p. 021001.

Long, Hannah K, Sara L Prescott, and Joanna Wysocka (Nov. 2016). "Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution". en. In: *Cell* 167.5, pp. 1170–1187.

Los, Georgyi V, Lance P Encell, et al. (June 2008). "HaloTag: a novel protein labeling technology for cell imaging and protein analysis". en. In: *ACS Chem. Biol.* 3.6, pp. 373–382.

Los, Georgyi V and Keith Wood (2007). "The HaloTag: a novel technology for cell imaging and protein analysis". en. In: *Methods Mol. Biol.* 356, pp. 195–208.

Lovén, Jakob et al. (Apr. 2013). "Selective inhibition of tumor oncogenes by disruption of super-enhancers". en. In: *Cell* 153.2, pp. 320–334.

Luger, K et al. (Sept. 1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution". en. In: *Nature* 389.6648, pp. 251–260.

Luger, Karolin, Mekonnen L Dechassa, and David J Tremethick (June 2012). "New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?" en. In: *Nat. Rev. Mol. Cell Biol.* 13.7, pp. 436–447.

Ma, Liang et al. (Apr. 2021). "Co-condensation between transcription factor and coactivator p300 modulates transcriptional bursting kinetics". en. In: *Mol. Cell* 81.8, 1682–1697.e7.

Mahat, Dig Bijay et al. (Aug. 2016). "Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq)". en. In: *Nat. Protoc.* 11.8, pp. 1455–1476.

Majello, B, L C Kenyon, and R Dalla-Favera (Dec. 1986). "Human c-myb protooncogene: nucleotide sequence of cDNA and organization of the genomic locus". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 83.24, pp. 9636–9640.

Mali, Prashant et al. (Feb. 2013). "RNA-guided human genome engineering via Cas9". en. In: *Science* 339.6121, pp. 823–826.

Malik, Sohail and Robert G Roeder (Nov. 2010). "The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation". en. In: *Nat. Rev. Genet.* 11.11, pp. 761–772.

Malin, Justin, Mohamed Radhouane Aniba, and Sridhar Hannenhalli (Aug. 2013). "Enhancer networks revealed by correlated DNAse hypersensitivity states of enhancers". en. In: *Nucleic Acids Res.* 41.14, pp. 6828–6838.

Mann, Matthias and Ole N Jensen (Mar. 2003). "Proteomic analysis of post-translational modifications". en. In: *Nat. Biotechnol.* 21.3, pp. 255–261.

Mansour, Marc R, Brian J Abraham, et al. (Dec. 2014). "Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element". In: *Science* 346.6215, pp. 1373–1377.

Mansour, Marc R, Takaomi Sanda, et al. (July 2013). "The TAL1 complex targets the FBXW7 tumor suppressor by activating miR-223 in human T cell acute lymphoblastic leukemia". en. In: *J. Exp. Med.* 210.8, pp. 1545–1557.

Marasco, Luciano E and Alberto R Kornblihtt (Apr. 2023). "The physiology of alternative splicing". en. In: *Nat. Rev. Mol. Cell Biol.* 24.4, pp. 242–254.

Martin, Erik W and Alex S Holehouse (Dec. 2020). "Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof". en. In: *Emerg Top Life Sci* 4.3, pp. 307–329.

Martire, Sara and Laura A Banaszynski (Sept. 2020). "The roles of histone variants in fine-tuning chromatin organization and function". en. In: *Nat. Rev. Mol. Cell Biol.* 21.9, pp. 522–541.

Matranga, Christian et al. (Nov. 2005). "Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes". en. In: *Cell* 123.4, pp. 607–620.

Mifsud, Borbala et al. (June 2015). "Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C". en. In: *Nat. Genet.* 47.6, pp. 598–606.

Migneault, Isabelle et al. (Nov. 2004). "Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking". en. In: *Biotechniques* 37.5, pp. 790–6, 798–802.

Mignone, Flavio et al. (Feb. 2002). "Untranslated regions of mRNAs". en. In: *Genome Biol.* 3.3, REVIEWS0004.

Millán-Zambrano, Gonzalo et al. (Sept. 2022). "Histone post-translational modifications - cause and consequence of genome function". en. In: *Nat. Rev. Genet.* 23.9, pp. 563–580.

Mohandas, T, R S Sparkes, and L J Shapiro (Jan. 1981). "Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation". en. In: *Science* 211.4480, pp. 393–396.

Moore, Lisa D, Thuc Le, and Guoping Fan (Jan. 2013). "DNA methylation and its basic function". en. In: *Neuropsychopharmacology* 38.1, pp. 23–38.

Morrison, Olivia and Jitendra Thakur (June 2021). "Molecular Complexes at Euchromatin, Heterochromatin and Centromeric Chromatin". en. In: *Int. J. Mol. Sci.* 22.13.

Mousavi, Kambiz et al. (Sept. 2013). "eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci". en. In: *Mol. Cell* 51.5, pp. 606–617.

Mucenski, M L et al. (May 1991). "A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis". en. In: *Cell* 65.4, pp. 677–689.

Nabet, Behnam et al. (May 2018). "The dTAG system for immediate and target-specific protein degradation". en. In: *Nat. Chem. Biol.* 14.5, pp. 431–441.

Nair, Sreejith J et al. (Mar. 2019). "Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly". In: *Nat. Struct. Mol. Biol.* 26.3, pp. 193–203.

Ngoc, Phuong Cao Thi et al. (Oct. 2018). "Identification of novel lncRNAs regulated by the TAL1 complex in T-cell acute lymphoblastic leukemia". en. In: *Leukemia* 32.10, pp. 2138–2151.

O'Reilly, Francis J and Juri Rappsilber (Nov. 2018). "Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology". en. In: *Nat. Struct. Mol. Biol.* 25.11, pp. 1000–1008.

Ogryzko, V V et al. (Nov. 1996). "The transcriptional coactivators p300 and CBP are histone acetyltransferases". en. In: *Cell* 87.5, pp. 953–959.

Okamoto, Sachiko et al. (Mar. 2019). "Highly efficient genome editing for single-base substitutions using optimized ssODNs with Cas9-RNPs". en. In: *Sci. Rep.* 9.1, p. 4811.

Oksuz, Ozgur et al. (2022). "Transcription factors interact with RNA to regulate genes".

Paddison, Patrick J et al. (Apr. 2002). "Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells". en. In: *Genes Dev.* 16.8, pp. 948–958.

Paix, A, H Schmidt, and G Seydoux (2016). "Cas9-assisted recombineering in C. elegans: Genome editing using in vivo assembly of linear DNAs". In: *Nucleic Acids Res.* 44.15, e128.

Pal, Debosree et al. (June 2023). "H4K16ac activates the transcription of transposable elements and contributes to their cis-regulatory function". en. In: *Nat. Struct. Mol. Biol.*

Palomero, Teresa et al. (Aug. 2006). "Transcriptional regulatory networks downstream of TAL1/SCL in T-cell acute lymphoblastic leukemia". en. In: *Blood* 108.3, pp. 986–992.

Pandolfi, P P et al. (Sept. 1995). "Targeted disruption of the GATA3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis". en. In: *Nat. Genet.* 11.1, pp. 40–44.

Parker, D et al. (1999). "Role of secondary structure in discrimination between constitutive and inducible activators". In: *Mol. Cell. Biol.* 19.8, pp. 5601–5607.

Pekowska, Aleksandra et al. (Nov. 2010). "A unique H3K4me2 profile marks tissue-specific gene regulation". en. In: *Genome Res.* 20.11, pp. 1493–1502.

Phillips-Cremins, Jennifer E et al. (June 2013). "Architectural protein subclasses shape 3D organization of genomes during lineage commitment". en. In: *Cell* 153.6, pp. 1281–1295.

Pradeepa, Madapura M et al. (June 2016). "Histone H3 globular domain acetylation identifies a new class of enhancers". en. In: *Nat. Genet.* 48.6, pp. 681–686.

Rada-Iglesias, Alvaro et al. (Feb. 2011). "A unique chromatin signature uncovers early developmental enhancers in humans". en. In: *Nature* 470.7333, pp. 279–283.

Rahnamoun, Homa et al. (Aug. 2018). "RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation". en. In: *Nat. Struct. Mol. Biol.* 25.8, pp. 687–697.

Ran, F Ann et al. (Nov. 2013). "Genome engineering using the CRISPR-Cas9 system". en. In: *Nat. Protoc.* 8.11, pp. 2281–2308.

Reuter, Jessica S and David H Mathews (Mar. 2010). "RNAstructure: software for RNA secondary structure prediction and analysis". en. In: *BMC Bioinformatics* 11, p. 129.

Ribeiro, Diogo M et al. (Jan. 2018). "Protein complex scaffolding predicted as a prevalent function of long non-coding RNAs". en. In: *Nucleic Acids Res.* 46.2, pp. 917–928.

Robertson, Gordon et al. (Aug. 2007). "Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing". en. In: *Nat. Methods* 4.8, pp. 651–657.

Robinson, Philip J J et al. (Sept. 2008). "30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction". en. In: *J. Mol. Biol.* 381.4, pp. 816–825.

Robson, Michael I, Alessa R Ringel, and Stefan Mundlos (June 2019). "Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D". en. In: *Mol. Cell* 74.6, pp. 1110–1122.

Sabari, Benjamin R, Alessandra Dall'Agnese, et al. (July 2018). "Coactivator condensation at super-enhancers links phase separation and gene control". en. In: *Science* 361.6400.

Sabari, Benjamin R, Zhanyun Tang, et al. (Apr. 2015). "Intracellular crotonyl-CoA stimulates transcription through p300-catalyzed histone crotonylation". en. In: *Mol. Cell* 58.2, pp. 203–215.

Sakamoto, K M et al. (July 2001). "Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 98.15, pp. 8554–8559.

Sanda, Takaomi et al. (Aug. 2012). "Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia". en. In: *Cancer Cell* 22.2, pp. 209–221.

Sandberg, Mark L et al. (Feb. 2005). "c-Myb and p300 regulate hematopoietic stem cell proliferation and differentiation". en. In: *Dev. Cell* 8.2, pp. 153–166.

Sano, Y and S Ishii (Feb. 2001). "Increased affinity of c-Myb for CREB-binding protein (CBP) after CBP-induced acetylation". In: *J. Biol. Chem.* 276.5, pp. 3674–3682.

Sanyal, Amartya et al. (Sept. 2012). "The long-range interaction landscape of gene promoters". en. In: *Nature* 489.7414, pp. 109–113.

Sartorelli, Vittorio and Shannon M Lauberth (June 2020). "Enhancer RNAs are an important regulatory layer of the epigenome". en. In: *Nat. Struct. Mol. Biol.* 27.6, pp. 521–528.

Sato, Kengo, Manato Akiyama, and Yasubumi Sakakibara (Feb. 2021). "RNA secondary structure prediction using deep learning with thermodynamic integration". en. In: *Nat. Commun.* 12.1, p. 941.

Schwalb, Björn et al. (June 2016). "TT-seq maps the human transient transcriptome". en. In: *Science* 352.6290, pp. 1225–1228.

Sentmanat, Monica F et al. (Jan. 2018). "A Survey of Validation Strategies for CRISPR-Cas9 Editing". en. In: *Sci. Rep.* 8.1, p. 888.

Shalem, Ophir et al. (Jan. 2014). "Genome-scale CRISPR-Cas9 knockout screening in human cells". en. In: *Science* 343.6166, pp. 84–87.

Sheiness, D and M Gardinier (July 1984). "Expression of a proto-oncogene (proto-myb) in hemopoietic tissues of mice". en. In: *Mol. Cell. Biol.* 4.7, pp. 1206–1212.

Shogren-Knaak, Michael et al. (Feb. 2006). "Histone H4-K16 acetylation controls chromatin structure and protein interactions". en. In: *Science* 311.5762, pp. 844–847.

Shrinivas, Krishna et al. (Aug. 2019). "Enhancer Features that Drive Formation of Transcriptional Condensates". In: *Mol. Cell* 75.3, 549–561.e7.

Sigova, Alla A et al. (2015). "Transcription factor trapping by RNA in gene regulatory elements". In: *Science* 350.6263, pp. 978–981.

Silverman, L B et al. (Mar. 2001). "Improved outcome for children with acute lymphoblastic leukemia: results of Dana-Farber Consortium Protocol 91-01". en. In: *Blood* 97.5, pp. 1211–1218.

Skene, Peter J and Steven Henikoff (Jan. 2017). "An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites". en. In: *Elife* 6.

Smith, Charlotte et al. (Jan. 2023). "Harnessing the MYB-dependent TAL1 5'super-enhancer for targeted therapy in T-ALL". en. In: *Mol. Cancer* 22.1, p. 12.

Smola, Matthew J, Greggory M Rice, et al. (Oct. 2015). "Selective 2â²−hydroxylacylationanalyzedbypri In: *Nat. Protoc.* 10.11, pp. 1643–1669.

Smola, Matthew J and Kevin M Weeks (June 2018). "In-cell RNA structure probing with SHAPE-MaP". en. In: *Nat. Protoc.* 13.6, pp. 1181–1195.

Soboleski, Mark R, Jason Oaks, and William P Halford (Mar. 2005). "Green fluorescent protein is a quantitative reporter of gene expression in individual eukaryotic cells". en. In: *FASEB J.* 19.3, pp. 440–442.

Spitz, François and Eileen E M Furlong (Sept. 2012). "Transcription factors: from enhancer binding to developmental control". en. In: *Nat. Rev. Genet.* 13.9, pp. 613–626.

Strahl, B D and C D Allis (Jan. 2000). "The language of covalent histone modifications". en. In: *Nature* 403.6765, pp. 41–45.

Sungalee, Stephanie et al. (May 2021). "Histone acetylation dynamics modulates chromatin conformation and allele-specific interactions at oncogenic loci". en. In: *Nat. Genet.* 53.5, pp. 650–662.

Sur, Inderpreet and Jussi Taipale (Aug. 2016). "The role of enhancers in cancer". en. In: *Nat. Rev. Cancer* 16.8, pp. 483–493.

Tan, Shi Hao et al. (July 2019). "The enhancer RNA ARIEL activates the oncogenic transcriptional program in T-cell acute lymphoblastic leukemia". en. In: *Blood* 134.3, pp. 239–251.

Taylor, Gillian C A et al. (Dec. 2013). "H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction". en. In: *Genome Res.* 23.12, pp. 2053–2065.

Tessarz, Peter and Tony Kouzarides (Nov. 2014). "Histone core modifications regulating nucleosome structure and dynamics". en. In: *Nat. Rev. Mol. Cell Biol.* 15.11, pp. 703–708.

Thanos, D and T Maniatis (Dec. 1995). "Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome". en. In: *Cell* 83.7, pp. 1091–1100.

Thomas, Henry F et al. (Jan. 2021). "Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements". en. In: *Mol. Cell* 0.0.

Thompson, Paul R et al. (Apr. 2004). "Regulation of the p300 HAT domain via a novel activation loop". en. In: *Nat. Struct. Mol. Biol.* 11.4, pp. 308–315.

Thoms, Matthias et al. (Aug. 2015). "The Exosome Is Recruited to RNA Substrates through Specific Adaptor Proteins". en. In: *Cell* 162.5, pp. 1029–1038.

Tie, Feng et al. (Sept. 2009). "CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing". en. In: *Development* 136.18, pp. 3131–3141.

Tropberger, Philipp et al. (Feb. 2013). "Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer". en. In: *Cell* 152.4, pp. 859–872.

Tsai, Shengdar Q et al. (June 2017). "CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets". en. In: *Nat. Methods* 14.6, pp. 607–614.

Ule, Jernej et al. (Dec. 2005). "CLIP: a method for identifying protein-RNA interaction sites in living cells". en. In: *Methods* 37.4, pp. 376–386.

Van Vlierberghe, Pieter and Adolfo Ferrando (Oct. 2012). "The molecular basis of T cell acute lymphoblastic leukemia". en. In: *J. Clin. Invest.* 122.10, pp. 3398–3406.

Vaquerizas, Juan M et al. (Apr. 2009). "A census of human transcription factors: function, expression and evolution". en. In: *Nat. Rev. Genet.* 10.4, pp. 252–263.

Venkatesh, Swaminathan and Jerry L Workman (Mar. 2015). "Histone exchange, chromatin structure and the regulation of transcription". en. In: *Nat. Rev. Mol. Cell Biol.* 16.3, pp. 178–189.

Verdin, Eric and Melanie Ott (Apr. 2015). "50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond". en. In: *Nat. Rev. Mol. Cell Biol.* 16.4, pp. 258–264.

Vermeulen, Michiel et al. (Oct. 2007). "Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4". en. In: *Cell* 131.1, pp. 58–69.

Vernimmen, Douglas et al. (Aug. 2011). "Polycomb eviction as a new distant enhancer function". en. In: *Genes Dev.* 25.15, pp. 1583–1588.

Voronova, A F and F Lee (June 1994). "The E2A and tal-1 helix-loop-helix proteins associate in vivo and are modulated by Id proteins during interleukin 6-induced myeloid differentiation". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 91.13, pp. 5952–5956.

Walf-Vorderwülbecke, V et al. (Apr. 2018). "Targeting acute myeloid leukemia by drug-induced c-MYB degradation". en. In: *Leukemia* 32.4, pp. 882–889.

Wang, Yu-Chieh, Suzanne E Peterson, and Jeanne F Loring (Feb. 2014). "Protein post-translational modifications and regulation of pluripotency in human stem cells". en. In: *Cell Res.* 24.2, pp. 143–160.

Wang, Zhibin et al. (July 2008). "Combinatorial patterns of histone acetylations and methylations in the human genome". en. In: *Nat. Genet.* 40.7, pp. 897–903.

Wang, Zhong et al. (Mar. 2022). "Prediction of histone post-translational modification patterns based on nascent transcription data". en. In: *Nat. Genet.* 54.3, pp. 295–305.

Wee, Hee-Jun et al. (Nov. 2008). "PEBP2-beta/CBF-beta-dependent phosphorylation of RUNX1 and p300 by HIPK2: implications for leukemogenesis". en. In: *Blood* 112.9, pp. 3777–3787.

Wessels, Hans-Hermann et al. (Mar. 2020). "Massively parallel Cas13 screens reveal principles for guide RNA design". In: *Nat. Biotechnol.*

Westin, E H, K M Gorse, and M F Clarke (Aug. 1990). "Alternative splicing of the human c-myb gene". en. In: *Oncogene* 5.8, pp. 1117–1124.

Whitson, R H, T Huang, and K Itakura (May 1999). "The novel Mrf-2 DNA-binding domain recognizes a five-base core sequence through major and minor-groove contacts". en. In: *Biochem. Biophys. Res. Commun.* 258.2, pp. 326–331.

Whyte, Warren A et al. (Apr. 2013). "Master transcription factors and mediator establish super-enhancers at key cell identity genes". en. In: *Cell* 153.2, pp. 307–319.

Wilsker, Deborah et al. (2002). "ARID Proteins: A Diverse Family of DNA Binding Proteins Implicated in the Control of Cell Growth, Differentiation, and Development1". In: *Heart* 28, p. 30.

Wingender, Edgar, Torsten Schoeps, and Jürgen Dönitz (Jan. 2013). "TFClass: an expandable hierarchical classification of human transcription factors". en. In: *Nucleic Acids Res.* 41.Database issue, pp. D165–70.

Wolffe, A P and D Guschin (Apr. 2000). "Review: chromatin structural features and targets that regulate transcription". en. In: *J. Struct. Biol.* 129.2-3, pp. 102–122.

Yao, T P et al. (May 1998). "Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300". en. In: *Cell* 93.3, pp. 361–372.

Yap, Karen, Tek Hong Chung, and Eugene V Makeyev (Jan. 2022). "Hybridization-proximity labeling reveals spatially ordered interactions of nuclear RNA compartments". en. In: *Mol. Cell* 82.2, 463–478.e11.

Yong, Wai Khang et al. (Jan. 2023). "ChIP-MS reveals the local chromatin composition by label-free quantitative proteomics". en.

Zaret, Kenneth S and Susan E Mango (Apr. 2016). "Pioneer transcription factors, chromatin dynamics, and cell fate control". en. In: *Curr. Opin. Genet. Dev.* 37, pp. 76–81.

Zeng, Lei, Qiang Zhang, et al. (Apr. 2008). "Structural basis of site-specific histone recognition by the bromodomains of human coactivators PCAF and CBP/p300". en. In: *Structure* 16.4, pp. 643–652.

Zeng, Lei and Ming Ming Zhou (Feb. 2002). "Bromodomain: an acetyl-lysine binding domain". en. In: *FEBS Lett.* 513.1, pp. 124–128.

Zentner, Gabriel E, Paul J Tesar, and Peter C Scacheri (Aug. 2011). "Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions". en. In: *Genome Res.* 21.8, pp. 1273–1283.

Zhang, Jian-Ping et al. (June 2016). "Different Effects of sgRNA Length on CRISPR-mediated Gene Knockout Efficiency". en. In: *Sci. Rep.* 6, p. 28566.

Zhang, Jing et al. (May 2019). "Regulation of histone arginine methylation/demethylation by methylase and demethylase (Review)". en. In: *Mol. Med. Rep.* 19.5, pp. 3963–3971.

Zhang, Luqing et al. (Jan. 2015). "Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9". In: *PLoS One* 10.3, e0120396.

Zhao, Zhihu et al. (Nov. 2006). "Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions". en. In: *Nat. Genet.* 38.11, pp. 1341–1347.

Zheng, Meizhen et al. (Feb. 2019). "Multiplex chromatin interactions with single-molecule precision". en. In: *Nature* 566.7745, pp. 558–562.

Zhou, Yan et al. (Dec. 2013). "Chromatin looping defines expression of TAL1, its flanking genes, and regulation in T-ALL". In: *Blood* 122.26, pp. 4199–4209.

Chapter 8

Appendix

8.1 DNA co-ordinates and sequences used for CRISPR/Cas9 mediated genome editing

8.1.1 gRNA targeting locations

Table 8.1: Genomic locations of gRNA targets

gRNA name	Sequence	Genomic target
CBP C-terminus gRNA	CACGCTAGAGAAGTTTGTGG	chr16:3,727,729
MYB C-terminus gRNA	TGCATTCTCAGCCCGGACGC	chr6:135,217,950
Scramble gRNA	GCTGATCTATCGCGGTCGTC	N/A
MYB KO gRNA 1	CAGCATATATAGCAGTGACG	chr6:135,185,922
MYB KO gRNA 2	GAAGCAGCCCATCATAGTCA	chr6:135,185,946
MYB KO gRNA 3	AGTCTGGAAAGCGTCACTTG	chr6:135,185,996

8.1.2 CRISPR/Cas9 HDR sequences

Homology arm
PAM site
Insert sequence
Homology arm

8.1.2.1   CBP-FTH ssODN sequence

TGGTCGGGGACACCACGGGGGACACGCTAGAGAAGTTTGTGGAAGGCTTGATGggtggaggttccgacta
caaggacgacgatgacaaaGAAAACCTGTATTTTCAGGGCcatcaccatcaccatcacggtggaggttcc
TAGCATTGTGAGAGCATCACCTTTTCCCTTTCATGTTCT

8.1.2.2 CBP-FTHH repair template sequence

```
TACCCACCGGCCATGCAGCAGCAGCAGCGCATGCAGCAGCATCTCCCCCTCCAGGGCAGCTCCAT
GGGCCAGATGGCGGCTCAGATGGGACAGCTTGGCCAGATGGGGCAGCCGGGGCTGGGGGCAGACA
GCACCCCCAACATCCAGCAAGCCCTGCAGCAGCGGATTCTGCAGCAACAGCAGATGAAGCAGCAG
ATTGGGTCCCCAGGCCAGCCGAACCCCATGAGCCCCCAGCAACACATGCTCTCAGGACAGCCACA
GGCCTCGCATCTCCCTGGCCAGCAGATCGCCACGTCCCTTAGTAACCAGGTGCGGTCTCCAGCCC
CTGTCCAGTCTCCACGGCCCCAGTCCCAGCCTCCACATTCCAGCCCGTCACCACGGATACAGCCC
CAGCCTTCGCCACACCACGTCTCACCCCAGACTGGTTCCCCCCCACCCCGGACTCGCAGTCACCAT
GGCCAGCTCCATAGATCAGGGACACTTGGGGAACCCCGAACAGAGTGCAATGCTCCCCCAGCTGA
ACACCCCCAGCAGGAGTGCGCTGTCCAGCGAACTGTCCCTGGTCGGGGACACCACGGGGGACACG
CTAGAGAAGTTTGTGGAAGGCTTGATGGGAGGCGGCAGCGACTATAAGGATGACGACGACAAAGA
AAACCTGTACTTCCAAGGTCATCACCATCACCATCACGGCGGCGGCTCCTTGGAGGTTCTGTTCC
AAGGCCCGGGCAGCGAAATCGGAACTGGGTTCCCCTTTGATCCGCATTACGTGGAAGTCCTCGGT
GAGCGGATGCACTATGTAGACGTTGGCCCACGCGATGGTACCCCTGTGCTTTTTTTTGCATGGTAA
CCCTACCTCCAGCTACGTTTGGAGAAACATAATCCCACATGTCGCCCCGACCCACAGATGCATTG
CACCTGACCTGATAGGAATGGGAAAATCAGACAAGCCAGATCTTGGTTATTTTTTTGACGATCAC
GTAAGGTTCATGGATGCGTTTATAGAGGCGTTGGGTCTTGAAGAGGTGGTTCTCGTGATACATGA
TTGGGGGTCCGCTCTCGGTTTTCATTGGGCGAAAAGAAACCCGGAACGGGTCAAGGGAATAGCTT
TTATGGAATTTATCCGACCTATACCTACCTGGGATGAGTGGCCTGAATTCGCACGAGAAACATTC
CAAGCCTTTAGGACCACCGACGTAGGTAGGAAACTCATCATTGACCAAAACGTGTTCATAGAGGG
GACACTCCCTATGGGGGTTGTGCGCCCTCTTACCGAGGTCGAAATGGACCATTACCGAGAGCCCT
TCCTGAATCCTGTGGATCGAGAACCGTTGTGGAGGTTCCCGAATGAACTTCCAATTGCGGGCGAA
CCGGCTAACATAGTTGCTCTTGTCGAGGAATATATGGACTGGCTTCATCAGAGCCCCGTGCCAAA
GCTCTTGTTCTGGGGAACCCCAGGAGTCTTGATACCTCCTGCCGAGGCCGCGCGGTTGGCTAAAT
CACTTCCGAATTGTAAGGCGGTCGACATAGGCCCCGGGCTGAATCTTTTGCAGGAAGATAATCCA
GATCTCATAGGTTCAGAAATTGCCCGGTGGCTGTCAACTCTTGAGATTAGTGGCTAGCATTGTGA
GAGCATCACCTTTTCCCTTTCATGTTCTTGGACCTTTTGTACTGAAAATCCAGGCATCTAGGTTC
TTTTTATTCCTAGATGGAACTGCGACTTCCGAGCCATGGAAGGGTGGATTGATGTTTAAAGAAAC
AATACAAAGAATATATTTTTTTTGTTAAAAACCAGTTGATTTAAATATCTGGTCTCTCTCTTTGGT
TTTTTTTTGGCGGGGGGGGTGGGGGGGGGTTCTTTTTTTTTCCGTTTTGTTTTTGTTTGGGGGGAGGG
GGGTTTTGTTTGGATTCTTTTTGTCGTCATTGCTGGTGACTCATGCCTTTTTTTTAACGGGAAAAA
CAAGTTCATTATATTCATATTTTTTATTTGTATTTCAAGACTTTAAACATTTATGTTTAAAAGTA
AGAAGAAAAATAATATTCAGAACTGATTCCTGAAATAATGCAAGCTTATAATGTATCCCGATAAC
TTTGTGATGTTTCGGGAAGATTTTTTTCTATAGTGAACTCTGTGGGCGTCTCCCAGTATTACCCT
GGATGATAGGAATTGACTCCGGCGTGCACACACGTACACACCCACACACATCTATCT
```

8.1.2.3 MYB-HTHAS repair template sequence

```
aaggaaaaaaattaaactatactgataaatatttctaaagtaattgatgaaaacatatatatgacactt
agcaggtgttgatcaattccagtgcgtgcatgagctgctttgcactgtggctacagcactcagcagttt
gcttagtgagtagatggccatagctcactttaggtatctggagaataatgctcattcctgtctccagct
ttgtaaaattgcctaatagacggaaggcattaaaatatgaagatgggattatatcttatgaagctgctt
ggaatgggttttccctcttaacaaataatttctgatcatccatatcccttaagaacaccggatttctgg
gggccagggaggtaagattctatctgacaaagccttcctggtgtcaaccacttgccatctgttggtcag
tgctggccctgctgggtctatagaatgagcttctttgtctgacgctcctgttgccatccctttctccat
cagCCTTGTAGCAGTACCTGGGAACCTGCATCCTGTGGAAAGATGGAGGAGCAGATGACATCTTCCAGT
CAAGCTCGTAAATACGTGAATGCATTCTCAGCCCGGACGCTAGtcATGggtggagttcccatcaccatc
accatcacGAAAACCTGTATTTTCAGGGCTATCCCTATGACGTGCCCGATTACGCCggtggaggttcct
ggagccacccgcagttcgagaaaggtggaggttccggaggtggatcgggaggttcggcgtggagccacc
cgcagttcgaaaaaTAGGACATTTCCAGAAAAGCATTATGGTTTTCAGAACACTTCAAGTTGACTTGGG
ATATATCATTCCTCAACATGAAACTTTTCATGAATGGGAGAAGAACCTATTTTTGTTGTGGTACAACAG
TTGAGAGCAGCACCAAGTGCATTTAGTTGAATGAAGTCTTCTTGGATTTCACCCAACTAAAAGGATTTT
TAAAAATAAATAACAGTCTTACCTAAATTATTAGGTAATGAATTGTAGCCAGTTGTTAATATCTTAATG
CAGATTTTTTTAAAAAAAACATAAAATGATTTATCTGTATTTTAAAGGATCCAACAGATCAGTATTTTT
TCCTGTGATGGGTTTTTTGAAATTTGACACATTAAAAGGTACTCCAGTATTTCACTTTTCTCGATCACT
AAACATATGCATATATTTTTAAAAATCAGTAAAAGCATTACTCTAAGTGTAGACTTAATACCATGTGAC
ATTTAATCCAGATTGTAAATGCTCATTTATGGTTAATGACATTGAAGGTACATTTATTGTACCAAACCA
TTTTATGAGTTTTCTGTTAGCTTGCTTTAAAAATTATTACTGTAAGAAATAGTTTTATAAAA
```

8.1.2.4 MYB-Halo repair template sequence

```
aaggaaaaaaattaaactatactgataaatatttctaaagtaattgatgaaaacatatatatgac
acttagcaggtgttgatcaattccagtgcgtgcatgagctgctttgcactgtggctacagcactc
agcagtttgcttagtgagtagatggccatagctcactttaggtatctggagaataatgctcattc
ctgtctccagctttgtaaaattgcctaatagacggaaggcattaaaatatgaagatgggattata
tcttatgaagctgcttggaatgggttttccctcttaacaaataatttctgatcatccatatccct
taagaacaccggatttctgggggccagggaggtaagattctatctgacaaagccttcctggtgtc
aaccacttgccatctgttggtcagtgctggccctgctgggtctatagaatgagcttctttgtctg
acgctcctgttgccatccctttctccatcagCCTTGTAGCAGTACCTGGGAACCTGCATCCTGTG
GAAAGATGGAGGAGCAGATGACATCTTCCAGTCAAGCTCGTAAATACGTGAATGCATTCTCAGCC
CGGACGCTAGtcATGGGCAGCGAAATCGGAACTGGGTTCCCCTTTGATCCGCATTACGTGGAAGT
CCTCGGTGAGCGGATGCACTATGTAGACGTTGGCCCACGCGATGGTACCCCTGTGCTTTTTTTGC
ATGGTAACCCTACCTCCAGCTACGTTTGGAGAAACATAATCCCACATGTCGCCCCGACCCACAGA
TGCATTGCACCTGACCTGATAGGAATGGGAAAATCAGACAAGCCAGATCTTGGTTATTTTTTTGA
CGATCACGTAAGGTTCATGGATGCGTTTATAGAGGCGTTGGGTCTTGAAGAGGTGGTTCTCGTGA
TACATGATTGGGGGTCCGCTCTCGGTTTTCATTGGGCGAAAAGAAACCCGGAACGGGTCAAGGGA
ATAGCTTTTATGGAATTTATCCGACCTATACCTACCTGGGATGAGTGGCCTGAATTCGCACGAGA
AACATTCCAAGCCTTTAGGACCACCGACGTAGGTAGGAAACTCATCATTGACCAAAACGTGTTCA
TAGAGGGGACACTCCCTATGGGGGTTGTGCGCCCTCTTACCGAGGTCGAAATGGACCATTACCGA
GAGCCCTTCCTGAATCCTGTGGATCGAGAACCGTTGTGGAGGTTCCCGAATGAACTTCCAATTGC
GGGCGAACCGGCTAACATAGTTGCTCTTGTCGAGGAATATATGGACTGGCTTCATCAGAGCCCCG
TGCCAAAGCTCTTGTTCTGGGGAACCCCAGGAGTCTTGATACCTCCTGCCGAGGCCGCGCGGTTG
GCTAAATCACTTCCGAATTGTAAGGCGGTCGACATAGGCCCCGGGCTGAATCTTTTGCAGGAAGA
TAATCCAGATCTCATAGGTTCAGAAATTGCCCGGTGGCTGTCAACTCTTGAGATTAGTGGCTGAC
ATTTCCAGAAAAGCATTATGGTTTTCAGAACACTTCAAGTTGACTTGGGATATATCATTCCTCAA
CATGAAACTTTTCATGAATGGGAGAAGAACCTATTTTTGTTGTGGTACAACAGTTGAGAGCAGCA
CCAAGTGCATTTAGTTGAATGAAGTCTTCTTGGATTTCACCCAACTAAAAGGATTTTTAAAAATA
AATAACAGTCTTACCTAAATTATTAGGTAATGAATTGTAGCCAGTTGTTAATATCTTAATGCAGA
TTTTTTTAAAAAAAACATAAAATGATTTATCTGTATTTTAAAGGATCCAACAGATCAGTATTTTT
TCCTGTGATGGGTTTTTTGAAATTTGACACATTAAAAGGTACTCCAGTATTTCACTTTTCTCGAT
CACTAAACATATGCATATATTTTTAAAAATCAGTAAAAGCATTACTCTAAGTGTAGACTTAATAC
CATGTGACATTTAATCCAGATTGTAAATGCTCATTTATGGTTAATGACATTGAAGGTACATTTAT
TGTACCAAACCATTTTATGAGTTTTCTGTTAGCTTGCTTTAAAAATTATTACTGTAAGAAATAGT
TTTATAAAA
```