
THE UNIVERSITY
*of* EDINBURGH

# The TAZ2 domain of CBP binds DNA to regulate histone acetylation

**Thomas W. Sheahan**


Thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in

Cell and Molecular Biology

University of Edinburgh

2020

# Declaration of authorship

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where stated otherwise by reference or acknowledgment, the work presented is entirely my own.

Thomas W. Sheahan

*Ed elli: 'O frate, andar in sù che porta?*
*ché non mi lascerebbe ire a' martìri*
*l'angel di Dio che siede in su la porta.*

*Prima convien che tanto il ciel m'aggiri*
*di fuor da essa quanto fece in vita*
*per ch'io 'ndugiai al fine i buon sospiri…'*

'Brother,' he said, 'what point in going up?
God's angel, sitting at the gate up there,
would not admit me to the Penances.

The skies must circle first around me here
the length of time they did around my life,
since I delayed good sighs until the end…'

Dante Alighieri (tr. Robin Kirkpatrick)
*Purgatorio IV (127-132)*

# Abstract

DNA sequences that regulate gene expression are marked by distinct chromatin architectures that segregate these functional elements from the surrounding genome. In mammals, DNA methylation is prevalent throughout the genome and is repressive to transcription. However, the majority of gene promoters are associated with regions of CG- and CpG-rich DNA called CpG islands (CGIs), which are unmethylated and permissive to transcription. Although previous studies have shown that CGIs act as platforms for the recruitment of both active and repressive chromatin-modifying activities, it remains only partially understood how mechanistically the chromatin environment at CGIs is established. To address this, an unbiased proteomics approach was adopted to generate an inventory of CGI-binding proteins with the aim of understanding how binding to CGIs influences promoter function. In parallel to this, a candidate approach was used to gain a greater understanding of how the chromatin landscape is established at distal regulatory regions called enhancers. Enhancers are typically associated with monomethylation of H3K4, placed by the histone methyltransferases (HMTs) MLL3 and MLL4, and with acetylation of H3K27, placed by the histone acetyltransferase (HAT) CBP and its paralogue p300. To understand how MLL3 and MLL4 contribute to enhancer function, the MLL3/MLL4 complexes were purified from mammalian cells and interaction partners were analysed by mass spectrometry to determine complex composition. To address how CBP contributes to the chromatin environment at enhancers, an *in vitro* domain mapping approach was developed using purified CBP and recombinant nucleosome templates. This showed that the CBP TAZ2 domain, located downstream of the catalytic HAT domain, is required for efficient acetylation of H3K27 in chromatin and that this is mediated through the TAZ2 domain driving association with nucleosomes via sequence-independent interaction with DNA. Further work showed that the TAZ2 domain is important for stable binding to chromatin *in vivo* and facilitates specific acetylation of H3K27 to activate transcription from regulatory elements. Together, this work elucidates a novel mechanism by which CBP HAT activity is selective for H3K27, forming the basis of a model in which mechanisms that determine HAT substrate specificity are vital to ensure robust regulation of gene expression.

# Lay summary

Every cell in a multicellular organism contains the same genetic information. A major question in biology is how this identical information can be used to give rise to a large range of different cell types. Ultimately, this is achieved by controlling which genes are switched on and off in each cell, with each cell type expressing a different set of genes that defines their identity.

The genetic information in cells is encoded in DNA, which is wound around proteins to form a packaged structure called chromatin. The highly packaged nature of chromatin means that genes can become relatively inaccessible, so that the cell is unable to switch genes on. To overcome this inaccessibility, the chromatin undergoes chemical modification to generate regions of more open chromatin, where genes can be activated more readily. These modifications to chromatin are carried out by enzymes, but it is still unclear how these enzymes are controlled so that individual genes are switched on or off in the right cells and at the right time.

During my PhD, I used multiple techniques to try to understand how some of these chromatin-modifying enzymes are regulated. First, I tried to isolate regions of genes that control gene expression to generate an inventory of all of the proteins that bind to them. The purpose of this was to determine if there are factors present at gene regulatory elements that might be important but that had previously been overlooked. My work identified several possible proteins that might be involved in regulating gene expression that will be important to follow up in future studies.

Second, I focussed on a chromatin-modifying enzyme that is known to activate genes to find out how this protein contributes to the highly complex process of switching genes on. I found that this enzyme contains a region that binds to DNA, and controls how it modifies chromatin and alters gene expression. This chromatin-modifying enzyme has a role in diseases such as cancer, so this work helping to understand how it is regulated may help us uncover how the enzyme is involved in disease and design new therapies in the future.

# Acknowledgements

First of all, I would like to thank my supervisor, Philipp Voigt, for his support over the last four years. It wasn't easy for me re-starting a PhD, but I very quickly felt at home in your lab. I've learnt a great deal during my time there, but most importantly I have grown in confidence and independence, two things that I had lost before coming to Edinburgh. I hope I carry them with me into the future.

Second, I want to extend this thanks to all the members of the Voigt lab. Thank you to Kim, Elana, Marie, Stefania, Viktoria, Katy, and all the others who have passed through the lab over these years. All of you have discussed ideas, helped with things in the lab, shown me where to find things, and generally just been there whenever I have needed help. Thanks especially to Kim, who showed me how to do everything related to the baculovirus system, and Elana, for constantly supplying me with histone octamers.

Thank you to the many other people throughout the institute who helped me along the way. This includes Christos Spanos and Juan Zhou, for support and advice with mass spectrometry, and Alba Abad-Fernandez, for advice about crosslinking. Thank you also to Adrian Bird, for valuable support and criticism as part of my thesis committee, and to Sara Buonomo and Marcus Wilson for input during lab meetings. Thank you also to those in my academic life outside Edinburgh, like Louis Mahadevan in Oxford, who helped give me the confidence to re-start this PhD in the first place.

Outside of the lab, I would also like to thank the fantastic medical team who supported me in Edinburgh. This includes Dr David Millar, Dr Louise Duthie and Ms Louise Foley at the University Health Centre, and Dr Gwo-Tzer Ho at the Western General. I wasn't very well when I first arrived in Edinburgh, and there have been times since then when I wasn't very well either, but you have been amazing and my health has now improved enormously, to the point where I

might soon be able to try life without medication, something which I couldn't have imagined four years ago.

I would of course like to thank my family. My mum, my dad, my brother and my grandparents have been constantly supportive, from helping us get our lives back together at the start of 2016, to taking us on holiday, visiting us, and just being there when I needed to talk. Especially in this last year when we were moving house in Cambridge and struggling to cope with everything that was going wrong, you were there for us and kept us going.

I would also like to thank the new extended family that I've gained since I started this PhD. Thank you to Anca's mum and dad, to Bogdan, Ioana and Vladimir, for accepting a slightly awkward British boy into your lives and making me feel so much like a member of your family.

And finally, I would like to thank Anca. Thank you for always believing in me, even though I didn't believe in myself. For putting up with me when I was difficult. For being there when I needed you. For listening to me talk about histone acetylation far too much. And for giving me wings.

The last four years have been tough, and the last two have been especially hard. But in the last two years we also got married! And we have a cat! And now we are finally living together again and I can't wait to just spend the rest of my life with you.

And finally, finally I would like to thank Woolly, our cat, for making us both so happy since you arrived in our lives. For keeping me sane while I was writing this thesis, for making sure I take regular breaks (by meowing at me until I feed you), and telling me when you think I've written enough by sitting directly on my laptop.

# Table of contents

# List of figures

# List of tables

# List of abbreviations

| | |
|---|---|
| Acetyl-CoA | Acetyl coenzyme A |
| AML | Acute myeloid leukaemia |
| APS | Ammonium persulphate |
| ATP | Adenosine triphosphate |
| bp | Base pairs |
| BSA | Bovine serum albumin |
| °C | Degrees Celcius |
| C- | Carboxyl- |
| cAMP | Cyclic adenosine monophosphate |
| CBP | CREB-binding protein |
| CFP1 | CXXC finger protein 1 |
| CGI | CpG island |
| ChIP | Chromatin immunoprecipitation |
| ChIP-qPCR | Chromatin immunoprecipitation followed by qPCR |
| ChIP-seq | Chromatin immunoprecipitation followed by massively parallel sequencing |
| COMPASS | Complex of proteins associated with Set1 |
| CREB | cAMP response element binding protein |
| DAPI | 4,6-diamino-2-phenylindole |
| dNTP | deoxyribonucleotide |
| DMSO | Dimethyl sulfoxide |
| DNA | Deoxyribonucleic acid |
| DNase | Deoxyribonuclease |
| DNMT | DNA methyltransferase |
| DTT | Dithiothreitol |
| EDTA | Ethylenediaminetetraacetic acid |
| EMSA | Electrophoretic mobility shift assay |
| ES cell | Embryonic stem cell |
| *g* | Relative centrifugal force |
| GFP | Green fluorescent protein |
| h | Hours |
| HAT | Histone acetyltransferase |

| | |
|---|---|
| HDAC | Histone deacetylase |
| HEPES | N-2-hydroxyethylpeperazine-N'-2-ethanesulfonic acid |
| HMT | Histone methyltransferase |
| H2AK119ub1 | Histone H2A lysine 119 monoubiquitylation |
| H2BK5ac | Histone H2B lysine 5 acetylation |
| H3K4me1/3 | Histone H3 lysine 4 mono/trimethylation |
| H3K9ac | Histone H3 lysine 9 acetylation |
| H3K9me3 | Histone H3 lysine 9 trimethylation |
| H3K14ac | Histone H3 lysine 14 acetylation |
| H3K18ac | Histone H3 lysine 18 acetylation |
| H3K23ac | Histone H3 lysine 23 acetylation |
| H3K27ac | Histone H3 lysine 27 acetylation |
| H3K36me2/3 | Histone H3 lysine 36 di/trimethylation |
| H3K64ac | Histone H3 lysine 64 acetylation |
| H3K122ac | Histone H3 lysine 122 acetylation |
| H4K5ac | Histone H4 lysine 5 acetylation |
| IF | Immunofluorescence |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| kb | Kilobase pairs |
| kDa | Kilodaltons |
| KDM2A | Lysine-specific demethylase 2A |
| KDM2B | Lysine-specific demethylase 2B |
| KID | Kinase-inducible domain |
| KIX | KID-interaction domain |
| L | Litres |
| LIF | Leukaemia-inducible factor |
| M | Moles/litre |
| mA | Milli amperes |
| MAPK | Mitogen activated protein kinase |
| MBD | Methyl-CpG binding domain |
| 5mC/5hmC | 5-methyl/hydroxymethyl-cytosine |
| min | Minutes |
| MLL | Mixed lineage leukaemia |
| N- | Amino- |
| ng | Nanograms |

| | |
|---|---|
| nt | Nucleotides |
| NuRD | Nucleosome remodelling and histone deacetylase |
| OD | Optical density |
| p300 | E1A-binding protein, 300 kDa |
| PAGE | Polyacrylamide gel electrophoresis |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PcG | Polycomb group |
| PHD | Plant homeodomain |
| PIC | Protease inhibitor cocktail |
| PMSF | Phenylmethylsulfonyl fluoride |
| PRC1/2 | Polycomb repressive complex 1/2 |
| qPCR | Quantitative PCR |
| RNA | Ribonucleic acid |
| RNase | Ribonuclease |
| RT | Reverse transcriptase |
| SAM | S-adenosyl-L-methionine |
| SDS | Sodium dodecyl sulphate |
| SEM | Standard error of the mean |
| TAZ | Transcriptional adaptor zinc finger |
| TBE | Tris-borate-EDTA |
| TCA | Trichloroacetic acid |
| TEMED | Tetramethylethylenediamine |
| TF | Transcription factor |
| TG | Tris-Glycine |
| Tris | Tris(hydroxymethyl)aminomethane |
| TSS | Transcription start site |
| µg | Microgram |
| µL | Microlitre |
| µM | Micromoles/L |
| UTR | Untranslated region |
| V | Volts |
| wt | Wild type |
| ZF-CXXC | ZF-CXXC DNA binding domain |
| *Xist* | X inactive specific transcript |

# 1. Introduction

## 1.1 Chromatin and transcriptional regulation

### *1.1.1 Chromatin is a complex of DNA and histone proteins*

DNA, originally called “nuclein”, was first isolated by Friedrich Miescher in 1869 from the nuclei of lymphocytes (Miescher-Rüsch, 1871). Ten years later, in 1879, Walther Flemming used aniline dyes to stain structures within the nucleus that he named “chromatin” and believed to correspond to the DNA isolated by Miescher (Flemming, 1879). This chromatin was later found to be identical with chromosomes, which were found by Theodor Boveri to be continuous entities throughout the cell cycle and central in heredity (Boveri, 1904). In the 1880s, Albrecht Kossel purified from erythrocyte nuclei a highly basic protein that was bound to nucleic acid, which he named “histone” (Kossel, 1884). Later analysis of the components of chromosomes showed that the vast majority of chromosome mass corresponds to nucleic acid in complex with histone protein (Mirsky and Ris, 1947). This DNA-histone complex, which is now referred to as chromatin, is the form in which DNA is packaged in the eukaryotic nucleus. Chromatin is divided into two main classes, called euchromatin and heterochromatin, intially identified by their differential staining properties (Heitz, 1928). Heterochromatin is relatively condensed, and as early as 1929 was found to comprise the gene-poor fraction of chromosomes (Heitz, 1929). Euchromatin, by contrast, is decondensed, gene-rich and associated with the actively transcribed regions of the genome. The concept that chromatin is closely linked with the activity and function of DNA therefore has a well established history in chromatin biology.

### *1.1.2 Chromatin and transcription*

The fundamental repeating unit of chromatin is the nucleosome (Kornberg, 1974; Kornberg and Thomas, 1974), which comprises 147 base pairs (bp) of DNA wrapped twice around an octamer of two copies of each of the four core

histones H2A, H2B, H3 and H4 (Luger et al., 1997). Histone proteins are highly conserved between species, and are made up of a core histone fold domain and an unstructured N-terminal tail region that extends from the surface of the nucleosome (Luger et al., 1997). The N-terminal tails of histones H2A and H4 protrude from the top and bottom surfaces of the nucleosome, whilst the H3 and H2B tails protrude between the gyres of DNA, with the tail of histone H3 emerging proximal to the entry/exit site of DNA, and that of H2B emerging distal to this site (Davey et al., 2002; Luger et al., 1997). Adjacent nucleosomes *in vivo* are separated by linker DNA, which varies in length between approximately 20 bp and 90 bp (Szerlong and Hansen, 2011). This linker DNA can be bound by a linker histone called histone H1, which comprises a central winged helix DNA binding domain, a short unstructured N-terminal region, and a longer and highly basic C-terminal domain (Allan et al., 1980; Ramakrishnan et al., 1993). The winged helix domain binds to the nucleosome core DNA at the DNA entry/exit site and the C-terminal domain makes contacts with linker DNA (Bednar et al., 2017), with these linker histone interactions thought to be important for the formation of higher order chromatin structures such as the 30 nm chromatin fibre (Robinson and Rhodes, 2006).

The chromatin structure of eukaryotic DNA reduces access to the underlying sequence, and therefore affects all processes that use DNA as a substrate, including DNA repair, recombination, and transcription. Chromatin structure has an inherently repressive effect on transcription. This was first demonstrated with observations from pea embryos showing that whilst free DNA is able to support transcription following incubation with extracts containing RNA polymerase enzymes, a chromatin template is unable to do so (Huang and Bonner, 1962). This was confirmed by *in vitro* transcription assays using reconstituted chromatin templates, which showed that the presence of nucleosomes inhibits transcriptional initiation (Knezetic and Luse, 1986; Lorch et al., 1987). This repressive effect is also apparent *in vivo*, where depletion of histones and nucleosomes in *Saccharomyces cerevisiae* is sufficient to activate expression of the gene *PHO5* (Han and Grunstein, 1988). These observations of the effect of nucleosome occupancy on transcription showed the importance of access to DNA in transcriptional regulation and suggested the existence of mechanisms by which the cell influences chromatin structure to regulate transcription.

Transcription in higher eukaryotes is carried out by three RNA polymerase enzymes (RNAPs), RNAPI, RNAPII and RNAPIII, which were first purified by high salt extraction from sea urchin embryos (Roeder and Rutter, 1969). RNAPI is dedicated to the transcription of large ribosomal RNA (rRNA) genes, and RNAPIII to the transcription of 5S rRNA, transfer RNA (tRNA) and other small RNAs (reviewed in Roeder, 2019). Protein-coding genes, by contrast, are transcribed by RNAPII through binding of the transcriptional machinery to regulatory elements.

Genes transcribed by RNAPII in mammals typically possess two distinct types of regulatory elements (Maston et al., 2006). These elements are promoters, which are proximal to the transcription start site (TSS) of a gene and comprise a core promoter along with promoter proximal elements, and distal regulatory elements, including enhancers, silencers and insulators. Transcription requires the assembly of RNAPII on the core promoter together with a pre-initiation complex of associated general transcription factors, including TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, which in total comprise some 44 different subunits (Roeder, 2019). Recruitment of RNAPII and the initiation of transcription are regulated by sequence-specific DNA binding proteins called transcription factors (Blau et al., 1996), which bind at promoter proximal elements and enhancers. Transcription factors can recruit the large multi-subunit complex called Mediator, which is thought to function at least in part by acting as a bridge between the promoter and distal enhancer regions to facilitate regulation by enhancers (Allen and Taatjes, 2015). In addition, transcription factors bound at promoters and enhancers depend on coactivators to stimulate transcription. A major mechanism by which coactivators function is to modify chromatin, allowing the transcriptional machinery to bind chromatin more readily and to transcribe the underlying DNA.

## 1.2 Post-translational modification of histone proteins

A subset of transcriptional coactivator proteins comprises enzymes that function to post-translationally modify proteins. One target of coactivator proteins is histones, which are subject to extensive post-translational modification in the cell in ways that are thought to influence transcription. The unstructured N-terminal tails of histones, and to a lesser extent the histone cores, have been found to undergo modification, including acetylation, phosphorylation, methylation, and ubiquitylation (Kouzarides, 2007).

There are several hypotheses as to how histone modifications function in transcriptional regulation. Negatively charged groups such as acetylation and phosphorylation could have a direct effect on chromatin structure by weakening the interactions between basic histones and acidic DNA, thereby allowing transcription factors and the transcriptional machinery greater access to the underlying DNA sequence (Zentner and Henikoff, 2013). A wider range of histone modifications can also be bound by specific domains within effector proteins, which can in turn function to influence transcription. The vast array of possible combinations of histone modifications at regulatory regions led to the formulation of the “histone code” hypothesis (Jenuwein and Allis, 2001; Turner, 2000), in which combinatorial recognition of sets of histone modifications could give rise to particular transcriptional outputs. However, it remains unclear to what extent many histone modifications are causative in regulating transcription rather than simply reflecting or reinforcing an existing transcriptional state (Henikoff and Shilatifard, 2011).

The following sections will describe in further detail what is known about the establishment and function of the major histone modifications that will be referred to in this thesis, namely histone lysine acetylation and histone lysine methylation.

## 1.3 Histone lysine acetylation

### *1.3.1 Histone acetylation is associated with active transcription*

Acetylation of lysine residues in histone proteins was first demonstrated by Vincent Allfrey in 1964 (Allfrey et al., 1964) by labelling newly incorporated acetyl groups in cells with $^{14}C$. Histone acetylation is generally considered to be associated with active transcription, and this connection was already made in 1964 as Allfrey also showed that acetylation reduced the inhibitory effect of histones on transcription *in vitro* (Allfrey et al., 1964). A correlation was found between histone acetylation and gene transcription in cells by the observation that regions of DNA that are preferentially digested by DNaseI, and are therefore thought to be more accessible in the nucleus, are associated both with both active gene sequences and acetylated histones (Sealy and Chalkley, 1978; Vidali et al., 1978). Further work by Allfrey taking advantage of residue cysteine-110 in the core of histone H3, which is exposed in accessible but not inaccessible chromatin (Prior et al., 1983), showed that active gene sequences could be bound to an organo-mercury column and that associated histones were also acetylated (Allegra et al., 1987; Chen and Allfrey, 1987). Early chromatin immunoprecipitation (ChIP) experiments using antibodies recognizing acetylated histone H4 showed that histone acetylation was found at the actively transcribed α-D globin gene in chicken embryo erythrocytes but not at the inactive ovalbuminin gene (Hebbes et al., 1988). These results clearly indicated a link between histone acetylation and active transcription.

All four core histone proteins can be acetylated, with acetylation occurring primarily in the unstructured N-terminal tails, but also at several lysine residues in the histone core domains (reviewed in Shahbazian and Grunstein, 2007). The best-studied targets of histone acetylation are histones H3 and H4. H3 acetylation has been detected at numerous residues, including H3K9, H3K14, H3K18, H3K23 and H3K27 in the N-terminal tail region (Shahbazian and Grunstein, 2007), and H3K56, H3K64 and H3K122 in the globular domain (Di Cerbo et al., 2014; Masumoto et al., 2005; Tropberger et al., 2013). The histone H4 tail is acetylated at H4K5, H4K8, H4K12 and H4K16, and the globular domain of H4 is acetylated at H4K91 (Ye et al., 2005). These histone

acetylation modifications are placed by multiple histone acetyltransferase (HAT) enzymes, which will be discussed in more detail below.

The N-terminal tails of histones H3 and H4 are not essential in *S. cerevisiae*, suggesting that cells can live in the absence of histone tail acetylation (Durrin et al., 1991; Mann and Grunstein, 1992). However, histone acetylation is thought to be closely and causally linked to transcription, with deletion of the N-terminal tail of histone H4 leading to greatly reduced induction of gene expression (Durrin et al., 1991). Moreover, experiments in which the tails of H3 and H4 are swapped *in vivo* suggest that the two histone tails have overlapping but non-redundant roles in regulating gene expression in budding yeast (Ling et al., 1996). Consistent with the possibility that different histone tails have specific functions, certain histone acetylation marks are thought to have distinct roles in gene expression regulation. H4K16ac is thought to play a unique role in opposing the formation of higher order chromatin structures that would be expected to impact gene accessibility and expression (Shogren-Knaak et al., 2006). Moreover, H3K27ac is thought to distinguish active from inactive regulatory elements in mammalian cells (Creyghton et al., 2010; Rada-Iglesias et al., 2011). Indeed, mapping of H3K27ac, and of the CBP/p300 enzyme that places this mark, has been used to accurately predict novel functional enhancers (Heintzman et al., 2007, 2009). This suggests that H3K27ac might also have a unique function in allowing the cell to discriminate between genes that should be active or inactive. An alternative possibility, however, is that compared to other acetylation marks, H3K27ac can be mapped more closely to active regulatory elements because of a better availability of specific, ChIP grade antibodies against this modification. Nevertheless, work to understand how H3K27ac, which clearly correlates with gene regulatory elements, is regulated and what role this modification plays in the acetylation landscape will have a profound impact on our understanding of how chromatin modifications influence gene expression.

*1.3.1.1 Histone acetylation: charge neutralisation model*

Several models have emerged to provide a mechanistic basis for the correlation between histone acetylation and transcriptional activity (Zentner and Henikoff, 2013). As mentioned above, in the first model acetylation of positively charged lysine side chains weaken the interaction between histone proteins and negatively charged DNA, allowing the DNA to become more accessible to transcription factors and the transcriptional machinery. Consistent with this model in which charge neutralisation enhances DNA accessibility, reconstitution of chromatin using acetylated histones has been shown to facilitate access for the DNA-binding transcription factors USF, HSF and TFIIIA to their cognate sequences (Lee et al., 1993; Nightingale et al., 1998; Vettese-Dadey et al., 1996). Moreover, mutation of lysine residues in the N-termini of H3 and H4 in *S. cerevisiae* shows that whilst these lysines are important for upregulation of transcription, individual residues within the same tail are largely interchangeable and function cumulatively (Dion et al., 2005; Martin et al., 2004). Furthermore, acetylation of the histone H3 globular domain at H3K64 and H3K122, residues located on the lateral surface of the histone octamer close to DNA, facilitates nucleosome eviction and mobilisation by ATP-dependent chromatin remodelling proteins (Di Cerbo et al., 2014; Tropberger et al., 2013), consistent with a structural role for histone acetylation in transcriptional regulation.

*1.3.1.2 Histone acetylation: antagonism with methylation*

A second model for how at least some forms of histone acetylation function to regulate transcription is by competing with alternative histone modifications, particularly methylation. Several histone lysine residues, including H3K9, H3K27 and H3K64 can undergo methylation as well as acetylation (Daujat et al., 2009; Kouzarides, 2007), and methylation and acetylation are mutually exclusive at the same site on the same histone tail. Trimethylation of these residues (H3K9me3, H3K27me3, H3K64me3) is associated with the repression of gene expression, with H3K9me3 and H3K64me3 found at transcriptionally inactive regions such as pericentromeric heterochromatin (Daujat et al., 2009; Lachner et al., 2001; Lange et al., 2013; Nakayama et al., 2001; Rea et al.,

2000), and H3K27me3 found at genes that are repressed by the Polycomb repressive system (Boyer et al., 2006; Schwartz et al., 2006). One possibility, therefore, is that acetylation of these sites facilitates active transcription by preventing repression through methylation.

Such a competition model is consistent with observations of an antagonism between H3K27me3 and H3K27ac *in vivo*, such that genes switch between acetylation and methylation of H3K27 depending on their transcriptional status (Pasini et al., 2010). Nevertheless, mutation of the H3K27 residue to arginine in *Drosophila* results in derepression of Polycomb target genes such as *Hox* genes, leading to homeotic transformations that phenocopy mutations in repressive Polycomb proteins, but does not lead to phenotypes that might be predicted to arise from loss of transcriptional activation caused by the absence of H3K27ac (Pengelly et al., 2013). However, these observations from H3K27R mutations would also be consistent with a model in which, once transcription has been established by binding of transcription factors, H3K27ac is not absolutely required for its continued maintenance, and in which there is a degree of redundancy between histone acetylation marks, such that acetylation of other histone residues can compensate for the loss of H3K27ac. Moreover, the lack of a phenotype associated with loss of transcriptional activation in H3K27R mutant cells does not preclude the possibility of more subtle phenotypes that cannot be detected at the level of gross *Drosophila* anatomy. Together these results, although suggestive of a correlative antagonism between repressive methylation and active acetylation, do not clearly indicate a mechanistic link between prevention of methylation and activation of transcription. Furthermore, this model cannot account for the correlation of other acetylated histone residues, which are not otherwise methylated, with active gene expression.

#### *1.3.1.3 Histone acetylation: binding by bromodomains*

A third, though not mutually exclusive, model postulates that histone acetylation functions as part of a histone code by directly recruiting transcriptional effector proteins through domains that specifically bind histone acetylation. Histone

acetylation can be bound by a specific domain called the bromodomain (Dhalluin et al., 1999; reviewed in Marmorstein and Zhou, 2014). Bromodomains comprise a left-handed four-helix bundle in which inter-helical loops form a hydrophobic pocket that binds the acetyl-lysine residue with moderate affinity, with typical dissociation constants ($K_d$) in the range of 10-1000 μM (Filippakopoulos and Knapp, 2012).

Bromodomains are found in a large number of chromatin-binding proteins, including components of the core transcriptional machinery, chromatin remodelling proteins, and in HATs, the enzymes that place histone acetylation. The TFII250 subunit of the TFIID general transcription factor contains a pair of bromodomains arranged in tandem, with the binding pockets of the two bromodomains arranged such that each domain is thought to bind one acetyl-lysine residue in diacetylated histone tail peptides (Jacobson et al., 2000). Indeed, whilst H4 tail peptide singly acetylated at H4K16 is bound by the tandem bromodomains with $K_d$ of approximately 40 μM, doubly acetylated H4 peptides bind with affinity in the range of 1-6 μM (Jacobson et al., 2000). These results suggest that bromodomain binding to histone acetylation could recruit or stabilise the binding of the transcriptional machinery at acetylated gene regulatory elements.

Chromatin remodellers are ATPase enzymes that use the energy from ATP hydrolysis to slide, evict or exchange nucleosomes (Becker and Workman, 2013). The SWI/SNF family of chromatin remodellers, which includes BRM and BRG1 in humans, are characterised by the presence of a C-terminal bromodomain, and bind acetylated lysine tails, with some preference for H3K14ac (Morrison et al., 2017; Shen et al., 2007; Singh et al., 2007). Another role for histone acetylation, therefore, could be to increase the affinity of nucleosome remodelling proteins for substrate nucleosomes. This would aid nucleosome remodelling or depletion at regulatory elements and facilitate transcription factor binding to accessible DNA.

Bromodomains are also found in several HAT enzymes, including the enzymes GCN5 and its paralogue PCAF, as well as CBP and its paralogue p300 (Marmorstein and Zhou, 2014). The bromodomain of GCN5 binds H4 tail

peptides modified with H4K16ac, and binding of the bromodomain stabilises the HAT on acetylated nucleosome array substrates (Hassan et al., 2002; Owen et al., 2000). The CBP/p300 bromodomains have been shown to bind numerous acetylated lysines in both H3 and H4 tails (Filippakopoulos and Knapp, 2012). Bromodomains in HATs may therefore function to recruit these enzymes to sites that have previously been marked by acetylation, reinforcing a positive feedback loop.

The importance of acetyl-lysine binding has recently been highlighted by the emergence of bromodomains as key targets for pharmacological inhibition (Marmorstein and Zhou, 2014). JQ1, a potent and selective inhibitor of BET (bromodomain and extra-terminal) family bromodomains, reduces the binding of the tandem bromodomains of proteins such as BRD4 to acetylated lysines and chromatin, and has dramatic anti-proliferative effects in acute myeloid leukaemia (AML) (Filippakopoulos et al., 2010; Zuber et al., 2011). Recent work has shown that inhibition of the bromodomain of CBP/p300 leads to a reduction in histone acetylation *in vitro* and *in vivo*, and could be an effective therapy in castration-resistant prostate cancer (Jin et al., 2017; Raisner et al., 2018). These results suggest that bromodomain binding to histone acetylation could be an important mechanism by which acetylation regulates gene expression.

Nevertheless, it remains unclear to what extent acetyl-lysine binding by bromodomains contributes to histone acetylation function. The affinity of bromodomains for acetyl-lysine is relatively poor, and proteins such as GCN5 appear to depend on transcription factors for recruitment (Kuo et al., 2000). In addition, despite reports that the CBP/p300 bromodomain binds to multiple acetylated lysines in histone peptides (Filippakopoulos and Knapp, 2012), *in vitro* pull down assays using nucleosome arrays show that p300 binding is not enhanced by acetylation of H4, and is not reduced by mutation of the acetyl-lysine binding pocket or by a bromodomain-targeting compound (Raisner et al., 2018). This raises the possibility that bromodomains may have important functions in addition to binding of histone acetylation. One possibility is that bromodomains bind acetyl-lysine residues in non-histone substrates. Indeed the bromodomain of CBP/p300 can bind to an acetylated form of the transcription factor and tumour suppressor protein p53 (Mujtaba et al., 2004). Alternatively,

there is growing evidence to suggest that some bromodomains have additional roles beyond binding acetyl-lysine. The bromodomain of p300 contributes to chromatin association regardless of the acetylation status of nucleosomes (Manning et al., 2001). The bromodomains of the BRM/BRG1 chromatin remodellers bind to DNA in a sequence-independent manner (Morrison et al., 2017). Similarly, bromodomains of BET family proteins including BRDT, BRD2, BRD3 and BRD4 bind to DNA independent of sequence via a positively charged surface that is common to many bromodomains and distinct from the acetyl-lysine interaction pocket (Miller et al., 2016). These bromodomain-DNA interactions are important both for *in vitro* interactions with nucleosomes and for chromatin interactions *in vivo*, suggesting that chromatin binding by bromodomains is not solely dependent on binding to acetylated histones (Miller et al., 2016).

Together, these results suggest that acetyl-lysine binding by bromodomains represents a mechanism by which histone acetylation function is mediated. However, the relatively low affinity of these interactions, together with the reported degeneracy of many bromodomains in terms of specificity of binding (Filippakopoulos and Knapp, 2012), suggests that histone acetylation is unlikely to represent a highly specific code that mediates protein recruitment, and may rather play a role in stabilising protein binding through multivalent interactions with chromatin. Nevertheless, the clear therapeutic value of bromodomain inhibitors suggests that these domains play an important role *in vivo*, although it has not been conclusively shown that such inhibitors function primarily through the inhibition of histone acetyl-lysine binding rather than through other mechanisms that might affect additional bromodomain functions. The precise contributions, therefore, of the three models described here for histone acetylation function remain to be fully elucidated.

### *1.3.2 Histone acetyltransferase (HAT) enzymes*

There are three broad families of HAT enzymes in eukaryotes, classified according to structural similarity: the GCN5-related histone *N*-acetyltransferases (GNAT) family, the MYST family, and the CBP/p300 family (Fig. 1.1). (Marmorstein and Zhou, 2014). The following sections will briefly describe members of these families, their substrate specificities and roles, particularly in relation to regulation of gene expression.

#### *1.3.2.1 GNAT family*

The first GNAT family member to be identified and cloned was HAT1 from *S. cerevisiae* (Kleff et al., 1995), which was later found to be a cytoplasmic HAT important for acetylation of newly synthesised histones prior to their deposition on chromatin (Parthun et al., 1996). However, the major step forward in the understanding of HAT function in transcriptional activation came from the purification of a HAT enzyme from macronuclei of the ciliated protozoan *Tetrahymena thermophila* (Brownell and Allis, 1995). This *Tetrahymena* HAT was found to be a homologue of the *S. cerevisiae* protein GCN5 (general control non-derepressible 5) (Brownell et al., 1996), which had previously been identified as a transcriptional co-activator in budding yeast (Marcus et al., 1994). This provided the first mechanistic link between a HAT enzyme and transcriptional activation, and established GCN5 as the founder member of the GNAT family of HATs (Fig. 1A).

Crystal structures of the HAT1 and GCN5 catalytic domains have been solved, showing a shared core structure comprising a three-stranded β-sheet and a long parallel α-helix, with divergent flanking regions that may account for the different substrate preferences of the two enzymes (Marmorstein and Zhou, 2014). GCN5 uses a ternary complex mechanism for acetylation, in which acetyl coenzyme A (acetyl-CoA) and the substrate lysine residue are co-bound in the enzyme active site. A conserved glutamate residue in the catalytic site of GCN5 deprotonates the substrate lysine side chain, facilitating direct transfer

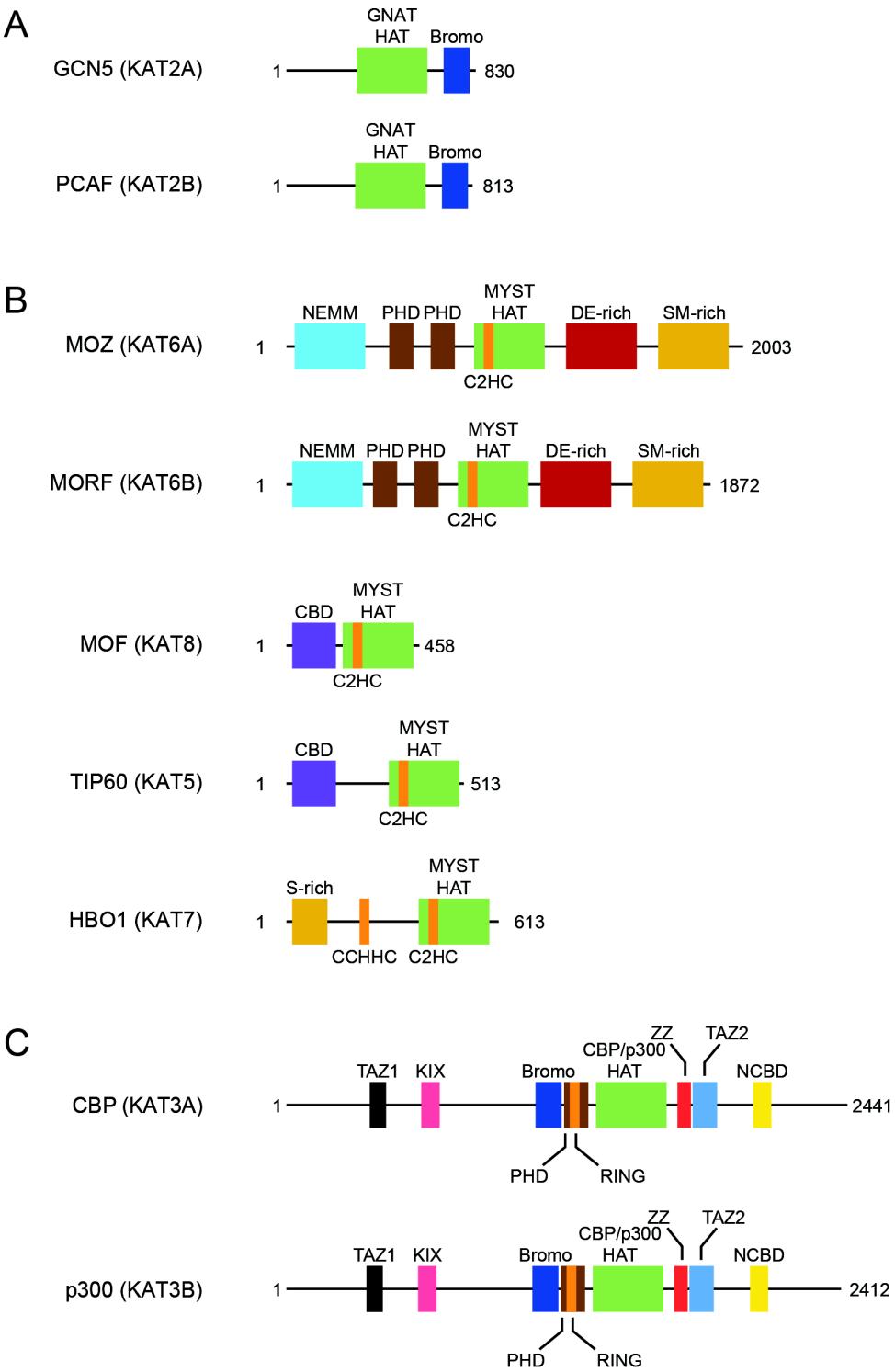



**Fig. 1.1: Mammals have three families of histone acetyltransferase enzymes.**
**(A)** Domain architecture of the GCN5 *N*-acetyltransferase HAT family in mice.
**(B)** Domain architecture of the MYST HAT family.
**(C)** Domain architecture of the CBP/p300 HAT family.
Abbreviations: bromodomain (Bromo); C2HC-type zinc finger (C2HC); chromobarrel domain (CBD) CCHHC-type zinc finger (CCHHC); aspartate/glutamate-rich region (DE-rich); histone acetyltransferase domain (HAT); kinase-inducible domain (KID)-interacting domain (KIX); nuclear coactivator binding domain (NCBD); N-terminal part of Enok, MOZ or MORF domain (NEMM); plant homeodomain (PHD); RING finger (RING); serine-rich region (S-rich); serine/methionine-rich region (SM-rich); transcription adaptor zinc finger domain (TAZ); ZZ-type zinc finger (ZZ).

of an acetyl group from the co-bound acetyl-CoA (Tanner et al., 1999; Trievel et al., 1999).

The budding yeast GCN5 protein acetylates histones (Kuo et al., 1996) and functions as part of the multi-subunit SAGA (Spt-Ada-GCN5 acetyltransferase) complex, with incorporation into this complex required for acetylation of nucleosomes (Grant et al., 1997). Mammalian GCN5, and its homologue PCAF (p300/CBP-associated factor), also form part of SAGA-like complexes (Martinez et al., 1998; Ogryzko et al., 1998). *In vivo*, GCN5/PCAF primarily acetylates H3K9 (Feller et al., 2015; Gates et al., 2017; Jin et al., 2011), and binds to most active genes genome-wide (Krebs et al., 2011; Wang et al., 2009). The SAGA complex is responsible for H3K9ac at the majority of active genes in both *S. cerevisiae* and human cells, and loss of GCN5 leads to reduced transcription at all tested genes in *S. cerevisiae* (Baptista et al., 2017; Bonnet et al., 2014). GCN5/PCAF complexes are thought to be recruited to target sites through interactions with transcription factors (Brown et al., 2001; McMahon et al., 1998). In budding yeast, GCN5 is targeted to promoters through interaction with transcription factors such as GCN4 (Kuo et al., 2000). Moreover, the SAGA subunit TRRAP (Tra1 in budding yeast) interacts with sequence-specific DNA binding proteins including c-Myc and E1A in mammals, and is required for transformation of fibroblasts by c-Myc and E1A (McMahon et al., 1998). This indicates that interaction with SAGA is required for transcription factors to execute their gene expression programme.

The importance of GCN5 in transcriptional regulation is underlined by observations that loss of GCN5 is lethal in mice, although PCAF mutants develop normally as a result of compensation by GCN5 (Xu et al., 2000; Yamauchi et al., 2000). Moreover, haematological malignancies driven by transcription factor fusion proteins, such as the MLL-AF9 fusion protein in acute myeloid leukaemia (AML), are dependent on GCN5 for proliferation, so that GCN5 could represent a valid pharmacological target in these cells (Tzelepis et al., 2016).

*1.3.2.2 MYST family*

The MYST (named for the founder members MOZ, Ybf2, Sas2 and TIP60) family of acetyltransferases includes the mammalian HATs MOZ and its paralogue MORF, MOF, TIP60, and HBO1 (Fig. 1.1B) (Marmorstein and Zhou, 2014). Structural elucidation of MYST family HATs showed a similar core structure to that of the GNAT family, with a conserved glutamate residue for lysine deprotonation (Yan et al., 2000). However, kinetic analysis of enzymatic activity demonstrated that MYST HATs catalyse acetylation via a different mechanism from GCN5, using a ping-pong mechanism in which a conserved cysteine residue forms an acetylated intermediate before transfer of the acetyl group to the substrate lysine (Yan et al., 2002).

MYST HATs acetylate a range of histone substrate residues. MOZ/MORF acetylates H3K23 in small cell lung cancer cells (Simó-Riudalbas et al., 2015). HBO1 also acetylates H3K23, together with H3K14 and the N-terminal tail of histone H4 (Feng et al., 2016; Lalonde et al., 2013; MacPherson et al., 2019). MOF acetylates H4K16, and its *Drosophila* homologue plays a key role in the MSL dosage compensation complex in flies (Rea et al., 2007). TIP60 forms part of the NuA4 complex, which is conserved from yeast to humans, and in budding yeast primarily acetylates the histone H4 tail (Allard et al., 1999; Suka et al., 2001). Similarly to GNAT HATs, MYST proteins such as MOF and TIP60 are recruited to promoters throughout the genome, with this general targeting of TIP60 potentially mediated by physical interactions with RNAPII (Wang et al., 2009). However, in budding yeast NuA4 is also specifically recruited to target promoters by transcription factors such as Hsf1 in response to heat shock, to facilitate acetylation and transcriptional activation of target genes (Reid et al., 2000). Such targeted recruitment is at least in part mediated by adaptor proteins such as TRRAP, which is a component of the NuA4 complex as well as SAGA (Brown et al., 2001).

MYST proteins have been implicated in cancer, particularly in leukaemias (Yang, 2004). Translocations between the MOZ locus and both the CBP and p300 loci can generate MOZ-CBP and MOZ-p300 fusion proteins that contain two HAT domains and are important in driving leukaemogenesis (Borrow et al.,

1996). Recent work has shown that HBO1 is required for leukaemia stem cell proliferation, and inhibition of HBO1 reduces H3K14ac and AML growth, so that HBO1 could be a feasible therapeutic target (MacPherson et al., 2019).

#### *1.3.2.3 CBP/p300 family*

p300 (E1A-associated 300 kDa protein) was first identified as an interaction partner of the adenoviral oncogenic transcription factor E1A (Whyte et al., 1989; Yee and Branton, 1985) and later found to have the properties of a transcriptional co-activator (Eckner et al., 1994). CBP (CREB-binding protein) was similarly identified as a co-activator protein that interacts with the phosphorylated form of the transcription factor CREB (cAMP response element binding protein) (Chrivia et al., 1993), and CBP and p300 were later found to be paralogues (Arany et al., 1994). Later work demonstrated that CBP and p300 are HAT enzymes, and that their HAT activity is required for transcriptional co-activation (Bannister and Kouzarides, 1996; Martinez-Balbás et al., 1998; Ogryzko et al., 1996)

CBP/p300 represent a distinct family of HATs (Fig. 1.1C), with little sequence similarity with GNAT or MYST enzymes (Dancy and Cole, 2015). Unlike GNAT and MYST HATs, which have homologues from *S. cerevisiae* to humans, CBP/p300 are found almost exclusively in metazoans (Marmorstein and Zhou, 2014), although examples of CBP/p300 homologues have been identified in protists such as *Capsaspora owczarzaki*, a close unicellular relative of animals (Sebé-Pedrós et al., 2011).

The importance of CBP/p300 in animal development is shown by observations that homozygous knockouts of either CBP or p300 are embryonic lethal in mice, with lethality occurring between E8.5 and E10.5 (Oike et al., 1999; Tanaka et al., 1997; Yao et al., 1998). This also suggests that, despite their similarity, CBP and p300 cannot entirely compensate for one another *in vivo*. Germline mutation of one allele of either CBP or p300 leads to the rare genetic disorder Rubinstein-Taybi Syndrome (RSTS) (Hennekam, 2006; Petrij et al., 1995). RSTS is characterised by clinical features including intellectual disability,

specific facial characteristics, and markedly broad thumbs and halluces (Hennekam, 2006). The association of RSTS with haploinsufficiency in CBP or p300 suggests that dosage of these genes is important in development, consistent with observations that amounts of CBP/p300 are limiting in cells (Kamei et al., 1996). Moreover, the lack of homozygous mutations in either CBP or p300 in patients suggests that total loss of either protein is incompatible with life.

Consistent with the divergence in sequence between CBP/p300 and the GNAT and MYST families, the HAT domain of CBP/p300 also has a variant structure (Delvecchio et al., 2013; Liu et al., 2008). Crystal structures of the HAT domain reveal that it comprises a central seven-stranded β-sheet surrounded by nine α-helices. The core of this domain contains a region that is structurally similar to other HAT domains, and functions in acetyl-CoA binding. The CBP/p300 HAT domain is preceded by a bromodomain and an unusual plant homeodomain (PHD)-RING finger, which interact closely with the HAT domain and form a catalytic core that is required for acetylation function (Delvecchio et al., 2013; Kalkhoven et al., 2002; Park et al., 2017). CBP/p300 has been proposed to catalyse acetylation via a Theorell-Chance “hit-and-run” ternary mechanism, which is distinguished from the ternary mechanism utilised by GNAT proteins in that, whilst an enzyme:substrate:acetyl-CoA ternary complex is generated as part of catalysis, this complex does not accumulate as a stable intermediate (Liu et al., 2008).

CBP/p300 catalytic activity is regulated through several distinct mechanisms. First, the HAT domain contains a highly basic loop that is thought to fold back into an electronegative pocket in the catalytic site and inhibit the activity of the enzyme (Liu et al., 2008; Thompson et al., 2004). Trans-autoacetylation of this loop neutralises its positive charge and relieves inhibition (Liu et al., 2008; Ortega et al., 2018). Autoregulation is also achieved via a mechanism involving the PHD-RING finger of CBP/p300 (Delvecchio et al., 2013; Ortega et al., 2018). The PHD and RING domains of CBP/p300 adopt non-canonical structures. PHD fingers are zinc finger domains that frequently contain an aromatic cage and function by binding to methylated lysine residues, particularly H3K4 (Fortschegger and Shiekhattar, 2011). The aromatic cage is

absent in the CBP/p300 PHD structure, and the PHD finger is interrupted by the insertion of the RING finger within a loop (Delvecchio et al., 2013). The RING finger is also non-canonical and contains only one zinc ion rather than the two that are typical for RING-type zinc fingers. Furthermore, the CBP/p300 RING finger possesses an extended loop that forms electrostatic interactions with the substrate-binding loop of the HAT domain and inhibits substrate binding. Interaction with activating transcription factors is thought to result in structural rearrangement of the RING finger that facilitates substrate access to the catalytic site (Delvecchio et al., 2013; Ortega et al., 2018).

Finally, it has also been proposed that CBP/p300 is regulated by binding of enhancer RNAs (eRNAs) (Bose et al., 2017), short bi-directional transcripts that arise from enhancer regulatory regions. CBP was found to crosslink to RNA in cells, with bound RNAs mapping to CBP binding sites at enhancers. Binding experiments suggested that RNA interacts with the positively charged autoinhibitory loop of CBP, so that addition of RNA to enzymatic reactions resulted in a small increase in catalytic activity, potentially through displacement of this loop from the catalytic site of the enzyme (Bose et al., 2017). However, the role of RNAs in regulation of CBP/p300 remains controversial, as later work suggested that such regulation was an artefact resulting from denaturation of the zinc fingers within CBP due to inclusion of the metal ion chelating agent EDTA in enzyme purification buffers (Ortega et al., 2018).

CBP/p300 has been shown to acetylate a wide range of histone substrates *in vitro*, including the N-terminal tails of H3 and H4, and the H3K64 and H3K122 residues in the globular domain of H3 (An et al., 2002; Di Cerbo et al., 2014; Ogryzko et al., 1996; Tropberger et al., 2013). *In vivo*, CBP/p300 acetylates globular domain H3 residues (Di Cerbo et al., 2014; Tropberger et al., 2013), but the enzymes appear to have more narrow specificity towards histone tail than typically found *in vitro*, specifically acetylating H3K18, H3K27 and the N-terminal tail of H2B (Jin et al., 2011; Pasini et al., 2010; Tie et al., 2009; Weinert et al., 2018). In addition to histone substrates, CBP/p300 acetylates a wide range of non-histone proteins (Dancy and Cole, 2015; Weinert et al., 2018). Non-histone substrates include p53, and acetylation appears to be required for full p53-mediated transactivation (Grossman, 2001), with mutation of CBP

contributing to a defective p53-mediated DNA damage response in progenitors of lymphoma (Horton et al., 2017). However, the mechanisms that underlie histone acetylation specificity *in vivo*, and the importance of non-histone substrate acetylation to CBP/p300 function, remain incompletely understood.

CBP/p300 are bound at promoter and enhancer regions genome-wide (Krebs et al., 2011; Wang et al., 2009), and the enzymes are recruited by a wide range of transcription factors, interacting directly with nuclear receptors such as the estrogen receptor (ER), the glucocorticoid receptor (GR) and the retinoic acid receptor (RAR), as well as dimeric transcription factors such as IRF3, STAT1 and AP-1 (Kamei et al., 1996; Ortega et al., 2018). CBP/p300 is also thought to play a scaffolding role in transcriptional regulation, in which multivalent interactions are formed with transcription factors and RNAPII to enhance recruitment of the transcriptional machinery to regulatory elements (Kim et al., 1998). Importantly, however, a scaffolding role is not sufficient to account for CBP/p300-mediated transcriptional activation, as catalytic activity is required for activation of transcription from chromatin templates *in vitro* (Lu et al., 2002). Recent work has shown that a domain downstream of the catalytic domain, a zinc finger called the ZZ domain, interacts directly with the N-terminal tail of histone H3 (Zhang et al., 2018). This report suggests that this interaction is important not only for the catalytic activity of p300 but also for recruitment of the protein genome-wide. It is unclear, however, how such a general histone-binding activity could generate a specific pattern of CBP/p300 recruitment across the genome, although the ZZ domain may play a role in stabilising CBP/p300 at target sites to which it is recruited by transcription factors.

CBP/p300 play multiple roles in cancers (Di Cerbo and Schneider, 2013; Iyer et al., 2004; Yang, 2004). CBP and p300 loss of function mutations are found in a range of cancers, including gastric, breast and colorectal cancer (Gayther et al., 2000; Muraoka et al., 1996), suggesting that CBP/p300 can act as tumour suppressor proteins. As described above, translocations in leukaemia generate MOZ-CBP and MOZ-p300 fusion proteins (Borrow et al., 1996). CBP and p300 translocations also generate fusion proteins in which the catalytic core and C-terminus of CBP/p300 is fused to the DNA binding domain of the protein MLL1, with MLL1-mediated DNA binding thought to drive aberrant gene activation by

CBP/p300 at target sites (Yang, 2004). Multiple transcription factors that drive tumourigenesis, including ER in breast cancer and the androgen receptor (AR) in prostate cancer, interact with CBP and p300 (Ianculescu et al., 2012; Mohammed et al., 2013; Papachristou et al., 2018). p300 is upregulated in late stage prostate cancers and in response to anti-androgen therapy (Comuzzi et al., 2004; Debes et al., 2003), and recent results suggest that inhibition of CBP/p300 through targeting of the bromodomain represents a promising avenue for therapy in castration-resistant prostate cancer (Jin et al., 2017). Together, these results suggest that CBP/p300 and the catalytic activity of these enzymes play an important role in processes including development and disease. Therefore, work to further our knowledge of how CBP/p300 function is vital both for a deeper understanding of the fundamental process of transcriptional regulation, but also to develop new therapies in disease.

### *1.3.3 Histone deacetylases (HDACs)*

Histone acetylation is a highly dynamic process in living cells (reviewed in Clayton et al., 2006). Indeed, Allfrey's experiments identified histone acetylation through the dynamic incorporation of new acetyl modifications (Allfrey et al., 1964). Histone acetylation can also be dynamically removed by a set of enzymes called histone deacetylases (HDACs) (Seto and Yoshida, 2014).

The first HDAC activity was identified from calf thymus extract in 1969 (Inoue and Fujimoto, 1969). However, it was not until 1996 that an HDAC was successfully cloned and isolated, with the identification of HDAC1 (Taunton et al., 1996). There are now 18 HDACs that have been identified in mammals, which are divided into four classes based on sequence and structural similarity.

Class I HDACs (HDAC1, HDAC2, HDAC3 and HDAC8) are those that display sequence similarity to the budding yeast protein Rpd3. Class II HDACS (HDAC4, HDAC5, HDAC6, HDAC7, HDAC9 and HDAC10) are homologous to the yeast protein Hda1. Class III enzymes (SIRT1-7) are related to yeast Sir2. A single Class IV HDAC (HDAC11) shares similarity with both Class I and Class II enzymes (Seto and Yoshida, 2014). These classes fall into two broad families.

Class I, II and IV HDACs utilize a common mechanism for the hydrolysis of the acetyl-lysine acetamide bond that is dependent on a zinc ion, whereas Class III sirtuin enzymes require $NAD^+$ as a cofactor.

Multiple studies have attempted to determine whether individual HDAC enzymes possess substrate specificity, although the results remain controversial. *In vivo* studies to define specific substrates are complicated by redundancy between homologous HDAC proteins, and *in vitro* studies are hindered by the inclusion of HDAC proteins in multi-subunit complexes that influence their activity (Seto and Yoshida, 2014). HDAC1/2 proteins, for example, are part of three distinct complexes called the Sin3 complex, the Co-REST complex, and the NuRD (nucleosome remodelling and histone deacetylase) complex, which also contains the chromatin remodelling proteins CHD3 or CHD4. To illustrate the difficulties in determining substrate specificity, a study examining the specificities of complexes containing either HDAC1/2 or HDAC3 in a reconstituted chromatin system suggested that the HDAC1/2 complexes can deacetylate both H3 and H4, whereas HDAC3 selectively deacetylates histone H3 (Vermeulen et al., 2004). This is in contrast with previous work suggesting that HDAC3 can deacetylate H4K5 and H4K12 in free histone and mononucleosome substrates (Johnson et al., 2002). The substrate specificity of Class III sirtuin HDACs, by contrast, is clearer, with the major targets of yeast Sir2 and SIRT1 being H3K9ac and H4K16ac (Imai et al., 2000; Vaquero et al., 2004).

HDACs are thought to work primarily as negative regulators of transcription through the removal of activation-associated histone acetylation. Indeed, tethering HDAC2 to a reporter gene via fusion to a GAL4 DNA binding domain led to repression of a reporter gene *in vivo* (Yang et al., 1996). However, earlier work showed that mutation of yeast Rpd3 leads to pleiotropic effects consistent with both activation and repression of target genes (Vidal and Gaber, 1991). In mammals, induction of gene expression with TPA in the presence of HDAC inhibitors leads to accumulation of histone acetylation at promoter elements but, surprisingly, to reduced gene expression compared to cells that are not treated with HDAC inhibitors (Hazzalin and Mahadevan, 2005). These results suggest that turnover of histone acetylation is highly dynamic, consistent with

observations that HDAC inhibition leads to rapid accumulation of histone acetylation and that HDACs are recruited to acetylated sites that are co-occupied by HATs genome-wide (Hazzalin and Mahadevan, 2005; Wang et al., 2009). This further suggests that, at least for some rapidly inducible genes such as immediate-early genes, histone deacetylation is important for activation of gene expression. One model to explain this would be that inhibition of HDAC activity leads to increased histone acetylation at regulatory sites throughout the genome, leading to increased accessibility that reveals cryptic transcription factor binding sites. This could lead to redistribution of transcription factors to non-target sites and therefore to repression of *bona fide* target genes by so-called “squelching” effects (see Gill and Ptashne, 1988; Meyer et al., 1989 for examples of negative gene regulation by transcriptional activators through squelching). This underlines the importance of histone acetylation, and regulation of chromatin accessibility, in control of gene expression.

## 1.4 Histone lysine methylation

Methylation of histone lysines was first identified in 1964 (Murray, 1964), and since that time a large number of histone lysine methylation sites have been identified (reviewed in Greer and Shi, 2012). Lysine residues can undergo mono-, di-, or trimethylation (me1, me2, me3), with lysine methylation generally carried out by the SET (Su(var) 3-9/E(z)/Trx) domain of histone methyltransferase (HMT) enzymes (Dillon et al., 2005; Herz et al., 2013; Rea et al., 2000). Unlike lysine acetylation, lysine methylation does not alter the charge of the amino acid side chain, and is therefore thought to function indirectly, primarily through recruitment or blocking of effector proteins (Taverna et al., 2007). Consistent with this possibility, numerous methyl-lysine binding domains have been identified, including chromodomains, PHD fingers, WD40 domains, MBT repeats and PWWP domains, that could recruit chromatin binding proteins to sites of histone lysine methylation (Taverna et al., 2007).

The best-studied histone lysine methylation sites are in the N-terminal tail of histone H3, at H3K4, H3K9, H3K27, and H3K36. H3K9 methylation is generally associated with constitutively repressed regions of the genome, such as those

found at pericentromeric heterochromatin (Dillon et al., 2005). H3K9me2/3 is thought to function through recruitment of proteins such as heterochromatin protein 1 α (HP1α), which binds to H3K9 methylation via its chromodomain (Bannister et al., 2001; Lachner et al., 2001) and is thought to condense the chromatin structure at heterochromatic regions. H3K36me2 is pervasive throughout the genome in mammals, with the exception of gene promoters where it is generally depleted, and H3K36me3 is deposited on the bodies of active genes (Mikkelsen et al., 2007). This inter- and intragenic H3K36me2/3 functions through suppression of spurious transcription initiation from cryptic transcription start sites through recruitment of the Rpd3S HDAC complex via the chromodomain of complex component Eaf3 (Carrozza et al., 2005). The following two sections will focus on methylation of H3K4 and H3K27, which are generally associate with active and repressed gene promoters, respectively, and on the proteins that place and bind to these marks.

## 1.5 H3K4 methylation and methyltransferases

### *1.5.1 H3K4 methyltransferase enzymes*

*S. cerevisiae* possesses a single H3K4 methyltransferase enzyme called Set1 (Roguev et al., 2001), which can catalyse all three H3K4 methylation states. *Drosophila* has three enzymes homologous to Set1, called Set1, Trithorax (Trx) and Trithorax-related (Trr) (Shilatifard, 2012). Mammals possess a total of six Set1 homologues: SETD1A and SETD1B are homologous to *Drosophila* Set1, MLL1 and MLL2 to Trx, and MLL3 and MLL4 to Trr (Fig. 1.2A) (Shilatifard, 2012). Set1 mutations are not lethal in budding yeast, suggesting that H3K4 methylation is not essential, but are associated with a slow growth phenotype (Miller et al., 2001). By contrast, all three H3K4 HMTs are essential in *Drosophila* (Hallson et al., 2012; Mazo et al., 1990; Sedkov et al., 1999). *Setd1A, Setd1B, Mll1* and *Mll2* knockout mice all show embryonic lethality (Glaser et al., 2006; Yu et al., 1995), consistent with non-redundant functions of these proteins in development. *Mll4* knockout embryos die at E9.5, whilst *Mll3* knockout is perinatal lethal, suggesting a degree of redundancy between MLL3 and MLL4 (Lee et al., 2013).

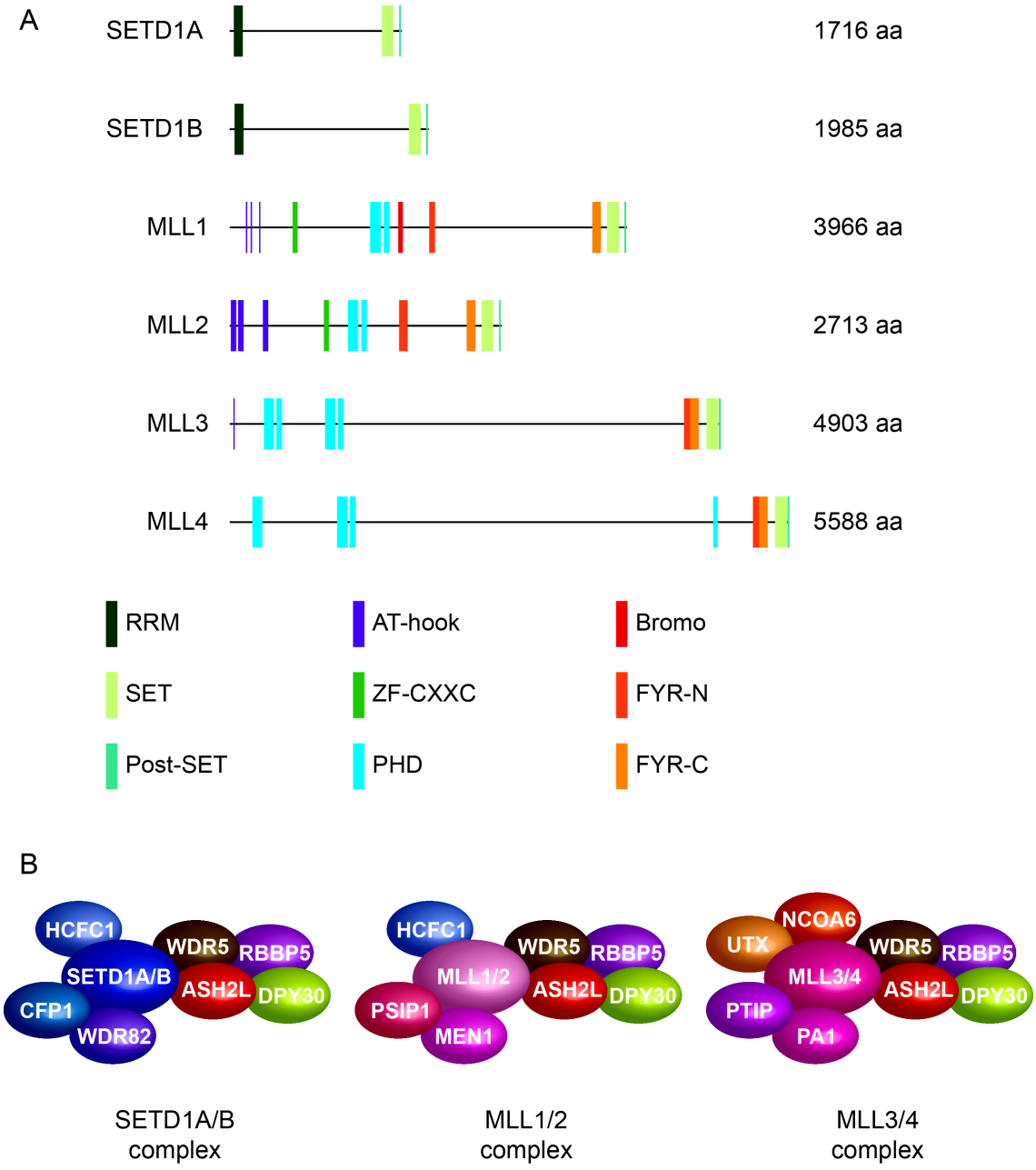



**Fig. 1.2: Mammalian H3K4 methyltransferase enzymes.**
**(A)** Domain schematics of SETD1A, SETD1B, MLL1, MLL2, MLL3 and MLL4. Abbreviations: RNA-recognition motif (RRM); catalytic SET domain (SET); ZF-CXXC DNA-binding domain (ZF-CXXC); plant homeodomain (PHD); bromodomain (bromo); phenylalanine/tyrosine-rich N- and C-terminal domains (FYR-N/FYR-C).
**(B)** Schematics representing the complexes formed by SETD1A/B (left). MLL1/2 (centre), and MLL3/4 (right).

All six mammalian Set1 homologues form part of complexes known as COMPASS (complex of proteins associated with Set1)-like complexes (Miller et al., 2001) (Fig. 1.2B). These complexes share four common subunits called WDR5, RBBP5, ASH2L and DPY30, which are required for proper catalytic activity (Zhang et al., 2015). The different COMPASS-like complexes are distinguished both by their catalytic subunits and by additional accessory proteins (van Nuland et al., 2013), and contribute to the H3K4 methylation landscape in different ways.

SETD1A and SETD1B are the complexes primarily responsible for the bulk of H3K4me3 in both mammals and *Drosophila* (Ardehali et al., 2011; Wu et al., 2008). H3K4me3 is found at both active and inactive gene promoters in mammals, with higher levels of H3K4me3 correlating with higher levels of transcription (Guenther et al., 2007). SETD1A/B complexes contain the additional subunits CFP1 (ZF-CXXC finger protein 1), WDR82, and HCFC1 (host cell factor 1) (Fig 1.2B). CFP1 contains both a PHD finger that binds H3K4me3, and a ZF-CXXC DNA binding domain, which recognizes unmethylated CpG promoters that overlap with unmethylated CpG island (CGI) elements (Eberl et al., 2013; Voo et al., 2000). WDR82 was found to interact with the initiating, serine-5 phosphorylated form of RNAPII (Lee and Skalnik, 2008). Genetic analysis shows that SETD1A/B complexes are required to generate high levels of H3K4me3 primarily at active promoters through recruitment by WDR82 and CFP1 (Brown et al., 2017; Clouaire et al., 2012, 2014; Lee and Skalnik, 2008).

MLL1/2 complexes also contain HCFC1, but are characterised by the presence of PSIP1 and Menin (also known as MEN1) (Fig. 1.2B) (van Nuland et al., 2013). Like CFP1, both MLL1 and MLL2 contain ZF-CXXC domains that bind to unmethylated DNA (Ma et al., 1993). Nevertheless, MLL1 is present only at a subset of active promoters in human lymphoma cells (Milne et al., 2005). MLL2 is bound at promoters genome-wide in embryonic stem cells (ES cells) (Denissov et al., 2014). However, MLL2 is primarily required for placement of H3K4me3 at a subset of promoters that are also associated with the Polycomb group (PcG) of transcriptional repressors and marked by H3K27me3 (Denissov et al., 2014; Hu et al., 2013a), although the mechanisms by which MLL2 activity

is restricted to such “bivalent” promoters remains unclear. It is thought that the H3K4me3 present at these repressed gene promoters is important for activation of developmental genes during differentiation, consistent with the lethal phenotype associated with loss of *Mll2* (Glaser et al., 2006). However, this “poising” of genes for activation remains controversial.

The MLL3/4 complexes are primarily involved in placement of H3K4me1, rather than H3K4me3 (Dorighi et al., 2017; Hu et al., 2013b). H4K4me1 has been mapped to enhancer regions in the mammalian genome, and is thought to mark both active and inactive enhancers (Heintzman et al., 2007, 2009). However, the predictive value of H3K4me1 in identifying putative enhancers is open to question, given that many identified enhancers have not been tested functionally, and that H3K4me1 is present on approximately 30% of all H3 molecules in ES cells (compared to 1.4% with H3K4me2 and 0.35% with H3K4me3) (Dorighi et al., 2017). Consistent with the possibility that generation of H3K4me1 may not be their primary function, MLL3/4 catalytic activity is largely dispensable for regulation of transcription, and MLL3/4 binding to enhancers is thought to play a direct role in binding of RNAPII and activation of transcription (Dorighi et al., 2017). MLL3/4 interact with four characteristic subunits in their COMPASS-like complex, namely UTX, NCOA6, PTIP, and PA1 (Fig. 1.2B) (van Nuland et al., 2013). NCOA6, PTIP, and PA1 are thought to have roles in recruitment of MLL3/4 complexes through interaction with transcription factors (Shilatifard, 2012). UTX (ubiquitously transcribed tetratricopeptide repeat, X chromosome) is a histone demethylase enzyme that removes the repressive H3K27me3 mark and is required for proper expression of developmental genes during *Drosophila* development (Copur and Müller, 2013). However, UTX is also thought to have roles independent of its catalytic activity, including bridging interactions between the MLL3/4 complex and the HAT p300 to enhance gene activation (Wang et al., 2017a). This is consistent with observations that at least some UTX functions can be compensated by its Y chromosome-encoded homologue UTY (Gozdecka et al., 2018), which is thought to be catalytically inactive (Hong et al., 2007; Lan et al., 2007; Shpargel et al., 2012).

### *1.5.2 Functions of H3K4 methylation*

H3K4 methylation is thought to function primarily through influencing the binding of effector proteins (Eberl et al., 2013). Amongst many other binders of H3K4me3, this modification is bound by PHD domains in the BPTF subunit of the NURF nucleosome remodelling complex, and by the TAF3 subunit of the general transcription factor TFIID (Vermeulen et al., 2007; Wysocka et al., 2006). H3K4me3 is therefore thought to promote transcriptional activation by increasing the accessibility of associated regulatory elements and by direct recruitment of the transcriptional machinery. H3K4me3 is also thought to function by blocking binding of repressive factors to promoters. H3K4me3 inhibits binding of the DNMT3L protein, which interacts with *de novo* DNA methyltransferases DNMT3A and DNMT3B, with nucleosomes (Ooi et al., 2007). H3K4me3 is also refractory to binding of the repressive NuRD complex, and inhibits the activity of Polycomb repressive complex 2 (PRC2) (Musselman et al., 2009; Schmitges et al., 2011).

The function of H3K4me1 is less clear. Screens for proteins that bind specifically to H3K4me1 have identified proteins, such as the BAF chromatin remodeller subunit BRG1, which preferentially bind H3K4 monomethylated mononucleosomes compared to nucleosomes carrying H3K4me3 (Local et al., 2018). However, such proteins typically bind to unmodified nucleosomes with similar affinity as to nucleosomes with H3K4me1, although the presence of H3K4me1 leads to slightly increased BAF-mediated chromatin remodelling *in vitro* (Local et al., 2018). Together, these results suggest that H3K4me3 is likely to play a role in maintaining active transcription at promoters. However, the extent to which H3K4me1 promotes active transcription at enhancers remains incompletely understood, and several lines of evidence point to roles for MLL3/4 complexes that are independent of their capacity to generate H3K4me1. Such functions may be mediated by MLL3/4 interaction partners that are not part of the canonical MLL3/4 COMPASS-like complex, and identification of such proteins would benefit from unbiased approaches to identifying the MLL3/4-associated proteome.

## 1.6 Chromatin modification by the Polycomb repressive system

### *1.6.1 Overview of the Polycomb system*

Polycomb group (PcG) proteins were first identified in *Drosophila* as developmental regulators whose absence leads to lethality and homeotic transformations (Lewis, 1978; Struhl, 1981). Segment identity in *Drosophila* is determined by expression of homeotic genes, such as the *Hox* genes, which are expressed in spatially restricted patterns in early development (Akam, 1987). In embryos mutant for a PcG protein, the initial pattern of expression of *Hox* genes, such as *Ultrabithorax* (*Ubx*), are unaltered compared to wild type animals, but show widespread misexpression at later developmental time points (Struhl and Akam, 1985). PcG proteins are therefore thought to act as negative regulators of gene expression by maintaining the repressed state of genes in cells where these genes were not expressed early in development.

PcG proteins assemble into several distinct complexes (Simon and Kingston, 2009). The best-studied PcG complexes are Polycomb repressive complexes 1 and 2 (PRC1 and PRC2), but additional complexes include the PRC1-like complex dRING-associated factors (dRAF), Polycomb repressive deubiquitylase (PR-DUB), and Pho repressive complex (PhoRC). The role of PhoRC will be briefly discussed, and PRC1, PRC2 and related complexes will be discussed in more detail in the sections that follow.

PhoRC is a *Drosophila* Polycomb complex comprising the sequence-specific DNA binding protein Pho (Pleiohomeotic) and its interaction partner Sfmbt (Scm-relate gene containing four MBT repeats) (Alfieri et al., 2013; Klymenko et al., 2006). The presence of a DNA binding protein in PhoRC suggested the possibility that this complex is responsible for recruitment of PcG proteins to their target sites in the fly genome, called Polycomb response elements (PREs). Consistent with this, Pho binding was found to be important for recruitment of other PcG proteins to *Hox* gene targets in *Drosophila* (Mohd-Sarip et al., 2002; Wang et al., 2004b). Mechanistically, PhoRC-mediated recruitment of PcG proteins is thought to occur through interaction of Sfmbt with the sub-stoichiometric PRC1 component Scm (Sex comb on midleg) via their sterile

alpha motif (SAM) domains (Frey et al., 2016; Grimm et al., 2009). However, mutation of *Pho* and its paralogue *Pho-like* (*Phol*) leads to lethality later in development than mutation of *Scm* and other PRC1 proteins, and is associated with only minor misregulation of target genes (Breen and Duncan, 1986; Brown, 2003), suggesting that PhoRC-mediated mechanisms cannot account for all PRC1 recruitment. This is consistent with observations that PRC1 stabilises recruitment of Pho at target sites rather than the reverse (Kahn et al., 2014; Schuettengruber et al., 2014), and that deletion of Sfmbt does not lead to widespread loss of PRC1 recruitment at PREs (T. Sheahan, J. Muller, unpublished observations). This suggests that whilst PhoRC might play a role in recruitment of other PcG proteins to a subset of classical Polycomb targets, such as *Hox* genes (Frey et al., 2016; Fritsch et al., 1999), PhoRC does not represent a general recruitment mechanism for PcG proteins in *Drosophila*. This is also consistent with observations that binding of the mammalian homologue of Pho, YY1, does not overlap with PcG targets genome-wide, and is therefore not thought to play a major role in Polycomb recruitment (Mendenhall et al., 2010).

### *1.6.2 PRC1, PRC2 and their histone modifications*

#### *1.6.2.1 PRC1 and H2AK119ub1*

PRC1 in flies consists of four core components, namely Pc (Polycomb), Ph (Polyhomeotic), Psc (Posterior sex combs) and Sce (Sex combs extra), which is also known as dRING (Shao et al., 1999). The PRC1-like dRAF complex contains the dRING and Psc subunits of PRC1, but lacks Pc and Ph, possessing instead additional subunits such as the H3K36 demethylase dKDM2 (Lagarou et al., 2008). The PRC1 core complex is largely conserved between flies and mammals, suggesting that mammalian PRC1 may function using similar mechanisms as *Drosophila* PRC1 (Levine et al., 2002). However, in mammals each PRC1 subunit has multiple paralogues, increasing the combinatorial complexity of the system (Fig. 1.3). There are two mammalian homologues of dRING (called RING1A and RING1B in mice), five homologues of Pc (CBX2, 4, 6, 7 and 8), three Ph proteins (PHC1, 2 and 3), and six homologues of Psc (PCGF1, PCGF2 (also known as MEL18), PCGF3, PCGF4

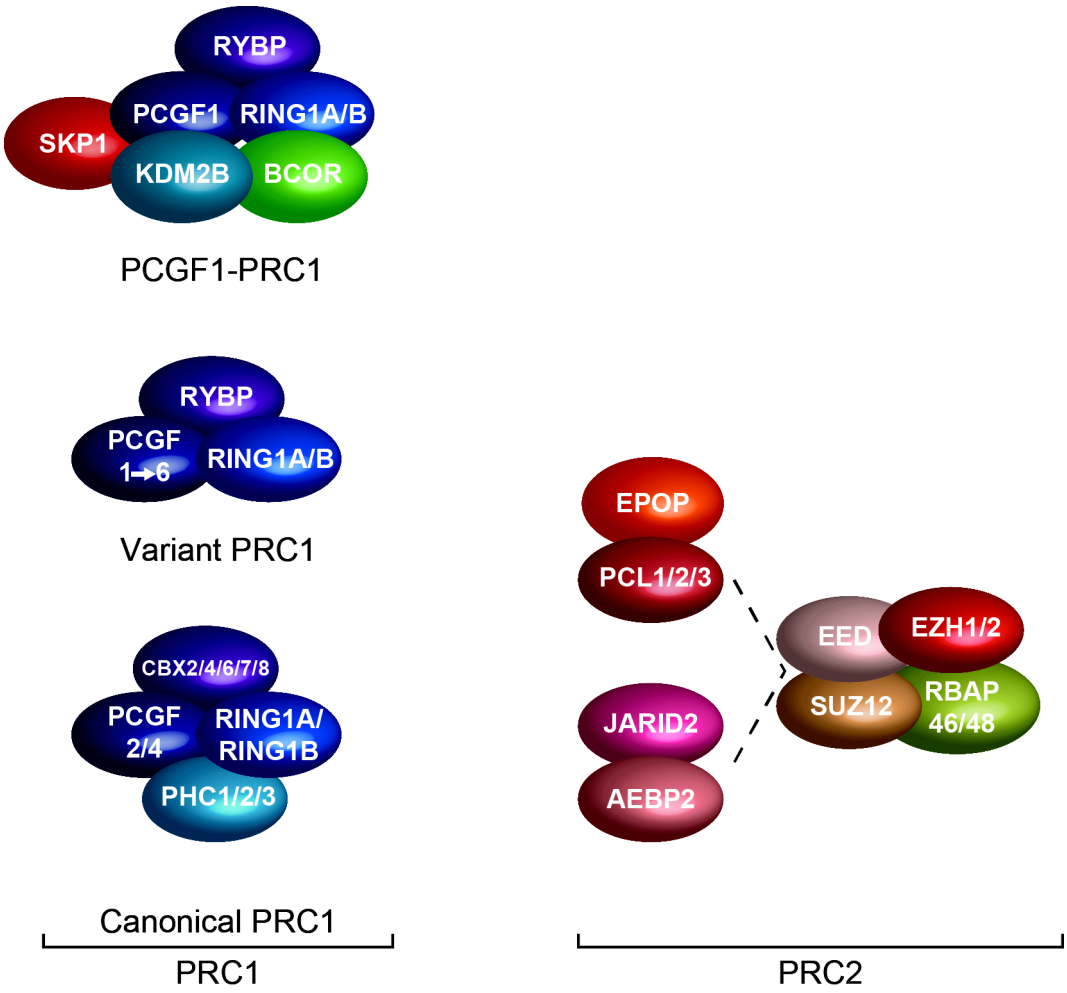



**Fig. 1.3: Polycomb group protein complexes in mammals.**
Left: Canonical and variant PRC1-type complexes, including the variant PCGF1-PRC1 complex that contains the ZF-CXXC protein KDM2B.
Right: PRC2 complexes.

(also known as BMI1), PCGF5 and PCGF6). Two PCGF proteins, PCGF2 and PCGF4, form complexes similar to the classical *Drosophila* PRC1 complex, containing CBX and PHC proteins, which are referred to as canonical PRC1. PCGF1, 3, 5 and 6 proteins form variant PRC1 complexes. These variant complexes lack CBX and PHC proteins, but contain additional common subunits, such as RYBP, and subunits that are unique to each complex (Farcas et al., 2012; Gao et al., 2012).

PRC1 is an E3 ubiquitin ligase, which monoubiquitylates H2AK119 in mammals and H2AK118 in *Drosophila* (H2AK119ub1/H2AK118ub1) (Cao et al., 2005; Wang et al., 2004a). The catalytic subunit of PRC1 is the RING1A/B subunit, but activity depends on interaction between the two RING fingers of the

RING1A/B subunit and the PCGF subunit (Buchwald et al., 2006; Cao et al., 2005). Importantly, the identity of the PCGF subunit in a PRC1 complex influences its catalytic activity, so that canonical complexes containing PCGF2 and PCGF4 have intrinsically lower catalytic activity than variant PRC1 complexes *in vitro* (Taherbhoy et al., 2015). Moreover, inclusion of the variant PRC1-specific subunit RYBP further stimulates the activity of RING1B-PCGF dimers (Rose et al., 2016). The greater catalytic activity of variant PRC1 has alsobeen confirmed *in vivo*, suggesting that canonical and variant PRC1 complexes may have distinct functions in cells (Blackledge et al., 2014, 2019).

PRC1 is thought to repress gene expression both by directly altering chromatin structure and by ubiquitylating H2A, although the extent to which each of these mechanisms contributes to PRC1 function remains incompletely understood. PRC1 was proposed to compact chromatin via a mechanism involving the long, unstructured C-terminal region of *Drosophila* Psc, a region which is not present in mammalian PCGF proteins (Francis et al., 2004; King et al., 2002). PRC1 is also thought to contribute to gene repression by engaging in long-range interactions between promoters through its PHC subunits, contributing to the maintenance of a repressive chromatin environment (Isono et al., 2013; Schoenfelder et al., 2015). Work to understand the role of PRC1 catalytic activity in PRC1 function showed that removal of RING1B in ES cells led to decompaction of repressed target loci, but that this decompaction phenotype was not observed in cells expressing a RING1B$^{I53A}$ mutant that should render the protein catalytically inactive (Buchwald et al., 2006; Eskeland et al., 2010). This suggestion that repression of target genes by PRC1 might be independent of its catalytic activity was further supported by work showing that mice expressing RING1B$^{I53A}$ mutations are viable and that mutation of H2AK118 to arginine in *Drosophila* larvae does not lead to misexpression of Polycomb target genes (Illingworth et al., 2015; Pengelly et al., 2015).

However, several lines of evidence point to a central role for H2AK119ub1 in Polycomb-mediated gene repression. Mutation of H2AK118 in flies is embryonic lethal (Pengelly et al., 2015). Mutation of the RING1B catalytic domain in ES cells in a RING1A-negative background leads to gene expression changes that disrupt the maintenance of cell identity (Endoh et al., 2012), suggesting that

previous work may have been complicated by compensation by RING1A. Moreover, the extensively used RING1B$^{I53A}$ mutation does not fully ablate PRC1 catalytic activity (Buchwald et al., 2006; Scheuermann et al., 2012), and recent work has shown that mutations that completely abolish RING1A/B catalytic activity are lethal in constitutively mutant ES cells (Blackledge et al., 2019). Moreover, in an inducible system, such catalytic mutations can recapitulate the gene misexpression phenotype of RING1A/B knockout cells (Blackledge et al., 2019). Mechanistically, H2AK119ub1 is thought to function primarily by directing the recruitment of PRC2, with genome-wide PRC2 binding greatly reduced in the absence of PRC1 catalytic activity (Blackledge et al., 2014, 2019; Cooper et al., 2016; Kalb et al., 2014).

### *1.6.2.2 PRC2 and H3K27me3*

The mammalian PRC2 complex comprises a core of four subunits, namely EZH2 (or its paralogue EZH1), EED, SUZ12 and RBAP46 or RBAP48 (also known as RBBP7 and RBBP4, respectively) (Fig. 1.3) (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002). PRC2 is an HMT enzyme that mono-, di- and trimethylates H3K27 (H3K27me1/2/3). The catalytic activity of PRC2 is contained in the SET domain of the EZH2 subunit, but interaction with the SUZ12 and EED subunits is required for efficient H3K27 methylation activity (Cao and Zhang, 2004; Pasini et al., 2004). The PRC2 core complex interacts with additional sub-stoichiometric components to form two sub-complexes, which may differ in their recruitment or enzymatic activities. One sub-complex comprises the PRC2 core components together with the additional proteins JARID2 and AEBP2, and a second sub-complex is formed with one of three proteins homologous to *Drosophila* Pcl (called PHF1, MTF2 and PHF19), EPOP (previously known as C17ORF96) and C10ORF12 (Alekseyenko et al., 2014; Conway et al., 2018; Grijzenhout et al., 2016; Liefke and Shi, 2015; Nekrasov et al., 2007).

PRC2 is thought to function in gene repression primarily through its catalytic activity. Mutation of the catalytic domain in the *Drosophila* EZH2 homologue E(z) (Enhancer of zeste) is sufficient for misexpression of Polycomb target

genes, and mutation of the H3K27 residue to arginine in *Drosophila* recapitulates a PRC2 mutant phenotype (Müller et al., 2002; Pengelly et al., 2013). However, recent work has also implicated PRC2 in mediating long-range chromatin interactions, facilitating future activation of repressed genes (Cruz-Molina et al., 2017), suggesting that PRC2 may have functions in addition to methylation of H3K27 and repression of gene expression.

Our understanding of how mechanistically H3K27me3 mediates gene repression remains incomplete. The CBX subunits in canonical PRC1 complexes bind H3K27me3 via their chromodomain (Czermin et al., 2002). This is sufficient for recruitment of PRC1 to sites marked by H3K27me3 (Blackledge et al., 2014), and may lead to transcriptional repression by compacting chromatin. In addition, H3K27me3 inhibits the activity of H3K4 methyltransferases *in vitro* (Kim et al., 2013), and may therefore reduce the capacity of these enzymes to generate a more active chromatin environment *in vivo*. This inhibition of H3K4 methylation may also lead to the generation of "bivalent" nucleosomes, in which one H3 molecule is marked by H3K27me3 and the other by H3K4me3 (Bernstein et al., 2006; Voigt et al., 2012). Such a chromatin structure may inhibit gene expression by preventing higher affinity binding of proteins such as TAF3 to nucleosomes symmetrically modified with H3K4me3.

#### *1.6.2.2 Recruitment of Polycomb proteins*

As discussed above, PcG proteins in flies are recruited to PREs. Although some PREs associated with classical Polycomb target genes, such as *Hox* genes, are located at a distance from the genes that they regulate, the majority of PREs in *Drosophila* and Polycomb target sites in mammals are associated with the transcription start sites (TSSs) of genes (Ku et al., 2008; Oktaba et al., 2008; Schwartz et al., 2006).

In mammals, several mechanisms have been proposed for PcG protein recruitment to target sites. As in flies, sequence-specific DNA binding proteins, such as REST, were identified as interaction partners of PcG proteins (Ren and

Kerppola, 2011). However, given that such transcription factors are not found to co-occupy most Polycomb target sites, this is unlikely to represent a general mechanism for Polycomb recruitment. Long noncoding RNAs (lncRNAs) have also been proposed to interact with PcG proteins and recruit them to targets such as the inactive X chromosome (Xi). However, it is unclear how mechanistically such recruitment might occur, as super-resolution microscopy studies suggest that PcG proteins do not directly interact with the *Xist* lncRNA which associates with Xi (Cerase et al., 2014). Moreover, PRC2 interacts with RNAs in a sequence non-specific manner, with such interactions thought to counteract rather than direct PcG recruitment genome-wide (Davidovich et al., 2013).

Genome-wide analysis of binding of PcG proteins and H3K27me3 in mouse and human ES cells showed that Polycomb binding occupies the promoters of approximately 2000 genes, overlapping almost exclusively with a subset of CpG islands (CGIs) (Bernstein et al., 2006; Boyer et al., 2006; Ku et al., 2008; Mikkelsen et al., 2007). Indeed, further work showed that introduction of CGI-like DNA from *E. coli* or randomly generated CGI-like DNA was sufficient to mediate recruitment of Polycomb (Mendenhall et al., 2010; Wachter et al., 2014). Furthermore, analysis of Polycomb binding sites in *Drosophila* showed that genes marked by H3K27me3 that have Polycomb target homologues in mammals have GC-rich promoter sequences, in marked contrast to typical *Drosophila* promoters which are AT-rich (Sharif et al., 2013). These results suggested that CGIs might play a direct, and potentially evolutionarily conserved, role in Polycomb recruitment.

For many years, PcG proteins were thought to be recruited by a “hierarchical” mechanism in which PRC2 was first recruited by proteins such as Pho, generating H3K27me3 at target sites. This mark could then be bound by the chromodomain of CBX proteins, leading to recruitment of PRC1 (Wang et al., 2004b). However, more recent work has shown that PRC1 can also be recruited to target sites in the absence of PRC2 (Tavares et al., 2012), suggesting that although H3K27me3 may play an important role in stabilising PRC1 binding and in gene repression, it is not essential for nucleation of PRC1 at target genes.

A direct mechanistic link between CGIs and PRC1 recruitment emerged in the form of the protein KDM2B. KDM2B, a mammalian homologue of the *Drosophila* dRAF complex subunit dKDM2, is an H3K36 demethylase enzyme that, like the CFP1 and MLL1/2 proteins described above, contains a ZF-CXXC DNA binding domain (Farcas et al., 2012; He et al., 2013; Wu et al., 2013). The ZF-CXXC domain of KDM2B facilitates its recruitment to CGIs genome-wide, but KDM2B is specifically enriched at Polycomb target sites (Farcas et al., 2012). Importantly, purification of KDM2B from ES cells demonstrated that this protein forms part of a variant PRC1 complex characterised by the presence of PCGF1. Knockdown of KDM2B or deletion of its ZF-CXXC domain led to loss of PRC1 binding and H2AK119ub1 genome-wide (Blackledge et al., 2014; Farcas et al., 2012; He et al., 2013; Wu et al., 2013), and reduction in PRC2 and H3K27me3 at approximately two-thirds of binding sites (Blackledge et al., 2014). Moreover, artificial tethering of a KDM2B protein was sufficient for the formation of a Polycomb domain, characterised by the presence of PRC1, PRC2, H2AK119ub1 and H3K27me3 (Blackledge et al., 2014). This showed that binding of a PRC1-associated protein to CGIs was sufficient for the recruitment of PRC1 and PRC2 in ES cells.

A further mechanistic link between Polycomb and CGIs has recently been uncovered in observations that the PRC2-associated proteins PHF1, MTF2 and PHF19 can bind specifically to unmethylated CpG dinucleotides via a winged helix domain (Li et al., 2017). Work *in vivo* further suggested that loss of this DNA binding activity leads to a partial loss of PRC2 binding at target sites, although there is little effect on H3K27me3. This suggests that the DNA binding activity of Pcl proteins cannot fully account for PRC2 recruitment genome-wide. Moreover, an additional report investigating the DNA binding activity of Pcl proteins found an alternative mechanism for PHF1 binding to DNA, indicating that PFH1 binds in a sequence non-specific manner, and does not specifically interact with CpG dinucleotides (Choi et al., 2017). Therefore, the contribution of Pcl homologues to PRC2 recruitment remains controversial, although these proteins are likely have a role in stabilising PRC2 on chromatin (Choi et al., 2017).

In addition to finding that KDM2B can mediate recruitment at Polycomb-repressed CGIs, careful analysis of the genome-wide ChIP data showed that the PRC1 catalytic subunit RING1B is recruited at low levels to essentially all CGIs in a KDM2B-dependent manner (Farcas et al., 2012). This surprising finding led to the hypothesis that all CGIs are dynamically sampled by KDM2B-mediated PRC1 binding, but that PRC1 is only stably recruited at repressed loci. This would provide a mechanism by which PcG proteins could perform their function of maintaining repression of previously repressed genes. Consistent with this, inhibition of transcription is sufficient for recruitment of PRC2 and generation of H3K27me3 at previously active genes (Riising et al., 2014). This suggests that PcG proteins bind both active and repressed CGIs, but that there are mechanisms in place at active CGIs to destabilise Polycomb binding. For example, the transcription-associated histone modification H3K4me3 inhibits the activity of PRC2 (Schmitges et al., 2011), the BAF chromatin remodelling complex is thought to evict PRC1 from chromatin (Stanton et al., 2017), RNA binding is thought to displace PRC2 from chromatin (Davidovich et al., 2013), and transcription through a PRE in flies is sufficient for loss of Polycomb binding (Erokhin et al., 2015; Schmitt et al., 2005). Through one or more such mechanism, CGIs can act as bistable switches, facilitating conversion between stably active and repressed states (Klose et al., 2013).

## 1.7 CpG islands and DNA methylation

### *1.7.1 DNA methylation and repression of transcription*

DNA, like histone proteins, undergoes modification in eukaryotic cells. The predominant modification of DNA is methylation, which in multicellular eukaryotes takes place at the 5-position of the cytosine ring to generate 5-methyl-cytosine (5mC). In vertebrates, 5mC is found almost exclusively in the context of CpG dinucleotides, with some 70-80% of CpG sites found methylated across the genome in mammalian cells (Lister et al., 2009).

DNA methylation is generated by DNA methyltransferase (DNMT) enzymes (Klose and Bird, 2006). The first DNMT enzyme to be identified was DNMT1

(Bestor and Ingram, 1983). This enzyme is significantly more active *in vitro* towards hemimethylated DNA than towards unmethylated DNA (Bestor and Ingram, 1983), and therefore plays a role as a maintenance methyltransferase, facilitating the propagation of symmetrical DNA methylation following DNA replication (Jeltsch, 2006). The importance of maintenance of DNA methylation is demonstrated by observations that knocking out *Dnmt1* leads to loss of approximately two-thirds of DNA methylation in ES cells, and to embryonic lethality (Li et al., 1992).

The retention of significant levels of DNA methylation in *Dnmt1* null cells suggested the presence of additional DNMT enzymes. Indeed, homology searches identified two proteins, named DNMT3A and DNMT3B, that can methylate CpG in both hemimethylated and unmethylated contexts (Okano et al., 1998). *Dnmt3a; Dnmt3b* double knockout ES cells are unable to methylate a newly integrated retroviral DNA sequence, indicating that these enzymes are *de novo* DNMTs, and double knockout embryos die before E11.5 (Okano et al., 1999).

DNA methylation is thought to be repressive to transcription (Klose and Bird, 2006), although importantly a strong transactivation signal is sufficient to overcome the repressive effects of DNA methylation (Thompson et al., 1986, 1988) . One mechanism through which DNA methylation inhibits transcription is by preventing the binding of sequence-specific transcription factors. DNA methylation can interfere with binding of transcription factors whose cognate sequence contains a CpG dinucleotide. In this way, DNA methylation is refractory to binding of CREB and CTCF, and methylation of a CTCF binding site can interfere with CTCF-mediated gene looping and gene expression (Bell and Felsenfeld, 2000; Liu et al., 2016; Mancini et al., 1999).

A second mechanism through which DNA methylation influences transcription is through binding by proteins that possess a methyl-CpG DNA-binding domain (MBD) (Hendrich and Bird, 1998; Klose and Bird, 2006). Several MBD proteins associate with transcriptional co-repressor complexes. The MBD protein MeCP2 associates with the NCoR/SMRT co-repressor complex, which contains HDAC3 (Lyst et al., 2013; Nan et al., 1998). The importance of this interaction is

underlined by observations that mutations that abolish either methyl-CpG binding or interaction with NCoR/SMRT are associated with the neurological disorder Rett syndrome (RTT) (Lyst et al., 2013). Moreover, treatment of cells with HDAC inhibitors led to derepression of a reporter gene, indicating that recruitment of HDAC activity is likely to play an important role in MeCP2-mediated transcription silencing. Similarly, the MBD2 and MBD3 proteins are thought to play a role in gene repression by their presence in the in the repressive nucleosome remodelling and HDAC complex NuRD complex (Zhang et al., 1999).

### *1.7.2 CpG islands have a transcriptionally permissive chromatin architecture*

Although DNA methylation is prevalent across the mammalian genome, there are short, contiguous regions that are free of this methylation, called CpG islands (CGIs) (reviewed in Blackledge and Klose, 2011). CGIs were originally identified as regions that were sensitive to digestion by the restriction enzyme *HpaII*, which is able to cut unmethylated but not methylated DNA, and account for approximately 1% of the vertebrate genome (Cooper et al., 1983). CGIs were found to be associated with gene sequences, and the lack of DNA methylation was thought to render these regions “available” for identification by proteins that could influence transcription (Bird, 1986).

CGIs have an average length of approximately 1 kb, and are characterised as being both GC-rich and CpG-rich compared to the bulk genome (Illingworth and Bird, 2009). CGIs are defined computationally as having a GC base composition greater than 50% and a CpG observed/expected ratio of greater than 0.6 (Gardiner-Garden and Frommer, 1987). CGIs are frequently associated with gene promoters in the mammalian genome, with some 70% of promoters overlapping CGIs (Saxonov et al., 2006). Indeed, even CGIs that are not associated with annotated TSSs exhibit promoter-like properties, including transcriptional initiation (Illingworth et al., 2010). It is therefore thought that CGIs contribute to promoter function, although the mechanisms by which they do so are not yet fully understood.

One mechanism by which CGIs are thought to influence transcriptional regulation is through the generation of a unique chromatin architecture that renders CGI-associated promoters permissive to transcription (Blackledge and Klose, 2011). Consistent with this, CGI chromatin released from the genome by digestion with restriction enzymes is comparatively nucleosome-free, enriched with histone acetylation and H3K4 methylation, and depleted of H3K9me3 and histone H1 (Tazi and Bird, 1990; Thomson et al., 2010). Furthermore, mapping of nucleosome occupancy *in vivo* by digestion with micrococcal nuclease (MNase) showed that CGI promoters are depleted of nucleosomes in a manner independent of transcription (Fenouil et al., 2012). Consistent with the possibility that CGIs promote a transcriptionally permissive chromatin environment, induction of gene expression by treatment of macrophages with lipopolysaccharide (LPS) reveals that the majority of rapidly induced primary response genes are associated with CGIs, whilst the secondary response genes are more likely to be non-CGI genes (Ramirez-Carrozzi et al., 2009). Moreover, induction of CGI genes was more likely to occur independently of chromatin remodelling activities, suggesting that CGIs are transcriptionally permissive without mechanisms employed by non-CGI genes to alter chromatin architecture.

### *1.7.3 ZF-CXXC DNA binding domain influences CGI chromatin architecture*

One mechanism by which the permissive chromatin environment at CGIs is thought to be established is through binding of proteins containing a domain that specifically recognises unmethylated CpG dinucleotides, called the ZF-CXXC DNA binding domain (Fig. 1.4A) (Long et al., 2013). ZF-CXXC proteins include enzymes that modify chromatin directly and subunits of larger chromatin-modifying complexes, and therefore contribute to the chromatin environment at CGIs.

The ZF-CXXC domain was originally identified in proteins including MLL1, DNMT1 and MBD1 (Bestor and Verdine, 1994; Cross et al., 1997; Ma et al., 1993). However, the selective binding activity for unmethylated CpG was not

A


B

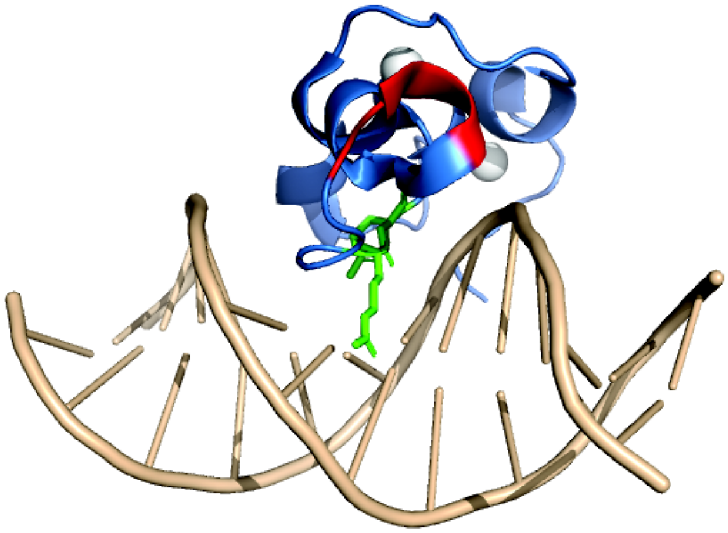


**Fig. 1.4: The ZF-CXXC selectively binds unmethylated DNA.**
**(A)** Alignment of ZF-CXXC domains from multiple proteins in mice. Greyscale shading represents the level of conservation, the KFGG motif is highlighted in red, and the KQ/RQ motif that interrogates the methylation status of the CpG dinucleotide in green.
**(B)** Cartoon representation of the ZF-CXXC domain of CFP1 bound to a DNA probe containing unmethylated CpG (PDB ID: 3QMG). The DNA is shown in wheat, and the protein domain in blue. The zinc ions are represented as grey spheres, the KFGG motif is shaded red, and the RQ motif residues are shown in stick representation inserting into the major groove of DNA.

definitively shown until the identification and cloning of the ZF-CXXC protein CFP1 (CXXC finger protein 1) (Lee et al., 2001; Voo et al., 2000). Structural analysis of the ZF-CXXC domains of MLL1, DNMT1 and CFP1 shows that the domain forms a compact crescent-like structure containing eight conserved cysteine residues that coordinate two zinc ions (Allen et al., 2006; Cierpicki et al., 2010; Xu et al., 2011). The linker region between the two cysteine-rich clusters contains a KFGG motif that is required to maintain the rigidity of the structure, and a KR/KQ motif in a DNA-binding loop that mediates base contacts in the major groove of DNA (Fig. 1.4B). The presence of methylation on a cytosine base contacted by the KR/KQ motif results in steric clashes and prevents the formation of hydrogen bonding, explaining the selectivity of ZF-CXXC for unmethylated CpG. Importantly, in addition to the contacts in the major groove, regions flanking the ZF-CXXC domain contact the minor groove on the opposite face of DNA, providing additional interaction energy. For this reason, ZF-CXXC domains are unable to bind DNA that is occluded by nucleosomes and can only bind to linker DNA (Zhou et al., 2012), as the presence of histone octamer on DNA is thought to be refractory to binding in both the major and minor grooves simultaneously.

ZF-CXXC proteins contribute to the chromatin environment at CGIs by recruiting chromatin-modifying activities. As discussed above, ZF-CXXC domains are found in CFP1, MLL1 and MLL2, which are associated with H3K4 methyltransferase activity and lead to H3K4me3 at CGIs (Brown et al., 2017; Clouaire et al., 2012; Denissov et al., 2014; Thomson et al., 2010). FBXL19 interacts with the Mediator complex and recruits the complex to a subset of gene promoters *in vivo* (Dimitrova et al., 2018). KDM2A and KDM2B both possess Jumonji C (JmjC) lysine demethylase domains and demethylate H3K36 at CGI-associated promoters (Blackledge et al., 2010). KDM2B, as described previously, interacts with PRC1 and, in the absence of transcriptional activation, is sufficient for nucleation of Polycomb domains (Blackledge et al., 2014; Farcas et al., 2012; He et al., 2013; Wu et al., 2013). In this way, CGIs are thought to act as platforms for the recruitment of both activating and repressive chromatin-modifying activities, enabling CGI-associated genes to be stabilised in either active or repressed states (Klose et al., 2013).

However, several important questions remain about how CGIs function. Little is known about how the unmethylated state of CGIs is established and maintained. The presence of H3K4me3, which inhibits chromatin binding by DNMT3L, is thought to prevent *de novo* DNA methylation once H3K4me3 has been established (Ooi et al., 2007). However, how this state is initially established remains poorly understood. Several proteins that may play a role in demethylation of DNA, namely TET1, TET3 and the TET2 interaction partner IDAX, possess ZF-CXXC domains. The TET (ten-eleven translocation) proteins are dioxygenase enzymes that convert 5mC to 5-hydroxymethyl-cytosine (5hmC) (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009), and subsequently convert 5hmC to 5-formyl-cytosine (5fC) and 5-carboxyl-cytosine (5caC) (Ito et al., 2011). The presence of 5hmC is thought to lead to the demethylation of DNA either through passive loss of 5mC maintenance or actively through the base excision repair pathway (Williams et al., 2012). However, the ZF-CXXC domains of TET1, TET3 and IDAX differ from the canonical ZF-CXXC (Fig. 1.4B), and biochemical and structural studies suggest that the TET3 ZF-CXXC has more degenerate DNA binding properties (Xu et al., 2012). Therefore, how precisely TET proteins or other factos are involved in generating the unmethylated status of CGIs remains unclear.

A second important question about how CGIs function that remains to be answered is how mechanistically the chromatin environment at a CGI gives rise to a specific transcriptional output. Many chromatin-binding proteins that can recognise the chromatin modifications found at CGIs have been identified, but it is unclear whether additional unknown proteins remain to be found. In particular, our understanding of how Polycomb proteins are recruited to CGIs and evicted from actively transcribed CGIs remains incomplete, and a mechanistic basis for the activation and repression of transcription at CGIs awaits further elucidation. Answering these questions would provide valuable insights into how CGIs contribute to promoter function and therefore into how chromatin architecture contributes to the regulation of gene expression.

## 1.8 Aims of this thesis

Proper regulation of transcription is fundamental to multicellular life. One of the major mechanisms by which transcription is regulated is through alteration of chromatin architecture at regulatory elements such as gene promoters and enhancers (Fig. 1.5). The overall aim of this thesis is therefore to understand how chromatin-modifying activities contribute to regulation of gene expression.

Work in Chapter 3 aims to understand how the chromatin environment at promoters and enhancers is established. To this end, a strategy was developed taking advantage of the affinity of the ZF-CXXC domain for unmethylated DNA, to purify proteins associated with CpG island promoters. An alternative ChIP-mass spectrometry approach was also utilised, and identified DNA damage proteins and chromatin modifying proteins as potential regulators of promoter function. To understand how the chromatin environment at enhancers is established, the enhancer-binding MLL3/4 H3K4 methyltransferases were endogenously tagged and purified to identify associated proteins in an unbiased manner.

To further our understanding of how histone acetyltransferase enzymes contribute to the chromatin environment at regulatory elements, a candidiate approach was also adopted to test how the activity of the CBP/p300 family of HATs is regulated. Work in Chapter 4 establishes an *in vitro* biochemical strategy to address how the histone acetyltransferase enzyme CBP specifically acetylates histones at the key H3K27 residue in the context of chromatin substrates. These experiments identified a domain in the C-terminal portion of CBP called TAZ2 as crucial for H3K27 acetylation activity. Work in Chapter 5 then focussed on understanding how mechanistically the TAZ2 domain of CBP influences CBP substrate selection, revealing that TAZ2 binds DNA in a sequence-independent manner, potentially orienting CBP to favour acetylation of H3K27. Finally, work in Chapter 6 identified mutations in the TAZ2 domain that abrogate DNA binding and reduce acetylation of H3K27 *in vitro*, and prevent specific H3K27 acetylation by CBP *in vivo*.

Together, this work elucidates a novel mechanism by which CBP/p300 HAT activity is selective for H3K27 *in vivo*, and forms the basis for a discussion of how mechanisms that determine HAT substrate specificity are vital to ensure robust regulation of gene expression.

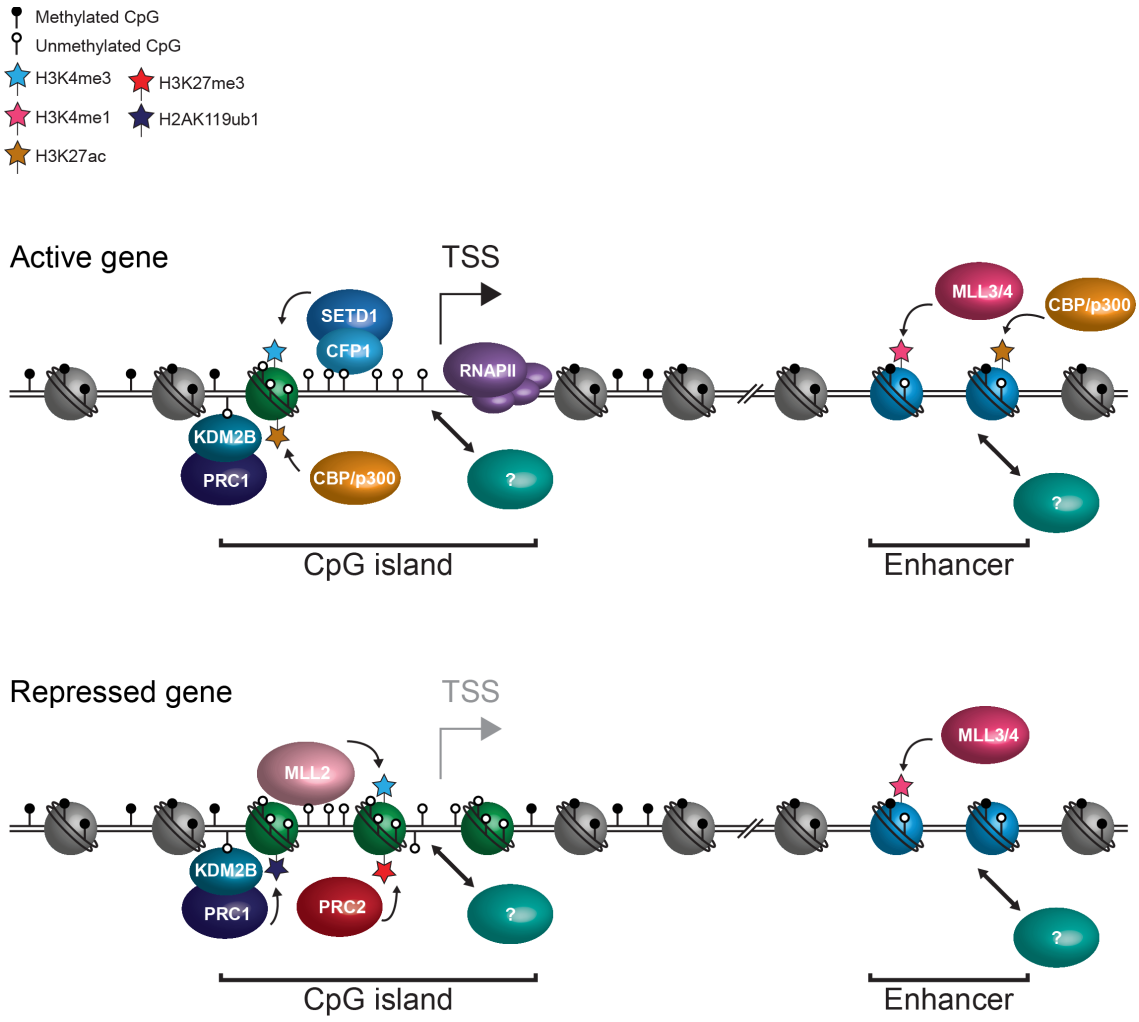



**Fig. 1.5: A model for chromatin-modifying activities in regulation of gene expression from promoters and enhancers.**
Top: At an active gene, the CpG island promoter is bound by the SETD1A/B complex, which places H3K4me3; by CBP/p300, which places H3K27ac; by PRC1, which is unable to establish a Polycomb domain; and by actively transcribing RNAPII. An active enhancer is bound by the MLL3/4 complex and CBP/p300, which place H3K4me1 and H3K27ac, respectively. Additional unknown proteins bind to promoters and enhancers, and mediate their function.
Bottom: At an inactive gene, the CpG island promoter is bound by the MLL2 complex, which places H3K4me3; and by PRC1 and PRC2, which place H2AK119ub1 and H3K27me3, respectively, generating a repressed Polycomb domain. An inactive enhancer is bound by the MLL3/4 complex, which places H3K4me1, but not by CBP/p300 HATs.

# 2. Materials and methods

## 2.1 DNA methods

### *2.1.1 DNA constructs used in this study*

**Table 2.1: Summary of DNA constructs used in this study**

| DNA inserts | | |
|---|---|---|
| Construct | Origin | Source |
| ZF-CXXC | Mouse KDM2B, codon optimised for *E. coli* expression | Synthesised by IDT with N-terminal 6xHis tag and C-terminal 3xFlag, 2xStrepII and 2xGCN4 tags |
| TEV protease | Tobacco etch virus | pRK793 (Addgene: 8827) |
| CBP | Mouse | A kind gift from Prof Shelley Berger (University of Pennsylvania) |
| CBP truncations and mutants | Mouse | Subcloned from full length CBP |
| p300 core | Human, residues 1048-1664 | pcDNA-dCas9-p300 Core (Addgene: 61357) |
| VP160 | Synthetic | pAC94-pmax-dCas9VP160-2A-puro (Addgene: 48226) |
| Vectors | | |
| Vector name | Experimental use | Source |
| pET22 | Bacterial expression | Voigt lab |
| pRK793 | Bacterial expression | Addgene: 8827 |
| pCAG | Mammalian protein expression | A kind gift from Dr Anca Farcaș (CRUK |

| | | |
|---|---|---|
| | | Cambridge Institute) |
| pFastBac | Insect cell expression | Voigt lab |
| pAC94 | dCas9 fusion protein expression | Addgene: 48226 |
| pmU6-gRNA | sgRNA expression | Addgene: 53187 |
| pX458 | Cas9 and sgRNA expression | Addgene: 48138 |
| CBP constructs | | |
| Construct | Residue numbers | Experimental use |
| CBP full length (FL) | 1-2441 | Insect cell expression |
| CBP core | 1082-1700 | Insect cell expression<br>Mammalian expression |
| CBP Nter-core | 1-1700 | Insect cell expression |
| CBP core-Cter | 1082-2441 | Insect cell expression |
| CBP core-ZZ (CZ) | 1082-1751 | Insect cell expression<br>Mammalian expression |
| CBP core-ZZ-TAZ2 ($CZT^{wt}$) | 1082-1873 | Insect cell expression<br>Mammalian expression |
| CBP core-ΔZZ-TAZ2 (CΔZT) | 1082-1873 (Δ1701-1758) | Insect cell expression |
| CBP core-ZZ-$TAZ2^{mut}$ ($CZT^{mut}$) | 1082-1873<br>R1769E/K1832E/K1850E | Insect cell expression<br>Mammalian expression |
| CBP core-ZZ-TAZ2 catalytically inactive ($CZT^{ci}$) | 1082-1873<br>D1436Y | Insect cell expression |
| ZZ | 1701-1751 | Bacterial expression |
| $TAZ2^{wt}$ | 1759-1873 | Bacterial expression |
| $TAZ2^{mut}$ | 1759-1873<br>R1769E/K1832E/K1850E | Bacterial expression |

Note that sequencing revealed that the original CBP full length sequence contains a P695L mutation, which was carried forward in the constructs containing the N-terminus. The mutation is in an unstructured region of the protein.

### *2.1.2 Polymerase chain reaction (PCR)*

#### *2.1.2.1 Analytical PCR*

Analytical PCRs were carried out in 10 µL final volume, for analysis by separation on an agarose gel, or in 50 µL final volume, for analysis by Sanger sequencing. DNA was prepared in QuickExtract DNA extraction solution (Epicentre) and PCRs were carried out with MangoTaq DNA polymerase (Bioline) in 1x MangoTaq buffer (Bioline), 1.5 mM $MgCl_2$ and 0.2 mM dNTP mix with 0.2 µM of forward and reverse primers. For problematic regions, 1x GC enhancer buffer (for Q5 polymerase, NEB) was added to reactions.

PCR reactions were mixed and briefly centrifuged. Reactions were then carried out in an Eppendorf Mastercycler Nexus thermal cycler, with cycling parameters varying depending on DNA template and primers. PCR products were then analysed by separation on an agarose gel and where necessary gel extracted using the EZNA Gel Extraction kit (Omega BioTek) and eluted in 30 µL volume for Sanger sequencing.

**Table 2.2: Pipetting scheme for analytical PCR.**

| Component | 10 µL final volume | 50 uL final volume |
|---|---|---|
| 5x MangoTaq buffer | 2 µL | 10 µL |
| Forward primer (5 µM) | 0.2 µL | 2 µL |
| Reverse primer (5 µM) | 0.2 µL | 2 µL |
| $MgCl_2$ (50 mM) | 0.3 µL | 1.5 µL |
| dNTP mix (10 mM) | 0.2 µL | 1 µL |
| DNA | 2 µL | 2 µL |
| MangoTaq (5 U/µL) | 0.2 µL | 1 µL |
| $H_2O$ | 4.5 µL | 30.5 µL |

**Table 2.3: Thermal cycling conditions for analytical PCR.**

| Cycling step | Temperature | Time |
|---|---|---|
| 1. Initial denaturation | 94°C | 3 mins |
| 2. Denaturation | 94°C | 20 s |
| 3. Annealing | 50-60°C | 20 s |
| 4. Extension | 72°C | 30s/kb |
| Repeat steps 2-4 for 35x cycles | | |
| 5. Final extension | 72°C | 5 mins |
| 6. Hold | 12°C | - |

*2.1.2.2 High fidelity PCR*

To generate constructs for expression in mammalian cells, insect cells or bacteria, inserts were PCR amplified using the Q5 DNA polymerase (NEB). Reactions were carried out in 1x Q5 buffer, usually with 1x GC enhancer buffer, in a 50 µL final volume. PCR reactions were mixed thoroughly and briefly centrifuged. Reactions were then carried out in an Eppendorf Mastercycler Nexus thermal cycler, with cycling parameters varying depending on DNA template and primers. PCR products were then analysed by separation on an agarose gel and were purified by gel extraction using the EZNA Gel Extraction kit (Omega BioTek) and eluted in 30 µL volume.

**Table 2.4 Pipetting scheme for high fidelity PCR.**

| Component | 50 uL final volume |
|---|---|
| 5x Q5 buffer | 10 µL |
| 5x GC enhancer | 10 µL |
| Forward primer (5 µM) | 4 µL |
| Reverse primer (5 µM) | 4 µL |
| dNTP mix (10 mM) | 1 µL |
| Plasmid DNA (10 ng/uL) | 1 µL |
| Q5 polymerase (2 U/µL) | 0.5 µL |
| $H_2O$ | 19.5 µL |

**Table 2.5: Thermal cycling conditions for high fidelity PCR.**

| Cycling step | Temperature | Time |
|---|---|---|
| 1. Initial denaturation | 98°C | 30 s |
| 2. Denaturation | 98°C | 10 s |
| 3. Annealing | 55-60°C | 30 s |
| 4. Extension | 72°C | 30s/kb |
| Repeat steps 2-4 for 35x cycles | | |
| 5. Final extension | 72°C | 10 mins |
| 6. Hold | 12°C | - |

## *2.1.2 Restriction cloning*

### *2.1.2.1 Insert digest*

Following high fidelity PCR, insert DNA was gel extracted, eluted in 30 µL and digested using restriction enzymes from NEB. Inserts were generally digested with two enzymes to generate sticky ends where possible, with digests carried out for 37°C for 1 h, unless manufacturer's instructions directed otherwise. The digest was then separated on an agarose gel and the digested insert purified by gel extraction.

**Table 2.6: Pipetting scheme for insert digest.**

| Component | Volume |
|---|---|
| Insert DNA | 30 µL |
| Enzyme X (10 U/µL) | 1 µL |
| Enzyme Y (10 U/µL) | 1 µL |
| Restriction digest buffer | 3.5 µL |

### *2.1.2.2 Vector digest*

The cloning vector was also digested using appropriate restriction enzymes, generally for 1 h at 37°C. The digest was then separated on an agarose gel and the vector backbone purified by gel extraction.

**Table 2.7: Pipetting scheme for vector digest.**

| Component | Volume |
|---|---|
| Vector DNA | 1 µL (500-1000 ng) |
| Enzyme X (10 U/µL) | 1 µL |
| Enzyme Y (10 U/µL) | 1 µL |
| Restriction digest buffer | 1 µL |
| $H_2O$ | 6 µL |

*2.1.2.3 Ligation reactions*

Following purification of digested insert and vector, ligation reactions were carried out using T4 DNA ligase (Promega, M1804) in Rapid ligation buffer for 1 h on ice, and 4 µL of the ligation reaction was used to transform XL-10 Gold chemically competent *Escherichia coli*.

**Table 2.8: Pipetting scheme for ligations.**

| Component | Volume |
|---|---|
| Insert DNA | 3 µL |
| Vector DNA | 1 µL |
| T4 ligase (1-3 U/µL) | 1 µL |
| 2x Rapid ligation buffer | 5 µL |

### *2.1.3 Overlap extension PCR*

Overlap extension PCR was used to generate deletion mutants, such as CBP core-ΔZZ-TAZ2, as described by (Kanoksilapatham et al., 2007). In the first step, two separate 50 µL PCR reactions were set up to amplify the two parts of the final desired construct by high fidelity PCR. Primers used in each reaction were designed so that one primer has a restriction enzyme overhang for cloning into a vector and the other has 30 nt primer with 15 nt that anneals to the part being amplified and 15 nt that would anneal to the part adjacent to the deletion. These PCR reactions generate “megaprimers” for use in the next step.

In the second step, 5 µL of the two megaprimers were mixed and a 25 µL low cycle number high fidelity PCR was carried out to generate a template carrying the deletion.

**Table 2.9: Pipetting scheme for overlap extension step.**

| Component | 25 uL final volume |
|---|---|
| 5x Q5 buffer | 5 µL |
| 5x GC enhancer | 5 µL |
| Megaprimer 1 | 5 µL |
| Megaprimer 2 | 5 µL |
| dNTP mix (10 mM) | 0.5 µL |
| Q5 polymerase (2 U/µL) | 0.5 µL |
| $H_2O$ | 4 µL |

**Table 2.10: Thermal cycling conditions for overlap extension step.**

| Cycling step | Temperature | Time |
|---|---|---|
| 1. Initial denaturation | 98°C | 30 s |
| 2. Denaturation | 98°C | 10 s |
| 3. Annealing | 72°C | 30 s |
| 4. Extension | 72°C | 30s/kb |
| Repeat steps 2-4 for 10x cycles | | |
| 5. Hold | 12°C | - |

For the final step, the overlap extension reaction was made up to 50 uL by addition of flanking primers carrying restriction enzyme overhangs (as used in the first step to generate the megaprimers) and appropriate volumes of PCR mix, and the newly generated template carrying the deletion was amplified by high fidelity PCR. The PCR product was then separated on an agarose gel and purified by gel extraction, digested, ligated into the desired vector and transformed into XL-10 Gold competent bacteria.

**Table 2.11: Pipetting scheme for amplification of deletion product.**

| Component | 25 uL final volume |
|---|---|
| 5x Q5 buffer | 5 µL |
| 5x GC enhancer | 5 µL |
| Flanking primer forward (5 µM) | 4 µL |
| Flanking primer reverse (5 µM) | 4 µL |
| dNTP mix (10 mM) | 0.5 µL |
| Q5 polymerase (2 U/µL) | 0.5 µL |
| $H_2O$ | 6 µL |

**Table 2.12: Thermal cycling conditions for amplification of deletion product.**

| Cycling step | Temperature | Time |
|---|---|---|
| 1. Initial denaturation | 98°C | 30 s |
| 2. Denaturation | 98°C | 10 s |
| 3. Annealing | 72°C | 30 s |
| 4. Extension | 72°C | 30s/kb |
| Repeat steps 2-4 for 10x cycles | | |
| 5. Final extension | 72°C | 10 mins |
| 6. Hold | 12°C | - |

### *2.1.4 DNA mutagenesis*

Mutagenesis was carried out by PCR using the Quikchange II (Stratagene) strategy. Two 32 nt complementary primers were designed, both annealing to the same sequence with the desired mutation(s) in the middle of the primer sequences. The entire vector was then amplified by high fidelity PCR, and the original vector was digested with *DpnI* (NEB) for 3 h at 37°C. The reaction was then directly transformed into XL-10 Gold competent bacteria.

**Table 2.13: Pipetting scheme for mutagenesis PCR.**

| Component | 50 uL final volume |
|---|---|
| 5x Q5 buffer | 10 µL |
| 5x GC enhancer | 10 µL |
| Plasmid DNA (10 ng/µL) | 1 µL |
| Flanking primer forward (5 µM) | 0.4 µL |
| Flanking primer reverse (5 µM) | 0.4 µL |
| dNTP mix (10 mM) | 1 µL |
| Q5 polymerase (2 U/µL) | 0.5 µL |
| $H_2O$ | 26.7 µL |

**Table 2.14: Thermal cycling conditions for mutagenesis PCR.**

| Cycling step | Temperature | Time |
|---|---|---|
| 1. Initial denaturation | 98°C | 30 s |
| 2. Denaturation | 98°C | 10 s |
| 3. Annealing | 60°C | 30 s |
| 4. Extension | 72°C | 30s/kb |
| Repeat steps 2-4 for 16x cycles | | |
| 5. Final extension | 72°C | 10 mins |
| 6. Hold | 12°C | - |

**Table 2.15: Pipetting scheme for DpnI digest.**

| Component | Volume |
|---|---|
| Mutagenesis PCR product | 50 µL |
| Cutsmart buffer | 6 µL |
| *DpnI* | 1 µL |

## *2.1.5 DNA manipulations*

### *2.1.5.1 Transformation of chemically competent bacteria*

XL-10 Gold chemically competent bacteria (prepared in-house) were transformed using PEG/DMSO. 1-4 µL of plasmid or ligation reaction was mixed on ice with 20 µL of sterile filtered 5x KCM buffer (500 mM KCl, 150 mM $CaCl_2$, 250 $MgCl_2$). 100 µL of competent bacteria were thawed on ice and added to the DNA in 5xKCM and incubated on ice for 10-30 mins. The transformations were then incubated at room temperature for 5-10 mins before 600 uL of Luria broth (LB) was added (10 g/L Tryptone, 5 g/L yeast extract, 10 g/L NaCl, dissolved in distilled H2O and pH adjusted to 7.0 prior to autoclaving). The bacteria were then allowed to recover for 45 mins at 37°C on a shaker. For circular plasmids, 100 uL of the transformation mix was then plated on LB agar plates with selective antibiotic and colonies were grown overnight at 37°C. For ligation reactions, after recovery the bacteria were centrifuged at 1,500 x *g* for 3 mins. The pelleted bacteria were then resuspended in 100 uL of LB and all of the bacteria were plated and colonies were grown overnight at 37°C.

### *2.1.5.2 Isolation of plasmid DNA*

For small scale plasmid preparations, individual bacterial colonies were picked using a pipette tip and used to inoculate 5 mL of LB supplemented with appropriate antibiotic. Cultures were grown overnight at 37°C shaking at 200 rpm and the next morning were collected by centrifugation at 3,750 x *g* for 10 mins. Plasmid DNA was then purified using the EZNA Plasmid Mini Kit (Omega BioTek) according to the manufacturer's instructions, with DNA eluted in 50 µL of elution buffer.

For large scale plasmid preparations (for example, for transfection into mammalian cells), individual bacterial colonies were used to inoculate 125 mL of LB in a 500 mL conical flask supplemented with appropriate antibiotic. Cultures were grown overnight at 37°C shaking at 200 rpm and the next morning were collected by centrifugation at 3,750 x *g* for 10 mins. Plasmid DNA was then purified using the EZNA Plasmid Maxi Kit (Omega BioTek) according

to the manufacturer's instructions. After elution, the isolated DNA was generally concentrated by precipitated with isopropanol. To this end, 0.1 volumes of 3 M sodium acetate pH 5.2, was added to the DNA followed by 0.7 volumes of 100% isopropanol. After mixing by inversion, the DNA was pelleted by centrifugation at 15,000 x *g* for 30 mins. The DNA pellet was then washed with 70% ethanol, centrifuged at 15,000 x *g* for 10 mins and air dried to evaporate residual ethanol. DNA was then resuspended in water, generally to a final concentration of 1 mg/mL.

### *2.1.5.3 Analytical restriction digest*

Isolated plasmids were analysed by restriction digest in 30 µL reaction volumes using restriction enzymes from NEB. Analytical digests were generally carried out at for 1 h at 37°C, unless otherwise specified by the manufacturer. Digests were then analysed by agarose gel electrophoresis.

**Table 2.16: Pipetting scheme for analytical restriction digest.**

| Component | Volume |
|---|---|
| Insert DNA | 2 µL |
| Enzyme X (10 U/µL) | 1 µL |
| Enzyme Y (10 U/µL) | 1 µL |
| Restriction digest buffer | 3 µL |
| $H_2O$ | 13 µL |

### *2.1.6 Sanger sequencing*

All cloned DNA was verified by Sanger sequencing, and sequencing was used to confirm positive clones in genome editing. Sequencing reactions were set up using approximately 20 ng of DNA for PCR products and 200 ng of DNA for plasmids in 1x BigDye sequencing buffer with 0.65 µM sequencing primer and BigDye terminator v3.1 (Applied Biosystems). Reactions were carried out in a thermal cycler and sequenced by Edinburgh Genomics.

**Table 2.17: Pipetting scheme for sequencing reactions.**

| Component | Volume |
|---|---|
| DNA | 1-5.7 µL |
| 5x BigDye buffer | 2 µL |
| BigDye Terminator v3.1 | 1 µL |
| Sequencing primer (5 µM) | 1.3 µL |
| $H_2O$ | to 10 µL |

**Table 2.18: Thermal cycling conditions for sequencing reactions.**

| Cycling step | Temperature | Time |
|---|---|---|
| 1. Initial denaturation | 96°C | 1 min |
| 2. Denaturation | 96°C | 10 s |
| 3. Annealing | 50°C | 5 s |
| 4. Extension | 60°C | 4 mins |
| Repeat steps 2-4 for 25x cycles | | |
| 5. Hold | 12°C | - |

## 2.2 Bacterial protein expression and purification

### *2.2.1 Bacterial expression constructs*

#### *2.2.1.1 ZF-CXXC construct*

The *E. coli* codon optimised 6xHis-CXXC-3F2S-2xGCN4 construct was synthesised by IDT and cloned into pET22 using *XbaI* and *SalI* sites in the ZF-CXXC construct and *XbaI* and *XhoI* sites in the vector.

#### *2.2.1.2 ZZ and TAZ2$^{wt}$ constructs*

Constructs for expression of ZZ and TAZ2 domains were amplified from pFastBac CBP FL with 5' *Acc65I* site and 3' *SalI* site and stop codons. The constructs were then cloned into pET22-CXXC into the *Acc65I* downstream of the N-terminal 6xHis and TEV sites and the *XhoI* site upstream of 3F2S, so that stop codons prevent expression of the C-terminal tags.

#### *2.2.1.3 TAZ2$^{mut}$ construct*

The TAZ2$^{mut}$ sequence carrying R1769E/K1832E/K1850E mutations was synthesised as a gBlock by IDT, amplified by PCR using the same primers as TAZ2$^{wt}$ and cloned into pET22-CXXC using *Acc65I* and *XhoI* sites.

### *2.2.2 Bacterial protein expression*

#### *2.2.2.1 Expression of ZF-CXXC, ZZ and TAZ2 constructs*

The bacterial expression vectors were transformed into BL21 competent bacteria by heat shock. 100 µL of bacteria were thawed on ice, 1 µL of plasmid was added and incubated on ice for 20 mins. The bacteria were then heat shocked at 42°C for 1 min 30 s, followed by incubation on ice for 2 mins. 600 µL of LB were then added to the transformations, and the bacteria were allowed to recover at 37°C on a shaker for 45 mins. 100 µL of the transformation reaction was then plated on LB agar plates supplemented with kanamycin, and colonies were allowed to grow overnight at 37°C.

For protein expression, a single colony was picked and used to inoculate a 25 mL starter culture overnight at 37°C, shaking at 200 rpm. The next morning, the entire starter culture was added to a 500 mL of LB supplemented with kanamycin and the culture was grown at 37°C at 200 rpm for 1-2 h until $OD_{600}$ was approximately 0.6. Protein expression was then induced by addition of Isopropyl β- d-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM and $ZnCl_2$ to a final concentration of 20 μM, and protein expression was allowed to proceed for 3 h at 37°C. The bacteria were then harvested by centrifugation at 6,000 x *g* for 15 mins at 4°C, washed with 1x PBS and either used immediately for protein purification or flash frozen and stored at -80°C.

#### *2.2.2.2 Expression of TEV protease*

TEV protease was expressed in bacteria essentially as described for ZF-CXXC, with some minor modifications. pRK793 was transformed into BL21 bacteria by heat shock and a starter culture and large scale culture were set up as described above. Protein expression was then induced only in the presence of 0.5 mM IPTG, and protein expression was allowed to proceed for 4 h at 30°C before bacteria were harvested.

### *2.2.3 Purification of 6xHis-tagged proteins from bacteria*

Bacterial pellets were resuspended in lysis buffer (20 mM Tris-Cl pH 8, 500 mM NaCl, 0.1% NP40, 0.5 mM PMSF), using 16 mL for a 500 mL culture, and lysed by sonication on ice at 40% power, three times for 33 s with pulses of 1 s on/0.1 s off. The lysate was then cleared by centrifugation at 23,000 x g for 30 mins at 4°C. The 6xHis tagged protein was then purified by IMAC (immobilised metal affinity column) affinity purification using Sepharose 6 Fast Flow Ni-NTA Resin (GE Healthcare). 0.5 mL of packed resin was used for 500 mL of bacterial culture, and the resin was washed twice with lysis buffer before addition of the lysate. The binding reaction was allowed to proceed at for 1 h at 4°C with end over end rotation. The beads were then pelleted by centrifugation at 800 x *g* for

2 mins at 4°C and the unbound flowthrough fraction was removed by aspiration. The beads were resuspended in 20 column volumes of low salt wash buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 20 mM imidazole, 0.1 mM PMSF, titrated to pH 8 with NaOH), added to a 10 mL poly-prep chromatography column (Bio-Rad) and the wash buffer allowed to flow through the column. The column was then washed with 20 column volumes of high salt wash buffer (50 mM $NaH_2PO_4$, 1 M NaCl, 20 mM imidazole, titrated to pH 8 with NaOH), and subsequently equilibrated into low salt with 10 column volumes of low salt wash buffer. Bound proteins were eluted 5-10 times with one column volume of elution buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 250 mM imidazole, titrated to pH 8 with NaOH). Protein concentration in the elution fractions was determined by Bradford assay or by SDS-PAGE followed by Coomassie staining with Instant Blue (Expedeon), and the most concentrated fractions were pooled. For ZF-CXXC, the protein was then cleaved with TEV protease using 1 mg of TEV protease for 30 mg of ZF-CXXC protein. The cleavage reaction was allowed to proceed at 4°C overnight with rotation.

The pooled fractions were then dialysed overnight at 4°C in BioDesign Dialysis Tubing (ThermoFisher) with molecular weight cut off (MWCO) of 8 kDa against 1 L of BC100 (20 mM HEPES pH 8, 100 mM KCl, 10% glycerol, 0.5 mM DTT). The dialysed protein was cleared of precipitate by centrifugation for 10 mins at 4°C at either 5,000 x g or 13,000 x g, for volumes greater or smaller than 1.5 mL, respectively. The concentration of soluble protein was then determined by Bradford assay and its purity assessed by SDS-PAGE followed by Coomassie staining. The purified protein was aliquoted and stored at -80°C.

### *2.2.4 Ion exchange chromatography purification of TEV protease*

Following 6xHis purification, TEV protease was further purified by ion exchange chromatography. TEV protease is expressed with a C-terminal 5xArg tag to impart positive charge to the protein, allowing it to be purified by cation exchange. The protein has a pI of approximately 9.6 and was dialysed into BC100 buffer with pH 7.5 (20 mM HEPES pH 7.5, 100 mM KCl, 10% glycerol, 0.5 mM DTT). Protein was applied to a 1 mL pre-packed MonoS column using a

fast protein liquid chromatography (FPLC) ÄKTA system in BC100. The resin was then washed with 20 column volumes of BC100, and protein was eluted over a linear gradient over 40 mL to a maximum concentration of NaCl of 1 M. Protein fractions were monitored by UV spectrometry at 280 nM, and analysed for concentration and purity by SDS-PAGE followed by Coomassie staining. Pooled fractions were dialysed overnight against 1 L of BC100 (20 mM HEPES pH 8, 100 mM KCl, 10% glycerol, 0.5 mM DTT), and the pure protein was made up to a final concentration of 2 mg/mL before aliquoting and storage at -80°C.

## 2.3 Insect cell protein expression and purification

### *2.3.1 Insect cell culture, freezing and storage*

Sf9 cells were maintained in suspension in serum- and antibiotic-free HyClone CCM3 media (GE Life Sciences) at 27°C and shaking at 125 rpm. Cell density was maintained between 0.5 x $10^6$ and 4 x $10^6$ cells/mL. Centrifugation of cells was avoided, but when necessary cells were pelleted at low speed at 200 x *g* for 5 mins before they were resuspended in media for passaging. To freeze cells for long term storage, cells were pelleted at 200 x *g*, resuspended in Bambanker serum-free cell freezing medium at a density of 32 x $10^6$ cells/mL, and 0.5 mL aliquoted in 2 mL cryovial tubes (Corning), frozen at -80°C overnight in Mr Frosty freezing containers (Nalgene) and transferred to liquid nitrogen. To thaw cells, one aliquot corresponding to 16 x $10^6$ cells was thawed in a 37°C water bath and transferred directly to 20 mL of media. Cell growth was monitored over the following days and cells were passaged when they reached a density of 4 x $10^6$ cells/mL.

### *2.3.2 Insect cell expression constructs*

#### *2.3.2.1 CBP truncation constructs*

CBP truncation constructs were amplified from pFastBac CBP FL using a forward primer that inserts 5' *AgeI*-1xFLAG-*Acc65I* and a reverse primer that inserts 3' stop codons and *HindIII* or *SalI* sites. The inserts were then cloned

into the pFastBacFS vector between the *AgeI* site, downstream of a 6xHis tag and TEV cleavage site, and either *HindIII* or *XhoI* sites.

#### *2.3.2.2 CBP core-ZZ-TAZ2$^{mut}$ (CZT$^{mut}$) construct*

The synthesised gBlock containing the TAZ2$^{mut}$ sequence with R1769E/K1832E/K1850E mutations was amplified with a forward primer containing a 5' *XhoI* site and a reverse primer containing 3' stop codons and a *HindIII* site. The insert was then cloned into the pFB CZT$^{wt}$ vector between an internal *XhoI* site within the coding sequence of the HAT domain of CBP and the *HindIII* site to generate pFB CZT$^{mut}$.

### *2.3.3 Insect cell protein expression*

#### *2.3.3.1 Generation of recombinant bacmids*

To generate recombinant baculoviruses for insect cell expression, recombinant bacmids were first generated using the Bac-to-Bac procedure. The insert of interest was cloned into the pFastBac vector and 1 µg of miniprep DNA was used to transform 50 µL of EmBacY competent cells by heat shock. EmBacY cells were thawed and incubated with the DNA on ice for 30 mins. The bacteria were then heat shocked at 42°C for 45 s. 200 µL of LB were immediately added to the transformation and the bacteria were allowed to recover for 4-6 h at 37°C on a shaker, before 80 µL of the transformation mixture was plated on an LB agar plate supplemented with 1 x kanamycin, 1 x tetracycline, 1 x gentamycin, 100 µg/mL Bluo-gal and 168 µM IPTG. Plates were incubated at 37°C overnight protected from light and blue colour was allowed to develop for a further 24 h. Successful transposition of the insert into bacmid DNA results in disruption of the lacZ reporter gene, resulting in white colonies. Therefore, individual white colonies were picked, re-streaked on selective plates and used to inoculate 5 mL cultures grown overnight at 37°C in the presence of 1 x kanamycin, 1 x tetracycline, 1 x gentamycin, with six individual colonies picked per construct. The following day, the bacteria were collected by centrifugation at 3,750 x g for 10 mins and recombinant bacmid DNA was prepared using the EZNA plasmid mini kit (Omega BioTek).

*2.3.3.2 Generation of recombinant baculoviruses*

To generate recombinant baculoviruses, the recombinant bacmids were transfected into Sf9 cells. Sf9 cells were seeded onto wells of a 6 well plate at a density of 1.8 x $10^6$ cells per well in 2 mL of CCM3 media and allowed to settle for 1 h. For the transfections, 8 µL of X-tremeGENE HP transfection reagent (Roche) and 1 µg of fresh bacmid DNA were mixed and incubated at room temperature in 100 µL of CCM3 media in a 1.5 mL Eppendorf tube for 30 mins. The transfection reagent:DNA complexes were then added to the cells dropwise, alongside an untransfected well as a negative control. The transfected cells were sealed with parafilm and transferred to a 27°C incubator and the transfections allowed to proceed for 4-7 days. The recombinant bacmids contain a GFP reporter gene to monitor the transfections. Therefore, once strong GFP signal could be observed, the recombinant baculoviruses that are released from lysed cells were harvested by collecting the culture medium. The virus was cleared of cells by centrifugation at 500 x *g* for 5 mins and the supernatant taken as the P1 virus. A small amount of P1 virus was retained and foetal bovine serum (FBS; Gibco) was added to a final concentration of 5% for long term storage at 4°C protected from light, and the rest was used for amplification of the virus.

*2.3.3.3 Amplification of recombinant baculoviruses*

To generate sufficient viral titre for efficient protein expression, the viruses underwent several rounds of amplification. 20 mL of Sf9 cells at a density of 2 x $10^6$ cells/mL were inoculated with the P1 virus, and viral infection was allowed to proceed for 5 days. The cells were harvested and pelleted by centrifugation at 500 x *g* for 5 mins and the supernatant was taken as the S1 virus. FBS was added to the virus to a final concentration of 5% for storage at 4°C. To generate S2 virus, 200 µL of S1 virus was used to inoculate 20 mL of Sf9 cells at a density of 2 x $10^6$ cells/mL, and the resulting virus was harvested and processed in the same way as for S1 virus. The amplification process was repeated until an S4 virus was generated, which could then be used for protein expression. Importantly, before generation of S4 virus, cells were taken from the S3 virus culture, boiled in 1xSDS loading buffer and analysed by SDS-

PAGE followed by western blot to ensure that cells were expressing the protein construct.

#### *2.3.3.4 Infection of insect cells for protein expression*

For large-scale expression of proteins in Sf9 cells, cultures were expanded to 500-1000 mL volumes at a density of 4 x $10^6$ cells/mL and were inoculated with S4 virus at a ratio of 1:100, so that 5 mL of S4 virus was added to a 500 mL culture. The infection was allowed to proceed for 48-72 h at 27°C and shaking at 125 rpm, and the cells were harvested by centrifugation in 250 mL conical tubes (Corning) at 1,500 x *g* for 15 mins at 4°C. The pellets were then snap frozen and stored at -80°C until required for protein purification.

### *2.3.4 Purification and concentration of FLAG-tagged proteins from insect cells*

Insect cell pellets from 500 mL culture were resuspended in 30 mL of lysis buffer (20 mM Tris-Cl pH 8.0, 350 mM NaCl, 10% glycerol, 10 µM $ZnCl_2$, 0.1% NP40, 0.5 mM PMSF, 1 mM DTT), divided between two 50 mL falcon tubes and lysed by sonication on ice for 3x 30 s at 30% power with pulses of 1 s on/0.1 s off. The lysate was cleared by centrifugation at 40,000 x *g* for 30 mins at 4°C and the supernatant filtered through a syringe-driven 0.45 µm PVDF membrane (Millipore, SLHVM25NS). To prepare the beads, 100 µL of packed FLAG M2 affinity resin (Sigma) was washed twice with 1 mL of BC100 (20 mM Tris-Cl pH 8.0, 100 mM NaCl, 10% glycerol, 1 mM DTT), once with 200 µL of 0.1 M glycine pH 2.5, twice with 200 µL of 1 M Tris-Cl pH 8 and twice with 1 mL of lysis buffer, with the resin collected after each wash step by centrifugation at 800 x *g* for 2 mins at 4°C. The resin was then resuspended in 0.5 mL of lysis buffer, added to the cleared lysate and incubated with rotation for 2 h at 4°C. The resin was collected by centrifugation at 800 x *g* for 2 mins at 4°C and the flowthrough fraction removed by aspiration. The flowthrough was then usually re-applied to fresh resin for further protein purification. The bound resin was then washed three times for 10 mins with 10 mL of BC350 wash buffer (20 mM Tris-Cl pH 8.0, 350 mM NaCl, 10% glycerol, 10 µM $ZnCl_2$, 0.5 mM PMSF, 1 mM DTT),

once in 1 mL of BC100 and transferred to a 1.5 mL protein LoBind tube (Eppendorf). Elution under native conditions was carried out by competition with 3xFLAG peptide (Sigma). 3xFLAG peptide was received as lyophilized powder and reconstituted to a final stock concentration of 5 mg/mL in Tris-buffered saline (TBS; 50 mM Tris-Cl, pH 7.9, 150 mM NaCl) and stored at -20°C. Bound protein was eluted three times for 30 mins in 0.3 mg/mL 3xFLAG peptide diluted in BC100. The concentration of purified protein was then measured by Bradford assay and purity was assessed by SDS-PAGE followed by Coomassie staining. Protein samples were then either processed further or aliquoted and stored at -80°C.

For crosslinking mass spectrometry (XL-MS), the presence of the primary amine Tris interferes with crosslinking reactions. Therefore, protein purified for XL-MS was dialysed three times against HEPES-containing BC100 buffer (20 mM HEPES pH 8, 100 mM KCl, 5% glycerol, 1 mM DTT). For XL-MS and electrophoretic mobility shift assays (EMSAs), proteins were required at higher concentrations. Therefore, after purification, protein samples were concentrated using Amicon Ultra 0.5 mL centrifugal filters with Ultracel-30 regenerated cellulose membrane and MWCO of 30 kDa (Millipore, UFC503008) or Amicon Ultra-15 centrifugal filters with Ultracel-3 regenerated cellulose membrane and MWCO of 3 kDa (Millipore, UFC900308). Samples were applied to the filter and centrifuged at 14,000 x *g* for 0.5 mL units or at 4,000 x *g* for 15 mL units for 5-15 mins at 4°C, until samples were concentrated to approximately 1 mg/mL. Protein samples were then aliquoted and stored at -80°C.

## 2.4 Mammalian tissue culture methods

### *2.4.1 Mammalian tissue culture media*

Medium for embryonic stem (ES) cells comprised Dulbecco's modified Eagle's medium (DMEM) (high glucose, without sodium pyruvate, Gibco, 41965062) supplemented with 15% foetal bovine serum (FBS; One Shot FBS, Gibco), 100 µg/mL penicillin/streptomycin (Gibco), 2 mM L-glutamine (Gibco), 1x non-essential amino acids (Gibco), 50 µM β-mercaptoethanol (Sigma) and 10 ng/mL leukaemia inhibitory factor (LIF). For antibiotic selection, puromycin (Gibco, A1113803) was used at 1.5 µg/mL concentration. For experiments using 4-hydroxytamoxifen (4-OHT) (Sigma, H7904-5MG), the powder was resuspended in ethanol to a final concentration of 8 mM for a 10,000x stock that was stored at -20°C. 4-OHT was then added to media to a concentration of 800 nM. Media was stored at 4°C and warmed before use.

### *2.4.2 Culturing, thawing and freezing ES cells*

ES cells were grown in tissue culture dishes (Greiner) coated in 0.1% gelatin (Sigma, G1393), incubated at 37°C in a 5% $CO_2$ atmosphere. To passage cells, the medium was aspirated and the cells were first washed with 1x PBS (Life Technologies, 70013065). 0.05% trypsin-EDTA reagent (Gibco, 25300-062) was added to cover the cells, usually 2.5 mL for a 10 cm plate and 5 mL for a 15 cm plate, and the cells were trypsinised at 37°C for 5 mins. The trypsin was quenched with an equal volume of media, cells were dispersed by pipetting, centrifuged at 500 x *g* for 5 mins where necessary, and added to a fresh tissue culture plate.

To freeze cells for storage and future culture, cells were trypsinised, quenched, counted and centrifuged as described above. Cells were then resuspended in Bambanker serum-free cell freezing medium at a density of 2 million cells/mL and 0.5 mL aliquoted in 2 mL cryovial tubes (Corning), frozen at -80°C overnight in Mr Frosty freezing containers (Nalgene) and transferred to liquid nitrogen.

To thaw cells, a cryovial containing 1 million cells was gently thawed in a 37°C water bath and added directly to either a 10 cm plate or one well of a 6-well plate, depending on the cell line. The cells were allowed to settle and grow overnight, and the following morning the media was changed to aspirate any dead cells that failed to attach to the tissue culture plate.

### *2.4.3 ES cell transfections*

For transfections to express proteins stably or transiently in ES cells, Lipofectamine 3000 reagent (Thermo Fisher) was used. Note that transfections for CRISPR/Cas9 genome editing and dCas9 targeting experiments are described in detail separately.

Transfections were carried out with a DNA:Lipofectamine:p3000 reagent ratio of 1:2:3. For one well of a 6-well plate this corresponds to 1 µg DNA + 2 µL P3000 + 3 µL lipofectamine, for a 10 cm plate to 5 µg DNA + 10 µL P3000 + 15 µL lipofectamine. DNA was diluted in 100 µL of Opti-MEM reduced serum media (Gibco) and mixed, and P3000 reagent was subsequently added to the same tube. Lipofectamine was separately diluted in 100 µL of Opti-MEM. The DNA/P3000 mix was then added to the lipofectamine, mixed and incubated at room temperature for 20 mins. The DNA:lipofectamine complexes were then added dropwise to cells which had been changed to medium from which penicillin/streptomycin was excluded at least 1 h before transfection. The transfection was then allowed to proceed overnight, and the following morning media was changed to regular complete media. For transient transfections, cells were harvested by trypsinising or scraping 48 h after transfection. To select stable cell lines, cells were passaged on to a 10 cm or 15 cm plate and puromycin applied at a concentration of 1 µg/mL.

### *2.4.4 Isolation of stable ES cell clones*

To isolate ES cell clones after antibiotic selection, the cells were washed with 1x PBS and then covered with 1x PBS whilst colonies were picked. ES cell colonies were picked by aspirating under a microscope with a 200 µL pipette. Each colony was then transferred directly into one well of a V-bottomed 96 well plate containing 30 µL of 0.05% trypsin-EDTA. The colonies were then trypsinised for 5 mins in a 37°C incubator, dispersed by pipetting and transferred to a flat-bottomed 96 well plate with ES cell medium without puromycin. After 24 h, the media was changed to media supplemented with 1 µg/mL puromycin.

### *2.4.5 Immunofluorescence (IF) of ES cells*

For IF of ZF-CXXC-3F2S-Venus-ERT2 cell lines, cells were seeded and grown on cover slips. Following 4-OHT treatment, cells were fixed with 4% formaldehyde for 20 mins and cells were permeabilised in 0.5% Triton X-100 for 10 mins. Cells were then stained with 4, 6-diamidino-2-phenylindole dihydrochloride (DAPI) nuclear stain and mounted on glass slides with SlowFade (Thermo Fisher) and imaged with a Zeiss fluorescence microscope.

## 2.5 ZF-CXXC affinity purification (CAP)

### *2.5.1 Chromatin preparation for CAP*

Ten 15 cm plates of ES cells were grown to confluency, harvested by trypsinisation, washed in 1x PBS, resuspended in 60 mL of 1x PBS and split between 6x 50 mL falcon tubes. Cells were then crosslinked by addition of 16% methanol-free formaldehyde (Pierce, 11586711) to a final concentration of 1% and incubated for 10 mins at room temperature. The crosslinking reactions were then quenched by addition of glycine to a final concentration of 125 mM, and the cells were pelleted by centrifugation at 800 x *g* for 4 mins.

To prepare nuclei, the cells were lysed in LB1 (50 mM HEPES pH 8, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP40, 0.25% Triton X-100, 0.5 mM PMSF) for 10 mins at 4°C. The nuclei were recovered by centrifugation at 800 x *g* for 4 mins, and were subsequently washed in LB2 (10 mM Tris-Cl, pH 8, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1 mM PMSF) for 10 mins at 4°C. The nuclei were again recovered by centrifugation at 800 x *g* for 4 mins, and resuspended in 6 mL of LB3 (10 mM Tris-Cl, pH 8, 200 mM NaCl, 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine, 0.1 mM PMSF) and split between six 15 mL hard plastic polystyrene falcon tubes for sonication. Sonication probes were inserted and the nuclei were sonicated using a Bioruptor (Diagenode) on the high power setting for 15 mins with pulses of 30 s on/30 s off, giving a total sonication time of 7 mins 30 s and average fragment sizes of 1 kb.

The sonicated chromatin was mixed and 10% Triton X-100 dissolved in LB3 was added to a final concentration of 1%. The chromatin was cleared by centrifugation at 15,000 x *g* for 10 mins at 4°C and the supernatant was taken as the chromatin extract. Chromatin was then either used immediately for CAP or aliquoted and stored at -80°C.

### *2.5.2 Chromatin CAP*

To purify chromatin using ZF-CXXC protein, ZF-CXXC was first immobilised on StreptactinXT Superflow resin (IBA) to saturate the beads with protein. For example, with 10 mg of ZF-CXXC, 1 mL of packed StreptactinXT resin was used to immobilise the protein. The beads were washed twice with BC100 (20 mM HEPES pH 8, 100 mM KCl, 10% glycerol, 0.5 mM DTT) and recovered by centrifugation at 800 x *g* for 2 mins and the protein was added and incubated with the beads with end-over-end rotation for 1 h at 4°C. The protein-bound beads were pelleted by centrifugation at 800 x *g* for 2 mins and washed once for 10 mins with CAP1000 buffer (50 mM HEPES pH 8, 10% glycerol, 0.1% Triton X-100, 1000 mM NaCl) to remove any bound DNA and twice with LB3 (10 mM Tris-Cl, pH 8, 200 mM NaCl, 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine, 0.1 mM PMSF) to equilibrate the beads. 3 mL of chromatin was added to each of the ZF-CXXC$^{wt}$ and control beads (either beads only or bound by ZF-CXXC$^{K616A}$), and the chromatin was incubated with the beads for 3 h. The beads were then washed three times for 10 mins with CAP150 (50 mM HEPES pH 8, 10% glycerol, 0.1% Triton X-100, 150 mM NaCl) for optimization experiments or CAP300 (50 mM HEPES pH 8, 10% glycerol, 0.1% Triton X-100, 300 mM NaCl) for later experiments.

For optimization experiments, bound chromatin was eluted by successive wash steps with one column volume of CAP300/500/700/1000 (50 mM HEPES pH 8, 10% glycerol, 0.1% Triton X-100, 300/500/700/1000 mM NaCl), followed by three elutions with one column volume of buffer BXT (IBA; 100 mM Tris-Cl pH 8, 150 mM NaCl, 1 mM EDTA, 50 mM D-biotin), and finally boiling the beads in one column volume of SDS-PAGE loading buffer. For later experiments, chromatin was directly eluted three times in 0.5 column volumes of buffer BXT.

The elution, input and flowthrough fractions were analysed by SDS-PAGE followed by western blot or by quantitative PCR (qPCR).

## 2.6 CRISPR-Cas9 genome editing

### *2.6.1 Constructs and guide RNAs*

To target the MLL3 (NCBI gene ID: 231051) and MLL4 (NCBI gene ID: 381022) loci to insert N-terminal tags using CRISPR/Cas9 genome editing technology, homology-directed repair (HDR) templates of 1 kb size were designed and synthesised by Dundee Cell Products for MLL3 or IDT for MLL4. The HDR templates were amplified by PCR and cloned into the pUC19 vector for transfection into ES cells.

Appropriate guide RNAs were identified using the CRISPR design tool (http://tools.genome-engineering.org). Guide oligos with 20 nt complementary regions and with cloning overhangs were synthesised by IDT and reconstituted in $H_2O$ at 100 µM concentration. To express the guide RNAs as single guide RNAs (sgRNAs) in cells, the oligos were annealed and cloned into the pX458 vector (Addgene plasmid number: 48138), which co-expresses an sgRNA from a U6 promoter together with a Cas9-T2A-GFP construct. Oligos were mixed in equimolar amounts, phosphorylated by T4 polynucleotide kinase (PNK; NEB) and annealed.

**Table 2.19: Pipetting scheme for phosphorylation/annealing reactions.**

| Component | Volume |
|---|---|
| Oligo top (100 µM) | 1 µL |
| Oligo bottom (100 µM) | 1 µL |
| 10x T4 PNK buffer | 1 µL |
| T4 PNK | 1 µL |
| $H_2O$ | 6 µL |

The oligos were phosphorylated and annealed in a thermal cycler by incubating at 37°C for 30 mins, boiling at 95°C for 5 mins, and then reducing the temperature in 5°C intervals, holding each temperature for 1 min, to a final temperature of 25°C. The phosphorylated and annealed oligos were then diluted 1 in 200 in $H_2O$ and ligated into pX458.

The oligo duplexes were ligated into pX458 in a digestion/ligation reaction. pX458 was digested with the restriction enzyme FastDigest *BbsI* (also known as *BpiI*; Thermo Fisher) and the oligo duplexes were annealed into the generated overhangs by T7 ligase (Thermo Fisher).

**Table 2.20: Pipetting scheme for digestion/ligation reactions.**

| Component | Volume |
|---|---|
| pX458 plasmid | X µL (100 ng) |
| Diluted oligo duplex | 2 µL |
| 10x Tango buffer | 2 µL |
| 10 mM DTT | 1 µL |
| 10 mM ATP | 1 µL |
| FastDigest *BbsI (BpiI)* | 1 µL |
| T7 ligase | 0.5 µL |
| $H_2O$ | to 20 µL |

**Table 2.21: Thermal cycling conditions for digestion/ligation reactions.**

| Cycling step | Temperature | Time |
|---|---|---|
| 1. Initial denaturation | 37°C | 5 min |
| 2. Denaturation | 21°C | 5 mins |
| Repeat steps 1-2 for 6x cycles | | |
| 3. Hold | 12°C | - |

Residual linearized DNA was digested with PlasmidSafe exonuclease (Thermo Fisher) at 37°C for 30 mins, followed by heat inactivation at 70°C for 30 mins.

**Table 2.22: Pipetting scheme for PlasmidSafe digest.**

| Component | Volume |
|---|---|
| Ligation reaction | 11 µL |
| 10x PlasmidSafe buffer | 1.5 µL |
| 10 mM ATP | 1.5 µL |
| PlasmidSafe exonuclease | 1 µL |

The ligated plasmid was then transformed into XL-10 Gold competent bacteria and colonies were grown on LB agar plates supplemented with ampicillin overnight. Single colonies were picked and 5 mL cultures were grown overnight in LB with ampicillin. Plasmids were isolated using the EZNA Plasmid Mini Kit (Omega BioTek) and were sequenced with U6 sequencing primer.

### *2.6.2 Co-transfection of pX458 and HDR template*

Transfections were performed in duplicate in two wells of a 6-well plate, using 400,000 cells per well. Cells were passaged and counted as normal, and each well was seeded with 400,000 cells in media without penicillin/streptomycin.

Transfections were carried out using Lipofecatmine 2000 (Thermo Fisher). For duplicate transfections, 10 µL of lipofectamine was added to 200 µL of OptiMEM. In a separate tube, 1 µg of pX458 and 1 µg of HDR plasmid was mixed with 200 µL OptiMEM. Both tubes were incubated for 5 mins at room temperature and then mixed and incubated together at room temperature for 20 mins. 200 µL of DNA:lipofectamine complexes were then added dropwise to the ES cells and the transfection was allowed to proceed overnight. The following morning, the media was changed to complete ES media.

### *2.6.3 FACS enrichment of transfected cells*

To isolate single ES cell clones that have been successfully genome edited, fluorescence-activated cell sorting (FACS) was used to enrich for cells that were transfected with GFP-expressing pX458 plasmid. Cells were trypsinised 48 h after transfection, quenched with complete media and pelleted. The cells were resuspended in 1 mL of FACS sorting media comprising serum-free media supplemented with 3x Antibiotic-Antimycotic (Gibco), and passed through a 70 µm cell strained into a FACS tube. GFP-positive cells were sorted by FACS (carried out by Dr Martin Waterfall, University of Edinburgh) into FACS sorting media. 15,000 sorted cells were seeded onto each of two 15 cm plates in complete ES media supplemented with 3x Antibiotic-Antimycotic and allowed to

form colonies over approximately seven days. A further 100,000 sorted cells were used to confirm that genome editing had been successful in bulk by PCR using mutation-specific primers and to optimise conditions for PCR-based genotyping.

### *2.6.4 Isolating ES cell clones*

To isolate ES cell clones, colonies were picked into 30 µL of 1x PBS in a V-bottomed 96 well plate. The colonies were trypsinised by addition of 40 µL of 0.5% trypsin-EDTA (Gibco). The trypsin was quenched by addition of complete media and dispersed by pipetting. 80 µL of cells were then added to a flat-bottomed 96 well plate pre-loaded with an additional 100 µL of media for continuing culture. The remainder of the cells were added to a duplicate 96 well plate for use in genotyping. Typically, 192 clones were picked per genome editing target to ensure multiple positive clones.

### *2.6.5 Genotyping ES cell clones*

When the 96 well plate set aside for genotyping reached confluency, the media was aspirated and DNA was extracted using QuickExtract DNA extraction Solution (Epicentre). 100 µL of QuickExtract was added to each well of the 96 well tissue culture plate, cells were resuspended and transferred to a 96 well PCR plate. The cells were vortexed, heated at 65°C for 6 mins, vortexed again and boiled at 98°C for 2 mins. Genotyping PCRs were then carried out using 2 µL of DNA per reaction, and the extracted DNA was stored at -80°C.

Initial screening PCRs were carried out using mutation-specific primers, in which one primer was designed that anneals to the newly integrated tag and the second primer anneals to a region of the endogenous locus outside the HDR template. PCRs were performed using MangoTaq DNA polymerase, in the presence of Q5 GC enhancer buffer for problematic PCRs, and analysed by agarose gel electrophoresis. The presence of a PCR product of correct size in

the analysed reaction suggested that the clone could have been successfully edited.

To verify successful genome editing, positive clones from the first PCR screen were used in a second set of PCR reactions using two primers that anneal to regions of the endogenous locus outside the HDR template. These reactions amplify the target region regardless of whether editing has taken place to reveal whether clones are homozygous or (trans)heterozygous for the insertion. Reactions were analysed by agarose gel electrophoresis and PCR products corresponding to both edited and unedited alleles were purified by gel extraction and analysed by Sanger sequencing.

For endogenous tagging of MLL3, however, the GC-rich nature of the locus precluded amplification of the target region even after extensive attempts at optimisation. Clones containing at least one successfully tagged allele were therefore confirmed by immunoprecipitation of the tagged protein with FLAG M2 beads from whole cell extracts.

## 2.7 Chromatin reconstitution

### *2.7.1 Nucleosome positioning DNA*

Nucleosome arrays for histone acetyltransferase (HAT) and histone methyltransferase (HMT) assays were reconstituted onto a circular plasmid called p177-601 containing 12 copies of the 601 strong nucleosome positioning sequence (Lowary and Widom, 1998) separated by 30 bp linker DNA sequences. For electrophoretic mobility shift assays (EMSAs) using nucleosome arrays, the 12x 601 array was excised from p177-601 by digestion with *EcoRV* (NEB) restriction enzyme overnight at 37°C. The digest was analysed by agarose gel electrophoresis and the upper band was purified by gel extraction using 12 DNA-binding columns per 250 µg of starting material, and eluting six columns in a single 60 µL elution volume.

**Table 2.23: Pipetting scheme for *EcoRV* digest of p177-601.**

| Component | Volume |
|---|---|
| p177-601 plasmid (1 mg/mL) | 250 µL |
| *EcoRV* (5 U/µL) | 50 µL |
| Cutsmart buffer | 35 µL |
| $H_2O$ | 15 µL |

For reconstitution of nucleosome core particles (NCPs),147 bp and 209 bp 601 DNA was amplified by PCR in 50 µL reactions in 96 well plate format from pBS-601 using a biotinylated forward primer. PCR products were purified using the EZNA Gel Extraction kit (Omega BioTek), using one DNA-binding column per five PCR reactions, and eluting three columns together in a single 50 µL elution to concentrate the DNA.

**Table 2.24: Pipetting scheme for 147 bp 601 PCR.**

| Component | 100x reactions |
|---|---|
| 5x MangoTaq buffer | 1,000 µL |
| Forward primer (50 µM) | 20 µL |
| Reverse primer (50 µM) | 20 µL |
| $MgCl_2$ (50 mM) | 150 µL |
| dNTP mix (10 mM) | 100 µL |
| pBS-601 template (10 ng/µL) | 100 µL |
| MangoTaq (5 U/µL) | 30 µL |
| $H_2O$ | 3,580 µL |

**Table 2.25: Thermal cycling conditions for 147 bp 601 PCR.**

| Cycling step | Temperature | Time |
|---|---|---|
| 1. Initial denaturation | 94°C | 2 mins |
| 2. Denaturation | 94°C | 30 s |
| 3. Annealing | 54°C | 30 s |
| 4. Extension | 70°C | 45 s |
| Repeat steps 2-4 for 35x cycles | | |
| 5. Final extension | 72°C | 3 mins |
| 6. Hold | 12°C | - |

**Table 2.26: Pipetting scheme for 209 bp 601 PCR.**

| Component | 100x reactions |
|---|---|
| 5x MangoTaq buffer | 1,000 µL |
| Forward primer (50 µM) | 100 µL |
| Reverse primer (50 µM) | 100 µL |
| $MgCl_2$ (50 mM) | 200 µL |
| dNTP mix (10 mM) | 100 µL |
| pBS-601 template (10 ng/µL) | 100 µL |
| MangoTaq (5 U/µL) | 30 µL |
| $H_2O$ | 3,370 µL |

**Table 2.27: Thermal cycling conditions for 209 bp 601 PCR.**

| Cycling step | Temperature | Time |
|---|---|---|
| 1. Initial denaturation | 94°C | 2 mins |
| 2. Denaturation | 94°C | 15 s |
| 3. Annealing | 48.9°C | 15 s |
| 4. Extension | 70°C | 10 s |
| Repeat steps 2-4 for 35x cycles | | |
| 5. Final extension | 70°C | 3 mins |
| 6. Hold | 12°C | - |

## *2.7.2 Histone expression and purification*

Histone expression and purification was carried out by other members of the Voigt lab.

Human histones H2A and H2B, and *Xenopus laevis* histones H3 and H4 were expressed in BL21 bacteria from pET3 bacterial expression vectors. A 50 mL starter culture was established overnight in LB and used to seed a 2 L culture. Expression was induced once $OD_{600}$ was approximately 0.6 by addition of 0.2 mM IPTG, and expression was allowed to proceed for 3 h at 37°C. Bacteria were pelleted at 5,000 x *g* for 10 mins at room temperature, washed in cold histone wash buffer (50 mM Tris-Cl pH 7.5, 100 mM NaCl, 1 mM EDTA, 1 mM

benzamidine, 5 mM β-mercaptoethanol), snap frozen in wash buffer and stored at -80°C until required.

To purify histones from inclusion bodies, the bacterial pellets were first thawed in warm water, and the freeze-thaw cycle repeated one more time. The cells were then lysed by sonication on ice at 40% power, three times for 33 s with pulses of 1 s on/0.1 s off. The samples were centrifuged at 23,000 x *g* for 60 mins at 4°C.

The supernatant was aspirated and the pellet was resuspended in 40 mL of histone wash buffer supplemented with 1% Triton X-100 and transferred to a Dounce homogenizer. The inclusion body pellet was first broken up with five strokes using the loose pestle, and the inclusion bodies were released with ten strokes using the tight pestle. The samples were centrifuged at 23,000 x *g* for 10 mins at 4°C to pellet the inclusion bodies.

Inclusion bodies were solubilised by addition of 500 µL of 100% DMSO, which was allowed to soak into the pellet for 15 mins. The pellet was then resuspended in 5 mL of unfolding buffer (Tris-Cl pH 7.5, 7 M guanidinium HCl, 10 mM DTT) and transferred to a Dounce homogenizer. The pellet was broken up first with the loose pestle followed by the tight pestle, as previously, and the sample was sonicated at 40% power for 15 seconds. The lysed inclusion bodies were stirred at room temperature for 1 h and centrifuged at 23,000 x *g* for 20 mins at 4°C. The supernatant was taken as the extracted histone fraction.

Histones were dialysed against Urea Buffer (10 mM Tris-Cl pH 8, 100 mM NaCl, 7 M urea, 1 mM EDTA, 5 mM β-mercaptoethanol) and precipitate was pelleted by centrifugation at 45,000 x *g* for 30 mins at 4°C. The supernatant was briefly sonicated for 15 s at 30% power to break up residual DNA contamination, was then filtered through a dual glass/PVDF filter (Millipore) and purified by tandem ion exchange chromatography. Histones were purified using an ÄKTA FPLC system using a 5 mL HiTrap Q anion exchange column (GE Life Sciences) and a HiTrap SP cation exchange column (GE Life Sciences). Protein was eluted in Urea Buffer using a salt gradient from 0.1 M to 1 M NaCl. Elution fractions were analysed by SDS-PAGE on 15% gels followed by Coomassie staining. Pure

histone-containing fractions were pooled, dialysed against $H_2O$ supplemented with 3 mM β-mercaptoethanol, snap frozen, lyophilized and stored at -80°C.

### *2.7.3 Refolding histone octamers*

Octamer refolding and purification were carried out by other members of the Voigt lab.

Histones were resuspended in unfolding buffer, mixed in equal amounts and dialysed against refolding buffer (10 mM Tris-Cl pH 8, 2 M NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol). Precipitate was pelleted by centrifugation at 15,000 x *g* for 10 mins at 4°C. Refolded octamers were purified using a 24 mL Superdex 200 gel filtration column (GE Life Sciences), elutions were analysed by SDS-PAGE followed by Coomassie staining and octamer fractions were pooled and stored at -80°C.

### *2.7.4 Nucleosome reconstitution*

For nucleosome arrays, DNA and histone octamers were mixed in refolding buffer at a DNA:histone mass ratio of 1:1.5 (typically 10 µg of DNA with 15 µg of histone octamer). For NCPs, DNA and histone octamers were mixed in refolding buffer at a DNA:histone mass ratio of 1:2 (typically 6 µg of DNA with 12 µg of histone octamer).

Reconstitution reactions were transferred to 0.5 mL Slide-A-Lyzer MINI Dialysis Devices 3.5 kDa MWCO (Thermo Fisher) and floated in 200 mL of refolding buffer. 800 mL of TE buffer (10 mM Tris-Cl pH 8, 1 mM EDTA) was then pumped into the refolding buffer using a peristaltic pump at 25 rpm to give a flow rate of 1 mL/min. In this way, the salt concentration of the buffer was gradually reduced over the course of dialysis. The reactions were then further dialysed against 200 mL of TE buffer alone for 5 h. The reconstituted nucleosomes were then stored at 4°C.

Proper reconstitution of NCPs was checked by analysis on a 6% native PAGE gel or 1% native agarose gel. Reconstitution of nucleosome arrays was verified by analysis on a 0.5% native agarose gel or by digestion of 100 ng (by DNA) of nucleosomes with *ScaI* (NEB) for 1 h followed by native PAGE.

## 2.8 *In vitro* protein biochemistry

### *2.8.1 Histone acetyltransferase (HAT) assays*

Enzyme and substrate (nucleosome or octamer) were mixed on ice in 1 x HAT buffer (50 mM Tris-Cl pH 7.5, 10% glycerol, 4 mM DTT) to a final concentration of 75 nM enzyme and 150 nM substrate (approximately 425 ng of histone protein) in a final reaction volume of 25 µL. For $^{3}$H-labelled asays, reactions were started by addition of [$^{3}$H]-acetyl coenzyme A (acetyl-CoA) (Hartmann Analytic) to a final activity of 1.5 µCi (corresponding to 7.8 µM). For unlabelled assays, unlabelled acetyl-CoA (Sigma) was added to a final concentration of 100 µM, from a stock solution of 5 mM acetylCoA made up in 10 mM sodium acetate and stored at -80°C. Reactions were allowed to proceed for 5-60 mins at 30°C in a ThermoMixer (Eppendorf) shaking at 550 rpm. Reactions were stopped by addition of 12.5 µL of 3xSDS loading buffer (to a final 1x concentration) and boiled for 5 mins at 95°C.

For [$^{3}$H]-labelled HAT assays, quenched reactions were then separated by SDS-PAGE on 18% gels to analyse histone proteins and 8% gels to analyse the enzymes and transferred to a PVDF membrane using the TurboTransfer system (Bio-Rad). The membrane was then stained with Coomassie solution (0.5% (w/v) Coomassie Brilliant Blue R-250 (AppliChem), 45% methanol, 10% acetic acid), destained with Coomassie destain solution (45% methanol, 10% acetic acid) and air dried. Once dry, the membrane was exposed to film (Carestream Kodak BioMax MS film, Sigma) through an intensifying screen (BioMax Transcreen, Sigma) at -80°C for 6 h to 5 days before the autoradiograph was developed. The stained membrane and autoradiograph were then imaged using the ChemiDoc Touch (Bio-Rad) imaging system.

For cold HAT assays, quenched reactions were also separated by SDS-PAGE on 18% gels to analyse histone proteins and 8% gels to analyse the enzymes, and were then transferred to a nitrocellulose membrane and analysed by western blot.

### *2.8.2 Histone methyltransferase (HMT) assays*

HMT assays were carried out as for HAT assays, except that reactions were carried out in 1x HMT buffer (50 mM Tris-Cl pH 7.5, 5 mM $MgCl_2$, 4 mM DTT) and reactions were started by addition of 1.5 µCi of [$^3$H]-labelled S-Adenosyl methionine (SAM; Hartmann Analytic). Reactions were incubated at 30°C for 4 h, and quenched and analysed as described for HAT assays.

### *2.8.3 DNA pull down*

For DNA pull down experiments, 20 µL of Streptavidin M-280 Dynabeads (ThermoFisher) were washed twice with 1 mL of 0.5% bovine serum albumin (BSA) in 1x PBS to block the beads and twice with 1 mL of TEN buffer (10 mM Tris-Cl pH 8.0, 1 mM EDTA, 1 M NaCl). The beads were then resuspended in 1 mL of TEN buffer and 1 µg of biotinylated 147 bp 601 DNA was added to the beads (or no DNA was added for control pull downs without DNA). DNA was bound to the beads with rotation for 1 h at 4°C. The beads were then washed once with 1 mL of TEN buffer and twice with 1 mL of binding buffer (50 mM Tris-Cl pH 8.0, 50 mM NaCl, 0.05% NP40, 0.5 µg/mL BSA, 0.5 mM DTT). 1 µg of protein was diluted in 1 mL of binding buffer, added to the beads and incubated with rotation for 2-3 h at 4°C. To remove any unbound protein, the beads were washed three times for 10 mins with 1 mL of wash buffer (50 mM Tris-Cl pH 8.0, 300 mM NaCl, 0.05% NP40, 0.5 µg/mL BSA, 0.5 mM DTT), and the remaining bound protein was then eluted by boiling for 5 mins in 60 µL of 1x SDS-PAGE loading buffer. Binding was then analysed by separation of 10% of input protein and 50% of pull down material on an 18% SDS-PAGE gel followed by western blot using anti-His antibody.

### *2.8.4 Electrophoretic mobility shift assay (EMSA)*

Binding reactions for EMSA experiments were set up in 10 µL volumes in 1x EMSA buffer (20 mM HEPES pH 8, 150 mM KCl, 7.5% glycerol, 0.5 mM DTT). DNA or nucleosomes were added to give a final concentration of 35 nM of DNA or mononucleosome equivalent (approximately 100 ng of 147 bp 601 DNA). Reactions were made up to volume with protein storage buffer and $H_2O$, the reactions were mixed by pipetting, and increasing concentrations of protein were then added to each reaction. The reactions were mixed again by pipetting and incubated on ice for 30 mins.

Binding experiments were analysed by separation on a native PAGE or agarose gel. ZF-CXXC EMSAs were analysed by native PAGE as follows. Native 8% PAGE gels were poured in 1.5 mm casettes (Invitrogen) in 0.5x Tris-Glycine (TG) buffer (1x TG buffer: 25 mM Tris-Cl pH 8, 190 mM glycine). The native PAGE gel was pre-run at 100 V for 1 h at 4°C in 1x TG buffer, then loaded and run for 2 h at 100 V for 1 h at 4°C in 1x TG buffer. The gel was then stained with 1x Sybr safe (Life Technologies) in 1x TG buffer for 15 mins, and imaged using the ChemiDoc Touch system.

All other EMSA experiments were analysed by native agarose gels. Gel pouring equipment was cleaned using diluted Decon90 (Fisher) to remove traces of Sybr safe that might interfere with binding to DNA. Agarose gels were then poured in 0.5x TBE, using 1.2% agarose for 29 bp dsDNA probes, 1% agarose for free 147 bp 601 DNA, 1% agarose for 147 or 209 bp nucleosome core particles (NCPs) and 0.5% agarose for 12x nucleosome arrays. 8.5 µL of each sample was loaded and the gel was run at 100 V (8.3 V/cm) for 20 mins for 1.2% gels, 30 mins for 1% gels and 40 mins for 0.5% gels, in 0.5x TBE buffer at room temperature. The gel was then stained in 1x Sybr safe for 15 mins and imaged using the ChemiDoc Touch system.

To generate 29 bp dsDNA probes for EMSA, complementary 29 nt oligos were each reconstituted in $H_2O$ at 100 µM concentration and 25 µL of each oligo were mixed for annealing. The oligos were annealed in a thermal cycler by incubating at 37°C for 30 mins, boiling at 95°C for 5 mins, and then reducing the

temperature in 5°C intervals, holding each temperature for 1 min, to a final temperature of 25°C. For binding experiments, the duplex DNAs were then diluted 1 in 10 and 3 µL of diluted dsDNA was used in each 10 µL binding reaction.

### *2.8.5 Crosslinking mass spectrometry (XL-MS)*

147 bp NCPs were reconstituted by mixing histone octamer and DNA at a 2:1 ratio as described above and dialysing against a peristaltic pump-derived gradient of HEK buffer (20 mM HEPES, 100 mM KCl, 1 mM EDTA), and then dialysed directly against 200 mL of HEK buffer. The NCPs were then concentrated using Amicon Ultra 0.5 mL centrifugal filters with Ultracel-100 regenerated cellulose membrane and MWCO of 100 kDa (Millipore, UFC510008) to a concentration of 1 mg/mL based on histone octamer. Purified $CZT^{wt}$ protein and concentrated NCP were mixed together at a molar ratio of $CZT^{wt}$:NCP=2:1 in XL buffer (20 mM HEPES pH 8, 100 mM KCl, 1 mM DTT), using 6.6 µg of $CZT^{wt}$ and 3.3 µg of NCP, and incubated for 30 mins on ice to allow complex formation.

Crosslinking reagents were reconstituted immediately before addition to the complex. EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride; Thermo Fisher) was reconstituted at 20 µg/µL in XL buffer, and sulfo-NHS (N-hydroxysulfosuccinimide; Thermo Fisher) was reconstituted at 46 µg/µL in XL buffer, and the two were mixed in a 1:1 volume ratio to give final concentrations of 10 µg/µL EDC and 23 µg/µL sulfo-NHS. The crosslinking reagents were added to the complex at a mass ratio of complex:EDC=1:7, using 70 µg of EDC. The crosslinking reaction was allowed to proceed for 2 h at room temperature in the dark and was quenched by addition of 100 mM Tris-Cl pH 8.

The products of the crosslinking reactions were boiled in 1x LDS-PAGE loading buffer (Invitrogen) supplemented with 50 mM DTT, and analysed by SDS-PAGE on a 4–12% Bis-Tris NuPAGE gel (Invitrogen) run in 1x MES buffer (Invitrogen) followed by Coomassie staining. High molecular weight bands indicative of successful crosslinking were excised and destained in acetonitrile (ACN).

Protein was reduced in-gel with 10 mM DTT (in ammonium bicarbonate (ABC)) for 30 mins at 37°C, and alkylated with 55 mM iodoacetamide (in ABC). Protein was then digested in-gel overnight in 13 ng/µL trypsin (Pierce) at 37°C.

Digested peptides were eluted from the gel in 0.1% trifluoroacetic acid (TFA). Peptides were loaded onto C18 stage tips that had been activated with methanol and washed twice with 0.1% TFA. The stage tips were washed twice more with 0.1% TFA and stored at -20°C until analysed by mass spectrometry.

Liquid chromatography tandem mass spectrometry analysis was carried out by Dr Juan Zhou (Wellcome Trust Centre for Cell Biology) as described in (Abad et al., 2019). Briefly, peptides were eluted from the stage tips in 80% ACN, concentrated, reconstituted in 0.1% TFA and introduced into an Orbitrap Fusion Lumos mass spectrometer. Data were acquired and matched to crosslinked peptides using Xi software (Mendes et al., 2019).

## 2.9 Protein methods

### *2.9.1 Whole cell extract preparation*

ES cells were harvested by centrifugation at 500 x *g* for 5 mins and washed once with 1x PBS. To prepare whole cell extracts, the cell pellet was resuspended in one pellet volume of RIPA buffer (50 mM Tris-Cl pH 8.0, 150 mM NaCl, 10% glycerol, 1% NP40, 1 mM DTT, 0.5 mM PMSF). Cells were lysed on ice for 30 mins with occasional agitation, centrifuged at 15,000 x *g* for 20 mins at 4°C and the supernatant was taken as whole cell extract.

### *2.9.2 Nuclear extract preparation*

ES cells were harvested by scraping in 1x PBS and pelleted at 500 x *g* for 5 mins. Cells were washed once with 1x PBS, resuspended in 10 pellet volumes of NE Buffer A (10 mM Hepes pH 8, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 0.5 mM PMSF) and equilibrated in this hypotonic buffer for 10 mins on ice. Cells were recovered by centrifugation at 1,500 x *g* for 5 mins and the pellet was

resuspended in 3 pellet volumes of NE Buffer A supplemented with 0.1% NP40. The cells were then lysed on ice for 10 mins with occasional agitation. The nuclei were recovered by centrifugation at 500 x *g* for 5 mins and resuspended in 1 pellet volume of NE Buffer B (5 mM HEPES pH 8, 26% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.1 mM PMSF). The salt concentration of buffer B was then raised to 400 mM NaCl by dropwise addition of 5 M NaCl. The nuclear extraction was allowed to proceed for 1 h on ice with occasional agitation, and insoluble material was pelleted by centrifugation at 15,000 x *g* for 20 mins. The supernatant was taken as the soluble nuclear extract.

For salt gradient nuclear extraction, after cell lysis the nuclei were resuspended in NE Buffer B supplemented with 150 mM NaCl and extraction proceeded for 30 mins on ice. The nuclei were recovered by centrifugation at 500 x *g* for 5 mins and the supernatant was taken as the 150 mM nuclear extract fraction. In the same way nuclei were then washed in 300 mM NaCl NE Buffer B, and finally in 450 mM NaCl, with the final extraction followed by centrifugation at 15,000 x *g* for 20 mins.

### *2.9.3 Histone acid extract preparation*

ES cells were harvested and washed once in 1x PBS. Cells were then washed in 1 mL of RSB (10 mM Tris-Cl pH 8, 10 mM NaCl, 3 mM $MgCl_2$) and subsequently lysed in 1 mL of RSB supplemented with 0.1% NP40 for 10 mins on ice. Nuclei were recovered by centrifugation at 500 x *g* for 5 mins and resuspended in 0.5 mL of 5 mM $MgCl_2$. To this, 0.5 mL of 0.8 M HCl was added and acid extraction was allowed to proceed for 20 mins on ice. Insoluble material was pelleted by centrifugation at 15,000 x *g* for 20 mins and the histone-containing supernatant was taken. Histone proteins were precipitated by addition of trichloroacetic acid (TCA) to a final concentration of 25% and incubated on ice for 30 mins. Precipitated histones were pelleted by centrifugation at 15,000 x *g* for 15 mins and washed twice with cold acetone. Follwing the final wash and centrifugation step, the acetone was aspirated and the histone pellet allowed to air dry at room temperature for 10-15 mins. The pellet was then resuspended in 100 µL of 1x SDS-PAGE loading buffer,

adjusted with 1 M Tris-Cl pH 8 if necessary, and boiled at 95°C for 5 mins. The histone extract was centrifuged once more at 15,000 x *g* for 5 mins to remove any insoluble material and the supernatant was taken as the histone extract.

### *2.9.4 Determination of protein concentration by Bradford assay*

The protein concentration of nuclear extracts and purified proteins was determined using the method developed by Bradford (Bradford, 1976). Concentrated Bradford reagent (Protein assay dye reagent, Bio-Rad, 5000006) was diluted 1 in 5 with $H_2O$ and 1 mL of diluted reagent was added to 1.5 mL eppendorf tubes. A series of volumes of the sample protein were added to 1 mL of diluted reagent, together with a standard curve of known concentrations of bovine serum albumin (BSA) ranging from 1-10 mg/mL. Absorbance was measured at 595 nm using a spectrophotomer (Cecil) and comparison of sample values with those of the standard curve allowed the protein concentration to be determined.

### *2.9.5 Small scale immunoprecipitation (IP)*

IPs were set up either using nuclear extract or whole cell extract. For nuclear extract, 0.5 mg of total nuclear extract was taken and made up to 500 µL with BC150 (50 mM HEPES pH 8, 150 mM NaCl, 10% glycerol, 0.5 mM EDTA, 0.5 mM DTT, 0.1 mM PMSF) and a 10% input fraction was taken. For whole cell extract, extracts were made up to 500 µL with RIPA buffer (50 mM Tris-Cl pH 8.0, 150 mM NaCl, 10% glycerol, 1% NP40, 1 mM DTT, 0.5 mM PMSF) and a 10% input fraction was taken. For FLAG IPs, 20 µL of packed FLAG M2 beads were washed twice in BC150, added to the extracts and incubated with rotation at 4°C either overnight or for 2 h. For other IPs, 3 µg of antibody were added to the extracts and the IPs were incubated with rotation overnight at 4°C.

For non-FLAG IPs, the next morning 25 µL of protein A Dynabeads (Thermo Fisher) were washed twice with BC150 and added to the IP reactions for 1 h at 4°C. For both FLAG and other IPs, the beads were then washed three times for

10 mins with BC300 (50 mM HEPES pH 8, 300 mM NaCl, 10% glycerol, 0.5 mM EDTA, 0.5 mM DTT, 0.1 mM PMSF). The beads were then boiled in 60 uL of SDS-PAGE loading buffer for 5 mins, centrifuged for 1 min at 15,000 x *g* and the supernatant was taken as the IP sample. The IP was then analysed alongside input samples by western blot.

### *2.9.6 Large scale purification of 3F2S-MLL4 complex*

To purify 3F2S-MLL4 complex from endogenously tagged ES cells, 50x 15 cm plates of cells were harvested by scraping and washed in 1x PBS. Nuclear extact was prepared as described above, and 10 mg of nuclear extract was used in each purification.

The 10 mg of nuclear extract was diluted in NE buffer B without NaCl (5 mM HEPES pH 8, 26% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.1 mM PMSF) to a final salt concentration of 150 mM. The diluted protein was then centrifuged at 5,000 x *g* for 5 mins to pellet any precipitate. 75 µL of packed FLAG M2 beads were washed once in BC150 (50 mM HEPES pH 8, 150 mM NaCl, 10% glycerol, 0.5 mM EDTA, 0.5 mM DTT, 0.1 mM PMSF), once in 150 µL of 100 mM glycine pH 2.5, twice in 300 µL of 1 M Tris pH 8, and twice more in BC150. Each wash step was followed by centrifugation at 800 x *g* for 2 mins. The prepared beads were then added to the diluted nuclear extract for 3 h and incubated at 4°C with end-over-end rotation for 2-3 h.

The beads were then pelleted by centrifugation at 800 x *g* for 2 mins and the flowthrough fraction taken by aspiration. The beads were then washed three times for 10 mins with BC300 (50 mM HEPES pH 8, 300 mM NaCl, 10% glycerol, 0.5 mM EDTA, 0.5 mM DTT, 0.1 mM PMSF). supplemented with 0.1% NP40, and then twice with BC350 alone. Bound protein was then eluted three times for 30 mins with 2 µg/µL 3xFLAG peptide diluted in BC150.

Purified material was analysed by SDS-PAGE followed by silver stain using the SilverQuest kit (Invitrogen) according to manufacturer's instructions, by western blot and by mass spectrometry. For mass spectrometry analysis, purified

protein was boiled in 1x LDS-PAGE loading buffer (Invitrogen) supplemented with 50 mM DTT and run for approximately 10 mins on a 4-12% Bis-Tris NuPAGE gel (Invitrogen). The protein was stained with Coomassie, excised, reduced by in-gel with 10 mM DTT, alkylated with 55 mM iodoacetamide and digested in-gel overnight in 13 ng/µL trypsin (Pierce) at 37°C. Peptides were then loaded on C18 stage tips and analysed by mass spectrometry by Dr Christos Spanos.

### *2.9.7 SDS polyacrylamide gel electrophoresis (SDS-PAGE)*

Proteins were analysed by SDS-PAGE to separate protein species by molecular weight using the method first described in (Laemmli, 1970). Gels were generally poured using 1.5 mm mini gel cassettes (Invitrogen, NC2015). The stacking gel comprised 5% acrylamide/bis solution (37.5:1, VWR), 125 mM Tris-Cl pH 6.8, 0.1% SDS, 0.1% ammonium persulphate (APS) and 0.1% *N*,*N*,*N*′,*N*′-Tetramethylethylenediamine (TEMED, Sigma). Separating gels were poured with an acrylamide concentration between 6% and 18%, 400 mM Tris-Cl pH 8.8, 0.1% SDS, 0.1% APS and 0.1% TEMED. Samples were made up in 1x SDS-PAGE loading buffer (63 mM Tris, 10% glycerol, 2% SDS, 50 mM DTT, 0.1% bromophenol blue) and boiled at 95°C for 5 mins before loading together with a protein marker (either Precision Plus Protein Dual Color Standards (Bio-Rad, 1610374) or HiMark Pre-stained Protein Standard (Invitrogen, LC5699)). Gels were run in 1x SDS-PAGE running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS) at 200 V constant voltage using the Mini gel tank system (Life Technologies, A25977) until the required separation was achieved. Gels were then analysed either by staining with Coomassie (Instant Blue, Expedeon) or by western blot.

### *2.9.8 Western blot*

Following SDS-PAGE, gels were transferred to 0.2 µm nitrocellulose membranes (Bio-Rad) for western blotting. For proteins of interest with sizes up to 120 kDa, gels were transferred using the Trans-Blot Turbo System (Bio-Rad). The membrane was soaked in transfer buffer comprising 1x Trans-Blot Turbo Buffer supplemented with 20% ethanol and placed on top of pre-soaked filter

paper. The gel was placed on top of this membrane and additional pre-soaked filter paper placed on top of the gel. Air bubbles were carefully removed from the transfer sandwich by rolling with a blot roller (Bio-Rad). Transfer was then carried out for 10 mins in a Trans-Blot Turbo Transfer instrument using the settings for 1.5 mm mini gels (1.3 A constant current or 2.3 A constant current for one or two gels, respectively, with maximum voltage of 25 V).

For higher molecular weight proteins, transfer was carried out either by traditional semi-dry transfer or by wet transfer. For semi-dry transfers, the transfer sandwich was set up using three pieces of pre-soaked Whatman filter paper on either side of the gel/membrane stack. Transfer was carried out in semi-dry transfer buffer (48mM Tris, 39mM glycine, 0.037% SDS, 20% methanol) in a Trans-Blot SD Semi-dry transfer cell (Bio-Rad) for 90 mins at 200 mA constant current and maximum voltage of 23 V.

For wet transfers, the transfer sandwich was similarly set up using three pieces of pre-soaked Whatman filter paper on either side of the gel/membrane stack. Transfer was carried out on ice in wet transfer buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, 20% ethanol) in a Mini Trans-Blot cell (Bio-Rad) for 90 mins at 90 V constant voltage.

Following transfer, membranes were blocked in 5% milk (non-fat milk powder, Marvel) made up in 1x PBS supplemented with 0.1% Tween-20 (VWR) (PBST) for 30 mins at room temperature. Primary antibody was made up to an appropriate dilution in 2% BSA/PBST and incubated with the membrane either overnight at 4°C or for 2 h at room temperature. Following incubation with primary antibody, the membrane was washed three times for 5 mins in PBST at room temperature. Species-specific secondary antibodies conjugated to horseradish peroxidase (HRP) were diluted 1 in 5000 in 2% BSA/PBST and incubated with the washed membrane for 1 h at room temperature. Depending on the species of the primary antibody, either Peroxidase-AffiniPure Donkey Anti-Rabbit IgG whole antibody (Jackson ImmunoResearch, 711-035-152) or Peroxidase-AffiniPure Donkey Anti-Mouse whole antibody (Jackson ImmunoResearch, 715-035-150) were used as secondary antibodies. Prior to

developing the western blot, the membrane was washed three times for 5 mins in PBST to remove excess secondary antibody.

Western blots were developed by enhanced chemiluminescence (ECL) using the Clarity Western ECL system (Bio-Rad). ECL peroxide solution was mixed in a 1:1 volume ratio with ECL luminol/enhancer solution and applied to the membrane for 1 min. The membrane was then imaged using the ChemiDoc Touch system (Bio-Rad).

## 2.10 Chromatin immunoprecipitation (ChIP)

### *2.10.1 Chromatin preparation for ChIP*

ES cells were harvested by trypsinisation, washed in 1x PBS, counted and 1 x $10^7$ cells were resuspended in 10 mL of 1x PBS. Cells were then crosslinked by addition of 16% methanol-free formaldehyde (Pierce, 11586711) to a final concentration of 1% and incubated for 10 mins at room temperature. The crosslinking reactions were quenched by addition of glycine to a final concentration of 125 mM (1.5 mL of 1.5 M glycine to each 10 mL crosslinking reaction), and the cells were pelleted by centrifugation at 800 x *g* for 4 mins.

Cells were lysed in 10 mL of LB1 (50 mM HEPES pH 8, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP40, 0.25% Triton X-100, 0.5 mM PMSF) for 10 mins at 4°C. The nuclei were recovered by centrifugation at 800 x *g* for 4 mins, and were subsequently washed in 10 mL of LB2 (10 mM Tris-Cl, pH 8, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1 mM PMSF) for 10 mins at 4°C. The nuclei were again recovered by centrifugation at 800 x *g* for 4 mins, and resuspended in 1 mL of LB3 (10 mM Tris-Cl, pH 8, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine, 0.1 mM PMSF) and transferred to a 15 mL hard plastic polystyrene falcon tube for sonication. Sonication probes were inserted and the nuclei were sonicated using a Bioruptor (Diagenode) on the high power setting for 30 mins with pulses of 30 s on/30 s off, giving a total sonication time of 15 mins and average fragment sizes of 200-300 bp. Following sonication, 10% Triton X-100 dissolved in LB3 was added to a final concentration of 1%. The chromatin was cleared by

centrifugation at 15,000 x *g* for 10 mins at 4°C and the supernatant was taken as the chromatin extract. Chromatin was then either used immediately for ChIP or aliquoted and stored at -80°C.

### *2.10.2 Chromatin size verification*

The size of sonicated chromatin fragments was checked by reversing chromatin crosslinks and analysing the resulting DNA by agarose gel electrophoresis. To this end, 50 µL of chromatin was made up to 100 µL with $H_2O$, 4 µL of 5 M NaCl and 1.5 µL of 20 mg/mL RNase was added, and the sample was incubated at 65°C for 30 mins. Following this incubation, 1 µL of 20 mg/mL proteinase K was added to the sample, followed by a further incubation of 65°C of at least 3 h 30 mins. Decrosslinked DNA was then recovered using the Monarch PCR DNA Cleanup Kit (NEB) and eluted in 20 µL of elution buffer, and approximately 1 µg of DNA was analysed on a 1% agarose gel and imaged using the ChemiDoc Touch.

### *2.10.3 ChIP immunoprecipitation step*

To immunoprecipitate chromatin, antibodies were first conjugated to protein A Dynabeads (Thermo Fisher). 25 µL of beads were aliquoted into 1.5 mL protein LoBind tubes (Eppendorf) and washed three times with 0.5% BSA/PBS. The beads were resuspended in 1 mL of 0.5% BSA/PBS and 1-5 µg of antibody was added. The beads were then blocked and the antibody conjugated to the beads for at least 4 h with end-over-end rotation at 4°C.

For each IP reaction, 100 µL of concentrated chromatin (corresponding to 1 x $10^6$ cells) was used. 100 µL of chromatin was mixed with 900 µL of ChIP dilution buffer (20 mM Tris-Cl pH 8, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1 mM PMSF). A larger mastermix was prepared for each chromatin type and 20 µL of diluted chromatin was taken as 2% input and stored at -20°C. The antibody:bead conjugates were washed three times with 0.5% BSA/PBS and 1

mL of diluted chromatin was added to the beads. The IP reactions proceeded overnight with end-over-end rotation at 4°C.

The following morning, the beads were magnetized and the flowthrough aspirated, and the beads were washed once each with low salt wash buffer (20 mM Tris-Cl pH 8, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), high salt wash buffer (20 mM Tris-Cl pH 8, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100) and LiCl buffer (10 mM Tris-Cl pH 8, 250 mM LiCl, 1 mM EDTA, 1% NP40, 1% sodium deoxycholate) and twice with TE buffer (10 mM Tris pH 8, 1 mM EDTA). The washed beads were then resuspended in 100 µL of ChIP elution buffer (0.1 M $NaHCO_3$, 1% SDS) and shaken viogorously in a ThermoMixer for 30 mins at 25°C. The beads were centrifuged at 15,000 x *g* for 1 min and magnetized, and the supernatant was taken as the ChIP elution.

The input sample was thawed and made up to 100 µL with $H_2O$, and both input and ChIP samples were decrosslinked. To decrosslink the input and ChIP chromatin, 4 µL of 5 M NaCl and 1.5 µL of 20 mg/mL RNase was added, and the sample was incubated at 65°C for 30 mins. Following this incubation, 1 µL of 20 mg/mL proteinase K was added to the sample, followed by a further incubation of 65°C of at least 3 h 30 mins. Decrosslinked DNA was then recovered using the Monarch PCR DNA Cleanup Kit (NEB) and eluted in 50 µL of elution buffer. The purified input and ChIP DNAs were then stored at -20°C until used for quantitative PCR (qPCR).

### *2.10.4 Quantitative PCR (qPCR)*

Quantitative PCR (qPCR) reactions were carried out using a LightCycler (Roche) instrument in either 96 or 384 well plate format. Each reaction was carried out in a total volume of 10 µL, comprising 2.5 µL of DNA, 5 µL of 2x Takyon No Rox SYBR MasterMix dTTP Blue (Eurogentech, UF-NSMTB0701), 0.625 µL of each of 10 µM forward and reverse primers, and 1.25 µL of $H_2O$. Reactions were amplified for 45 cycles. Importantly, a melt curve was generated for each reaction to verify specificity of amplification, each reaction was carried out in duplicate and negative control reactions without template were performed for each primer set in each qPCR experiment.

**Table 2.28: Pipetting scheme for qPCR reactions.**

| Component | 10 µL reaction volume |
|---|---|
| 2x SYBR master mix | 5 µL |
| Forward primer (10 µM) | 0.625 µL |
| Reverse primer (10 µM) | 0.625 µL |
| DNA | 2.5 µL |
| $H_2O$ | 1.25 µL |

**Table 2.29: Thermal cycling conditions for qPCR reactions.**

| Cycling step | Temperature | Time |
|---|---|---|
| 1. Initial denaturation | 95°C | 5 mins |
| 2. Denaturation | 95°C | 10 s |
| 3. Annealing | 60°C | 15 s |
| 4. Extension | 72°C | 15 s |
| Repeat steps 2-4 for 45x cycles | | |

### *2.10.5 qPCR primer design*

Primers for qPCR were designed using the Primer3 program (http://bioinfo.ut.ee/primer3-0.4.0) using DNA sequences obtained from the UCSC genome browser (https://genome.ucsc.edu). Primers were designed so that the product size was 80-120 bp, primer length was 18-27 bp and primers had melting temperature ($T_m$) 59-61°C with maximum $T_m$ difference of 1°C. Primers were checked for off-target amplification using the *in silico* PCR function of the UCSC genome browser. Notably, qPCR primers for reverse transcriptase qPCR (RT-qPCR) were designed to span intron sequences where possible to eliminate potential signal arising from genomic DNA contamination.

## 2.11 Gene expression analysis by reverse transcriptase qPCR

### *2.11.1 RNA extraction*

For gene expression analysis, ES cells were harvested by trypsinisation, washed once with 1x PBS and pelleted by centrifugation at 500 x *g* for 5 mins. Cells were resuspended in 1 mL of TriPure reagent (Roche) and incubated for 5 mins at room temperature. 200 µL of chloroform was then added to each sample and the mixtures were agitated vigorously for 15 s before centrifugation at 12,000 x *g* for 15 mins at 4°C. The upper aqueous fraction containing RNA was transferred to a fresh tube and 500 µL of isopropanol was added. The samples were incubated at room temperature for 10 mins to precipitate the RNA, which was then pelleted at 12,000 x *g* for 10 mins at 4°C. The supernatant was aspirated, the RNA pellet was washed once with 1 mL of 70% ethanol and centrifuged at 7,600 x *g* for 5 mins at 4°C. The supernatant was aspirated and the RNA pellet was air dried for 5-10 mins at room temperature. Finally the RNA pellet was resuspended in 100 µL of RNase-free water and stored at -80°C.

### *2.11.2 Complementary DNA (cDNA) synthesis*

cDNA was synthesised from RNA using the SuperScript IV reverse transcriptase (Invitrogen). 1 µg of RNA was annealed to oligo $(dT)_{20}$ primer at 65°C for 5 mins before incubation on ice for 1 min. The RT reaction was then carried out in a 20 µL volume in 1x SuperScript IV buffer with 5 mM DTT, 20 U of RNase inhibitor and 200 U of SuperScript IV RT enzyme, with three rounds of extension carried out for 10 mins each at 42°C, 50°C and 55°C, followed by heat inactivation of the enzyme for 10 mins at 80°C. The cDNA was then analysed by qPCR as described in section 2.10.4.

**Table 2.30: Pipetting scheme for RT annealing reaction.**

| Component | 13.5 µL reaction volume |
|---|---|
| Oligo $(dT)_{20}$ primer (50 µM) | 1 µL |
| dNTP mix (10 mM) | 1 µL |
| RNA template (100 ng/µL) | 10 µL |
| $H_2O$ | 1.5 µL |

**Table 2.31: Pipetting scheme for RT reaction.**

| Component | 20 µL reaction volume |
|---|---|
| Annealing reaction | 13.5 µL |
| 5x SuperScript IV buffer | 4 µL |
| 100 mM DTT | 1 µL |
| RNase inhibitor (40 U/µL) | 0.5 µL |
| Super script IV RT enzyme (200 U/µL) | 1 µL |

**Table 2.32: Thermal cycling conditions for RT reactions.**

| Cycling step | Temperature | Time |
|---|---|---|
| 1. Extension | 42°C | 10 mins |
| 2. Extension | 50°C | 10 mins |
| 3. Extension | 55°C | 10 mins |
| 4. Inactivation | 80°C | 10 mins |

## 2.12 Rapid immunoprecipitation-mass spectrometry of endogenous elements (RIME)

RIME experiments were carried out by Dr Anca Farcaș (CRUK Cambridge Institute) as part of a collaboration and under my supervision.

### *2.12.1 Chromatin preparation for RIME*

2x 15 cm plates of MCF7 cells were used for each RIME experiment. Cells were first crosslinked on the plate with 20 mL of 2 mM disuccinimidyl glutarate (DSG) (Santa Cruz Biotechnology, sc-285455A) in PBS for 20 mins at room temperature. The DSG was then aspirated and the cells were further crosslinked with 20 mL of 1% formaldehyde for 10 mins at room temperature before quenching with 125 mM glycine. The cells were washed twice in cold 1x PBS and then harvested by scraping in 500 µL of 1x PBS per 15 cm plate supplemented with 1x Complete Protease Inhibitor Cocktail EDTA-free (PIC; Roche). The cells were pelleted at 500 x *g* for 3 mins, washed once in 1 mL of 1x PBS with PIC per 15 cm plate, pelleted and either used directly for chromatin preparation or snap frozen and stored at -80°C.

To prepare chromatin, nuclei were first released from the cells by lysis with 10 mL of ChIP LB1 (see ChIP protocol, section 10.10) with PIC for 10 mins at 4°C. The nuclei were recovered by centrifugation at 800 x *g* for 4 mins and washed with 1 mL of ChIP LB2 with PIC for 5 mins at 4°C. The nuclei were again recovered by centrifugation at 800 x *g* for 4 mins. The nuclei were then resuspend in ChIP LB3 with PIC, using 300 µL per 15 cm plate and divided between 1.5 mL sonication tubes. The nuclei were sonicated using a Bioruptor Pico (Diagenode) using the high power setting for 15 mins mins with pulses of 30 s on/30 s off, giving a total sonication time of 7 mins 30 s and average chromatin fragments of 200-300 bp. To verify the size of chromatin fragments, a 10 µL aliquot of chromatin was taken, bolied at 95°C for 10 mins and analysed using the E-Gel agarose gel electrophoresis system (Invitrogen). 10% Triton X-100 in LB3 was added to the chromatin to a final concentration of 1%, samples were centrifuged at 20,000 x *g* for 10 mins at 4°C to remove insoluble material and the supernatant was taken as the chromatin extract.

### *2.12.2 RIME immunoprecipitation step*

For each RIME experiment, 50 µL of protein A or G Dynabeads (Thermo Fisher) were washed three times with 0.5% BSA/PBS and resuspended in 500 µL of 0.5% BSA/PBS. 5 µg of antibody was then added to the beads and the beads were blocked and conjugated to the antibody overnight at 4°C.

The following day, the beads were washed three times with 0.5% BSA/PBS. The chromatin extract was made up to 1.5 mL volume with 1% Triton X-100 in LB3 and added to the beads, and the IP reaction was allowed to proceed overnight at 4°C. The next morning, the beads were washed 10 times in RIPA buffer (50mM HEPES pH 7.6, 500 mM LiCl, 1mM EDTA, 0.7% sodium deoxycholate, 1% NP-40) and then twice in fresh 0.1 M $NaHCO_3$. The immunocomplexes were then on-bead digested with 10 µL of 15 ng/µL trypsin in $NaHCO_3$ at 37°C and the released peptides analysed by mass spectrometry, as performed in (Mohammed et al., 2013).

## 2.13 dCas9-based genomic targeting

### *2.13.1 Constructs and sequences*

CBP sequences were cloned into the pAC94-pmax-dCas9VP160-2A-puro vector described by (Cheng et al., 2013) by amplifying the desired CBP regions by high fidelity PCR with primers generating a 5' *BamHI* site and a 3' *AclI* site. The pAC94 vector was digested with *BamHI* and *ClaI*, and the amplified CBP sequences were ligated into the vector, taking advantage of the complementary overhangs generated by *AclI and ClaI*.

Guide sequences for the *Ascl1* locus were taken from (Black et al., 2016) and cloned into the pmU6 vector (Kabadi et al., 2014) for expression as sgRNAs. Correct ligation of the guide oligos was determined by sequencing with the mU6 sequencing primer.

### *2.13.2 Expression of dCas9 fusion proteins and sgRNAs in ES cells*

To express the dCas9 fusion proteins and sgRNAs, ES cells were transiently transfected with the constructs for 48 h, using one 15 cm plate per dCas9 fusion protein. Cells were seeded on a 15 cm plate 24 h before transfection and grown in ES media without penicillin/streptomycin. Transfections were carried out using Lipofectamine 3000 (Invitrogen). For each transfection, 30 µL of Lipofectamine 3000 was added to 200 µL of OptiMEM. In a separate tube, 15 µg of pAC94 plasmid DNA and 1.25 µg each of four different pmU6 plasmids were added to 200 µL of OptiMEM, followed 40 µL of P3000 reagent. After 5 mins of separate incubation at room temperature, the lipofectamine and DNA tubes were mixed and incubated for 20 mins together at room temperature. The lipofectamine:DNA complexes were then added dropwise to each 15 cm plate of ES cells and transfections were allowed to proceed overnight. The following morning, the media was changed on transfected cells to complete ES media and cells were cultured for a further 24 h. 48 h after transfection, cells were harvested by trypsinisation, washed in 1x PBS and split for extraction of whole cell extracts, RNA and chromatin.

## 2.14 List of reagents used in this study

### *2.14.1 qPCR primers*

**Table 2.33: List of qPCR primers used in this study.**

| Primer name | Primer sequence | Application |
|---|---|---|
| SUV420h1 CGI for | GACGTGGTTTCTTTGGTGGT | ChIP-qPCR |
| SUV420h1 CGI rev | CGGGAAGAGGCTGAGAGAC | ChIP-qPCR |
| SUV420h1 body for | TGAGGCTCTCAGCAAGACTG | ChIP-qPCR |
| SUV420h1 body rev | ATCTTCCAGGAGAACGAGCA | ChIP-qPCR |
| BCOR CGI for | GTAAAACCGAAAGCGAGCAA | ChIP-qPCR |
| BCOR CGI rev | GAGGGTTTCTCCTCCGACTT | ChIP-qPCR |
| Nanog enhancer for | CGCTCGGATCTTTCACCAGA | ChIP-qPCR |
| Nanog enhancer rev | CGGGTCAAAGGAGTCTGCTT | ChIP-qPCR |
| BRD2 CGI for | TGCTGGGCCTTAGAGAGAAA | ChIP-qPCR |
| BRD2 CGI rev | AGTGATTTTCCGGAATGCAG | ChIP-qPCR |
| ASCL1 -630 bp for | AGGAAGGTAGGAGGGGAGAG | ChIP-qPCR |
| ASCL1 -630 bp rev | TGCTCAGACAGGGTAGAACTTAC | ChIP-qPCR |
| ASCL1 -260 bp for | CAGCCTGGTTTGTTGTTGCA | ChIP-qPCR |
| ASCL1 -260 bp rev | CCCATTTCTAGAGCCACCCC | ChIP-qPCR |
| ASCL1 +100 bp for | CGCTCTCCCTTGCTCCAG | ChIP-qPCR |
| ASCL1 +100 bp rev | CGGTTAGGGAGGGCGAATT | ChIP-qPCR |
| ASCL1 for | GGAACAAGAGCTGCTGGACT | RT-qPCR |
| ASCL1 rev | GTTTTTCTGCCTCCCCATTT | RT-qPCR |
| Nanog for | AGGCTTTGGAGACAGTGAGGTG | RT-qPCR |
| Nanog rev | TGGGTAAGGGTGTTCAAGCACT | RT-qPCR |
| GAPDH for | CATGGCCTTCCGTGTTCCT | RT-qPCR |
| GAPDH rev | GCGGCACGTCAGATCCA | RT-qPCR |

## *2.14.2 Guide RNA oligos*

**Table 2.34: List of guide RNA oligos used in this study.**
Upper case corresponds to complementary regions of guides, and lowercase to overhangs.

| Oligo name | Oligo sequence | Vector |
|---|---|---|
| MLL3 top | caccgCCGACGACATCCTAGTCACC | pX458 |
| MLL3 bottom | aaacGGTGACTAGGATGTCGTCGGc | pX458 |
| MLL4 top | caccgTATCTTCAGCAGGCGGCTTC | pX458 |
| MLL4 bottom | aaacGAAGCCGCCTGCTGAAGATAc | pX458 |
| ASCL1-1 top | ttgtttgCAGCCGCTCGCTGCAGCAG | pmU6 |
| ASCL1-1 bottom | aaacCTGCTGCAGCGAGCGGCTGcaa | pmU6 |
| ASCL1-2 top | ttgtttgTGGAGAGTTTGCAAGGAGC | pmU6 |
| ASCL1-2 bottom | aaacGCTCCTTGCAAACTCTCCAcaa | pmU6 |
| ASCL1-3 top | ttgtttgCCCTCCAGACTTTCCACCT | pmU6 |
| ASCL1-3 bottom | aaacAGGTGGAAAGTCTGGAGGGcaa | pmU6 |
| ASCL1-4 top | ttgtttgCTGCGGAGAGAAGAAAGGG | pmU6 |
| ASCL1-4 bottom | aaacCCCTTTCTTCTCTCCGCAGcaa | pmU6 |

## *2.14.3 EMSA oligos*

**Table 2.35: List of EMSA oligos used in this study.**

| Oligo name | Oligo sequence |
|---|---|
| 60% GC-1 top | GTAGGCGGTGCTACACGGTTCCTGAAGTG |
| 60% GC-1 bottom | CACTTCAGGAACCGTGTAGCACCGCCTAC |
| 60% GC-2 top | AATGGGAACAACCACACCATAGCGATTCG |
| 60% GC-2 bottom | CGAATCGCTATGGTGTGGTTGTTCCCATT |
| 40% GC-1 top | TGAGGAATCCAAAAGGTGAACCAAGCCAG |
| 40% GC-1 bottom | CTGGCTTGGTTCACCTTTTGGATTCCTCA |
| 40% GC-2 top | TAGCCGCTATAATTGTCTCTTTGCCGACT |
| 40% GC-2 bottom | AGTCGGCAAAGAGACAATTATAGCGGCTA |

### *2.14.4 601 nucleosome positioning sequence PCR primers*

**Table 2.36: List of 601 nucleosome positioning sequence PCR primers used in this study.**

| Oligo name | Oligo sequence |
|---|---|
| 601 147 bp for (biotinylated) | ACAGGATGTATATATGTGACAC |
| 601 147 bp rev | CTGGAGAATCCCGGTGCC |
| 601 209 bp for (biotinylated) | GCTTCACCTCGTGACCC |
| 601 209 bp rev | CGCTCTAGACCATGATGC |

### *2.14.5 Antibodies*

**Table 2.37: List of antibodies used in this study.**

| Antigen | Source | Species | Application |
|---|---|---|---|
| FLAG | Sigma FLAG M2 | Mouse mAb | WB: 1 in 1000<br>ChIP: 1 µL |
| 6xHis | Sigma H1209 | Mouse mAb | WB: 1 in 1000 |
| HA | CST | Rabbit mAb | WB: 1 in 1000<br>ChIP: 2 µL |
| Lamin A/C | Santa Cruz sc20681 | Rabbit mAb | WB: 1 in 1000 |
| Tubulin | DHSBAA4-3-5 | Mouse mAb | WB: 1 in 1000 |
| ASH2L | CST 5019 | Rabbit mAb | WB: 1 in 2000 |
| UTX | CST 33510 | Rabbit mAb | WB: 1 in 1000 |
| SUZ12 | CST 3737 | Rabbit mAb | WB: 1 in 1000<br>RIME: 5 µg |
| SUZ12 | Bethyl A302_407A | Rabbit pAb | RIME: 5 µg |
| PSIP1/LEDGF | Abcam ab70641 | Rabbit pAb | RIME: 5 µg |
| PSIP1/LEDGF | Bethyl A300_848A | Rabbit pAb | RIME: 5 µg |
| RNAPII S5P | Abcam ab5131 | Rabbit pAb | RIME: 5 µg |
| CBP | Sanca Cruz sc369(A22) | Rabbit pAb | RIME: 5 µg |
| TBP | Abcam 51841 | Mouse mAb | WB: 1 in 2500 |
| H3K4me3 | CST 9751 | Rabbit mAb | WB: 1 in 1000 |
| H3K9me3 | Abcam ab8898 | Rabbit pAb | WB: 1 in 1000 |
| H3K27ac | CST 8173 | Rabbit mAb | WB: 1 in 1000<br>ChIP: 0.5 µL |
| H3K9ac | Abcam ab4441 | Rabbit pAb | WB: 1 in 1000<br>ChIP: 1 µL |
| H2BK5ac | Abcam ab40886 | Rabbit mAb | WB: 1 in 500 |
| H4K5ac | CST 8647 | Rabbit mAb | WB: 1 in 1000 |
| H2A | CST 3636 | Mouse mAb | WB: 1 in 1000 |
| H3 | Abcam ab1791 | Rabbit pAb | WB: 1 in 1000<br>ChIP: 2 µL |
| H4 | CST 13919 | Rabbit mAb | WB: 1 in 1000 |

## *2.14.6 General buffers and reagents*

### *2.14.6.1 SDS-PAGE loading buffer*

**Table 2.38: SDS-PAGE loading buffer.**

| Component | Final concentration | Amount for 50 mL of 3x SDS-PAGE buffer |
|---|---|---|
| Tris base | 190 mM | 1.151 g |
| Glycerol | 30% | 15 mL of 100% |
| SDS | 6% | 3 g |
| DTT | 150 mM | 1.156 |
| Bromophenol blue | 0.3% | 0.15 g |

### *2.14.6.2 SDS-PAGE running buffer*

**Table 2.39: SDS-PAGE running buffer.**

| Component | Final concentration | Amount for 2 L of 10x running buffer |
|---|---|---|
| Tris base | 250 mM | 60.6 g |
| Glycine | 1.92 M | 288 g |
| SDS | 1% | 20 g |

### *2.14.6.3 Transfer buffers*

**Table 2.40: Wet transfer buffer.**

| Component | Amount for 2 L of 1x wet transfer buffer |
|---|---|
| 10x SDS running buffer | 200 mL |
| Ethanol | 400 mL |

**Table 2.41: Semi-dry transfer buffer.**

| Component | Final concentration | Amount for 1 L of 1x semi-dry transfer buffer |
|---|---|---|
| Tris base | 48 mM | 5.8 g |
| Glycine | 39 mM | 2.9 g |
| SDS | 0.037% | 0.37 g |
| Methanol | 20% | 200 mL |

*2.14.6.4 DNA loading buffer*

**Table 2.42: DNA loading buffer.**

| Component | Final concentration | Amount for 50 mL of 6xDNA loading buffer |
|---|---|---|
| Tris-Cl pH 8 | 10 mM | 500 µL of 1 M |
| Glycerol | 30% | 15 mL of 100% |
| Xylene cyanol | 0.03% | 15 mg |
| Orange G | 0.15% | 75 mg |

*2.14.6.5 TBE*

**Table 2.43: TBE buffer.**

| Component | Final concentration | Amount for 2 L of 10x TBE |
|---|---|---|
| Tris base | 892 mM | 216 g |
| Boric acid | 892 mM | 110 g |
| EDTA | 20 mM | 14.5 g or 80 mL of 0.5 M |

*2.14.6.6 Antibiotics*

**Table 2.44: Antibiotics for prokaryotic culture.**

| Antibiotic | Concentration of 1000x stock |
|---|---|
| Ampicillin | 100 mg/mL |
| Kanamycin | 50 mg/mL |
| Chloramphenicol | 25 mg/mL |
| Tetracycline | 10 mg/mL |
| Gentamycin | 10 mg/mL |

# 3. Understanding the chromatin environment at regulatory elements

Gene expression in eukaryotes is controlled by *cis*-acting gene regulatory elements including gene promoters, present at the transcription start sites (TSSs) of genes, and distal regulatory elements called enhancers. These elements act as platforms for the binding of sequence-specific transcription factors that modulate the transcriptional activity of associated genes. One mechanism through which transcription factors regulate gene expression is via the recruitment of co-factors such as histone modifying proteins and chromatin remodellers that influence the chromatin architecture at promoters and enhancers, changing the accessibility of these regions to the transcriptional machinery.

Gene promoters are frequently associated with regions of unmethylated DNA called CpG islands (CGIs), with some 70% of promoters overlapping with CGIs in mammalian cells (Saxonov et al., 2006). In addition, many “orphan” CGIs that are not found at annotated gene promoters continue to exhibit promoter-like properties, including transcriptional initiation (Illingworth et al., 2010). There is therefore a clear correlation between CGIs and promoter regions, but the mechanisms by which CGIs contribute to promoter function remain incompletely understood.

One mechanism through which CGI function is mediated is via the binding of a domain called the ZF-CXXC domain, which specifically recognizes the unmethylated CpG dinucleotides found at CGIs (reviewed in Long et al., 2013). Indeed, the selectivity of ZF-CXXC for unmethylated DNA is such that ZF-CXXC has previously been used for ZF-CXXC affinity purification (CAP) to purify unmethylated DNA from the genome with high specificity (Blackledge et al., 2012; Illingworth et al., 2008). This domain is found within multiple chromatin-modifying proteins and complexes including the H3K4 methyltransferase complexes SETD1, MLL1 and MLL2, the H3K36 demethylase KDM2A, and the PRC1-associated protein KDM2B (Blackledge et

al., 2010; Farcas et al., 2012; Lee and Skalnik, 2005; Ma et al., 1993; Voo et al., 2000).

One way in which CGIs function at promoters, therefore, is to generate a unique chromatin architecture at these regions, through the recruitment of ZF-CXXC proteins and other chromatin-modifying proteins (Blackledge and Klose, 2011). CGIs are associated with chromatin accessibility, with the presence of histone modifications linked with transcriptional activation such as H3K4 trimethylation (H3K4me3) and histone acetylation, and with the depletion of histone modifications that are antagonistic to transcription initiation such as H3K36 dimethylation (H3K36me2) (Blackledge et al., 2010; Tazi and Bird, 1990; Thomson et al., 2010). This chromatin architecture distinguishes CGIs from the surrounding bulk chromatin and generates an environment that is permissive to transcription. In addition, however, a subset of CGIs is also subject to transcriptional repression. Indeed, the presence of an unmethylated CGI is sufficient for the recruitment of Polycomb repressive complex 2 (PRC2) in embryonic stem cells (ES cells), which generates the repressive histone modification H3K27 trimethylation (H3K27me3) (Mendenhall et al., 2010; Wachter et al., 2014), and endogenous CGIs recruit Polycomb repressive complex 1 (PRC1) to generate monoubiquitylation of H2AK119 (Farcas et al., 2012; He et al., 2013; Wu et al., 2013).

Similarly to CGIs, enhancer regions are thought to function through generating a chromatin environment that influences transcription (reviewed in Calo and Wysocka, 2013). Enhancers are bound by the H3K4 methyltransferases MLL3 and MLL4, which influence transcription directly and through placing H3K4 monomethylation (H3K4me1) (Dorighi et al., 2017; Hu et al., 2013c). Enhancer regions are also associated with binding of histone acetyltransferases (HATs), including CBP and its paralogue p300 (CBP/p300), GCN5 and its paralogue PCAF (GCN5/PCAF), TIP60, and MOF (Krebs et al., 2011; Wang et al., 2009). Enhancers are therefore correlated with multiple forms of histone acetylation, including acetylation of H3K9, H3K14, H3K27 and H4K16 (H3K9ac, H3K14ac, H3K27ac and H4K16ac) (Ernst et al., 2011; Karmodiya et al., 2012; Wang et al., 2009), and the presence of H3K4me1 and H3K27ac is thought to distinguish

active enhancers from the surrounding genome (Creyghton et al., 2010; Rada-Iglesias et al., 2011).

However, although much is now known about the chromatin environment of promoters and enhancers, our understanding of how precisely the chromatin architecture at these regions is established and how it is interpreted to give rise to transcriptional outputs remains incomplete. One way to further our understanding of these processes is to use unbiased proteomics approaches to generate an inventory of proteins bound at gene regulatory regions, as has previously been applied to chromatin regions such as telomeres, pericentromeric heterochromatin, and ribosomal RNA genes (rDNA) (Déjardin and Kingston, 2009; Ide and Dejardin, 2015; Saksouk et al., 2014). Work in this chapter made use of the ZF-CXXC domain to attempt to purify the chromatin associated with CGIs in an unbiased manner to identify the CGI-associated proteome, together with an alternative chromatin immunoprecipitation (ChIP)-mass spectrometry approach.

Complementary to this, candidate approaches can be used to understand better how known regulatory element binding proteins contribute to gene regulatory function. To this end, the MLL3 and MLL4 enhancer-binding proteins were endogenously tagged and purified to understand the complexes they form and how their enzymatic activity is regulated. These techniques permit the identification of novel players at promoters and enhancers to allow a greater understanding of how transcription is regulated from these regions.

## 3.1 A ZF-CXXC affinity purification (CAP) approach to identifying the CGI-associated proteome

To generate an inventory of CGI-associated proteins, an approach was used exploiting the affinity and selectivity of ZF-CXXC for unmethylated DNA (Fig. 3.1). Previous work has used ZF-CXXC domains from MBD1 (Illingworth et al., 2008) and KDM2B (Blackledge et al., 2012) to purify CGI DNA from bulk genomic DNA. To purify CGI-associated proteins, this approach was extended to use recombinantly expressed ZF-CXXC domain from KDM2B to purify the chromatin at CGIs and identify associated proteins by mass spectrometry.

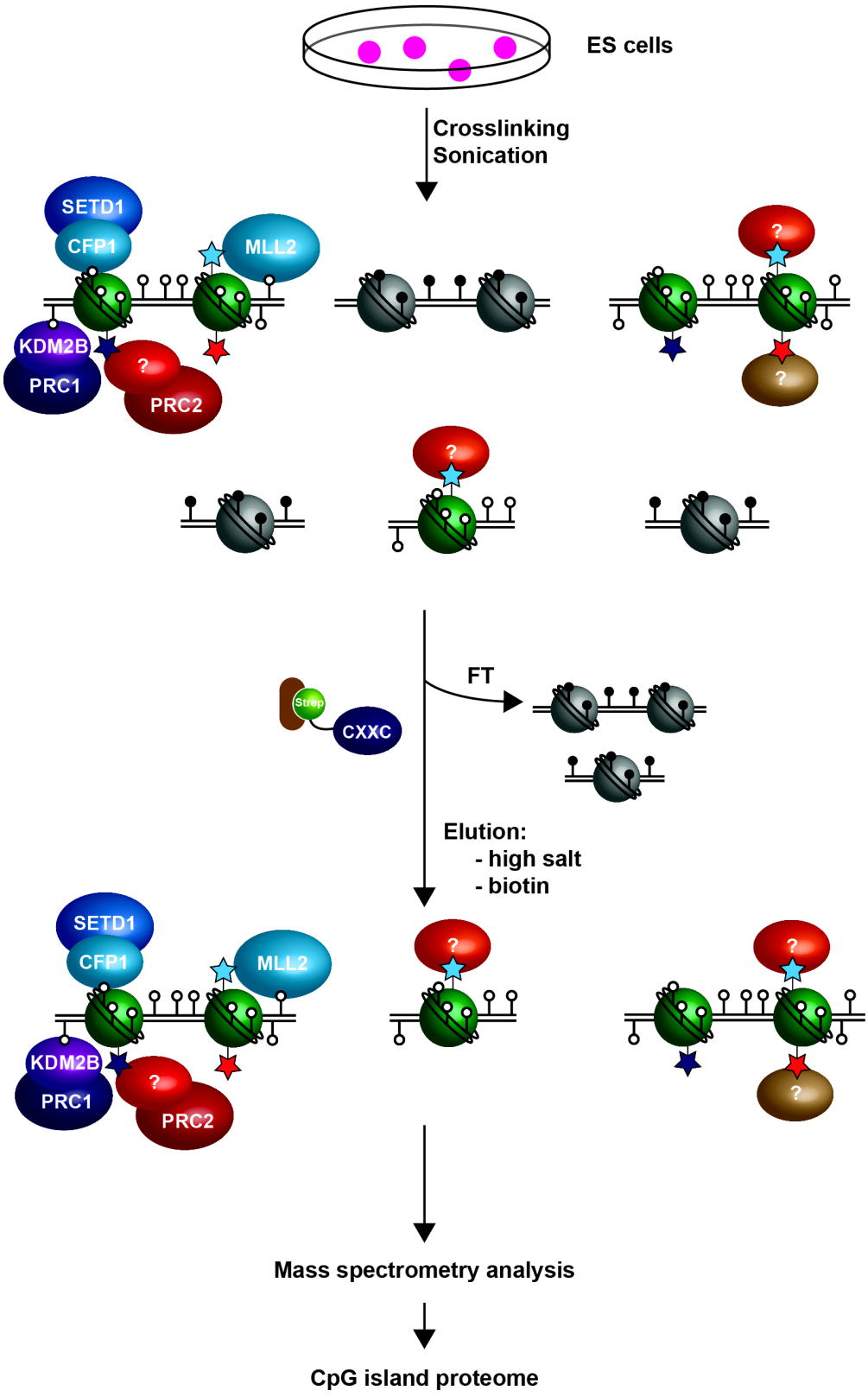



**Fig. 3.1: ZF-CXXC affinity purification (CAP) to identify the CpG island proteome.** A flow diagram illustrating the CAP approach to purify CGI-associated proteins. ES cells were crosslinked and chromatin solubilised by sonication to yield fragments containing regions with DNA methylation (black circles) and with unmethylated DNA at CGIs (white circles), and associated with histone modifications (stars) and both known and unknown proteins. Purified ZF-CXXC protein was then immobilised on StreptactinXT resin via a StrepII tag and used to affinity purify CGI chromatin with associated proteins, which could be identified by mass spectrometry.

First, to purify the large quantities of ZF-CXXC protein required for this approach, the domain was cloned into a bacterial expression vector with an N-terminal 6xHis tag for purification of the protein, and C-terminal 3xFLAG, 2xStrepII and 2xGCN4 tags for immobilisation of the domain on beads during the CAP procedure (Fig. 3.2A). Importantly, as a control for non-specific binding of chromatin to the immobilised protein, a mutant ZF-CXXC that is unable to bind to unmethylated CpG, K616A in mouse KDM2B (ZF-CXXC$^{K616A}$) (Zhou et al., 2012), was generated in addition to the wild type domain (ZF-CXXC$^{wt}$).

The ZF-CXXC proteins were expressed in bacteria and purified via the 6xHis tag on a nickel-NTA (Ni-NTA) affinity column (Fig. 3.2B, C). Previous work has shown, however, that the presence of the N-terminal 6xHis tag interferes with the DNA binding activity of the domain (Blackledge et al., 2012). The 6xHis tag was therefore removed prior to use in downstream applications by cleavage with the tobacco etch virus (TEV) protease (Fig 3.2D). To this end, TEV protease itself was first expressed and purified in bacteria (Fig. 3.3). TEV was expressed as previously described (Tropea et al., 2009) as a self-cleaving maltose binding protein (MBP) fusion protein to aid solubility, and purified via the 6xHis tag on a Ni-NTA affinity column and then further purified by cation exchange to yield essentially pure TEV protease.

After cleavage of the 6xHis tag, the suitability of ZF-CXXC as a module for use in CAP was verified by successfully binding the protein to StreptactinXT beads via the StrepII tags, yielding beads saturated with ZF-CXXC but free of contaminant proteins (Fig. 3.2D). The binding of ZF-CXXC$^{wt}$ specifically to unmethylated DNA was then confirmed by electrophoretic mobility shift assay (EMSA) using a 209 bp DNA probe containing 18 CpG dinucleotides that were either unmethylated or methylated *in vitro* using the *M.SssI* CpG methyltransferase (Fig. 3.2E). ZF-CXXC was able to bind to the unmethylated but not to the methylated DNA probe, and, importantly, that this binding was dependent on ZF-CXXC maintaining an intact structure, as denatured protein was unable to bind to unmethylated DNA. These results show that purified ZF-CXXC is a suitable module for purifying CGIs.

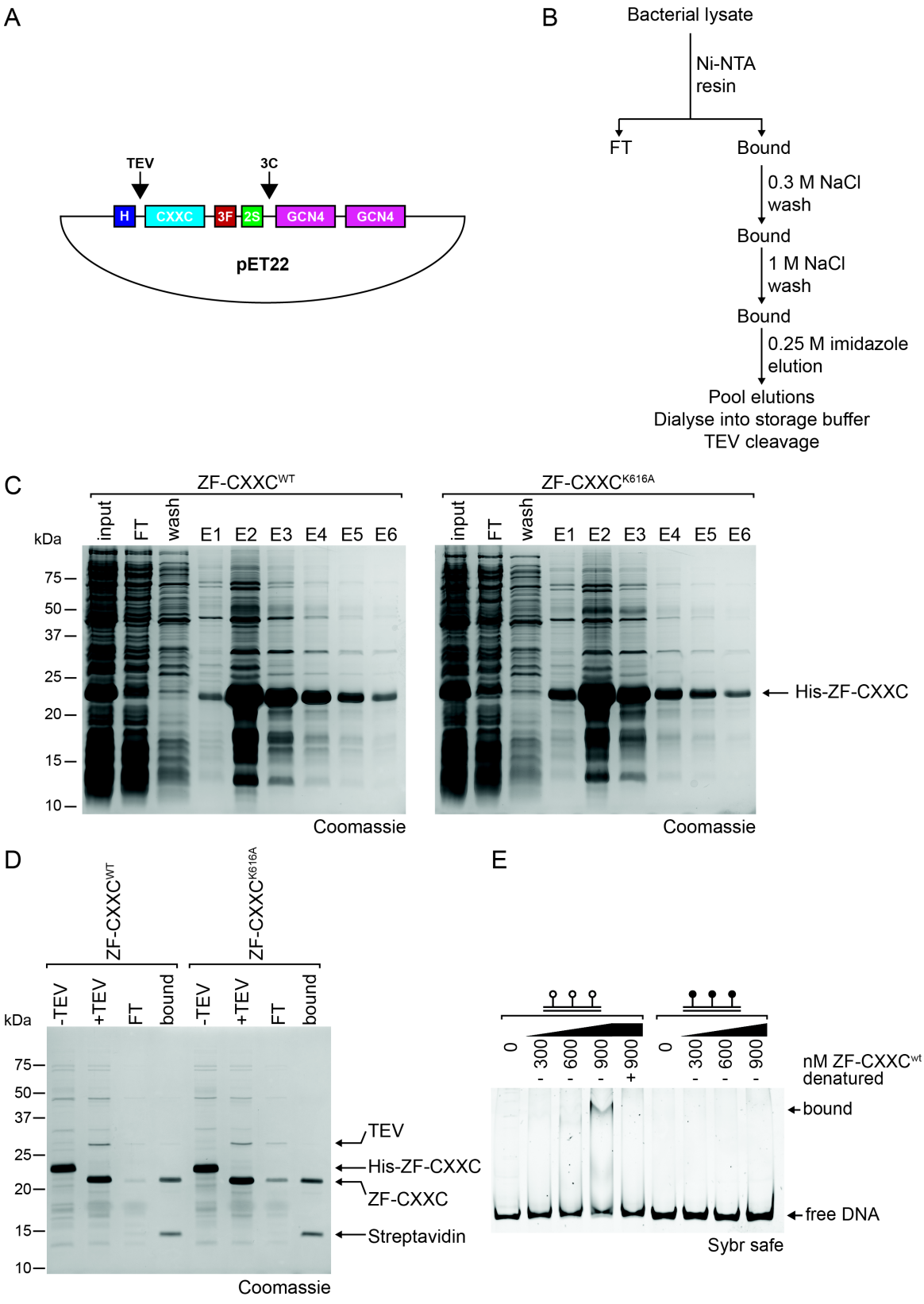



**Fig. 3.2: Purification of ZF-CXXC protein for CAP.**
**(A)** ZF-CXXC (light blue box) was cloned into a bacterial expression and expressed with an N-terminal 6xHis tag (H, dark blue box) separated by a TEV cleavage site, and C-terminal 3xFLAG (3F, red box), 2xStrepII (2S, green box) and 2xGCN4 tags (magenta boxes), the latter of which are separated by a 3C protease site.

## 3.2 Optimisation of CAP to purify CGI chromatin

To attempt to purify CGI chromatin, 25 µg of ZF-CXXC$^{wt}$ were immobilised on StreptactinXT beads and incubated with crosslinked chromatin extracted from mouse embryonic stem (ES) cells. After washing the beads, bound chromatin was eluted using a salt gradient followed by competitive elution of ZF-CXXC$^{wt}$ with biotin. Bound material was analysed by quantitative PCR (qPCR) and by western blot for the CGI-associated histone modification H3K4me3, the CGI-associated protein SUZ12 and H3K9 trimethylation (H3K9me3) as a negative control histone modification that is not found at CGIs (Fig. 3.4A, B). The qPCR analysis of the DNA content of purified chromatin shows that DNA from the CGI of the Suv420h1 gene was enriched in the presence of ZF-CXXC$^{wt}$ but absent from of the beads only control. Moreover, a non-CGI region in the body of the same gene was not enriched by ZF-CXXC$^{wt}$. This suggests that ZF-CXXC$^{wt}$ can specifically enrich for CGI chromatin. However, the levels of enrichment achieved with ZF-CXXC$^{wt}$ were relatively low, and CGI chromatin was only mildly depleted in the unbound flowthrough fraction, suggesting that the majority of CGI chromatin was not captured by ZF-CXXC$^{wt}$. This interpretation is supported by the western blot analysis. Although H3K4me3 was specifically enriched in the ZF-CXXC$^{wt}$ elutions, the level of enrichment was relatively low, and the non-histone CGI-associated protein SUZ12 could not be detected in the purified chromatin. This indicates that the CAP approach had successfully purified CGI chrmatin, but relatively inefficiently.

---

**(Fig. 3.2 cont.)**
**(B)** Flow diagram showing the purification scheme for ZF-CXXC. The protein was expressed in bacteria and the bacterial lysate applied to a nickel-NTA (Ni-NTA) column to purify the protein via the 6xHis tag. The bound protein was washed in 0.3 M and 1 M NaCl conditions to remove protein and DNA contaminants, and eluted in 0.25 M imidazole. The 6xHis tag was then removed by cleavage with TEV protease and the proteins was dialysed into storage buffer.
**(C)** Input, flowthrough (FT), wash and elution fractions for both ZF-CXXC$^{wt}$ and ZF-CXXC$^{K616A}$ proteins were analysed by SDS-PAGE followed by Coomassie staining.
**(D)** ZF-CXXC$^{wt}$ and ZF-CXXC$^{K616A}$ proteins were analysed by SDS-PAGE and Coomassie staining before (-TEV) and after (+TEV) TEV cleavage. Protein was then incubated with StreptactinXT resin and both bound and unbound (FT) material was analysed.
**(E)** Binding reactions with increasing concentrations of intact or denatured ZF-CXXC$^{wt}$ protein with unmethylated (white circles) or methylated (black circles) 209 bp DNA, analysed by EMSA.

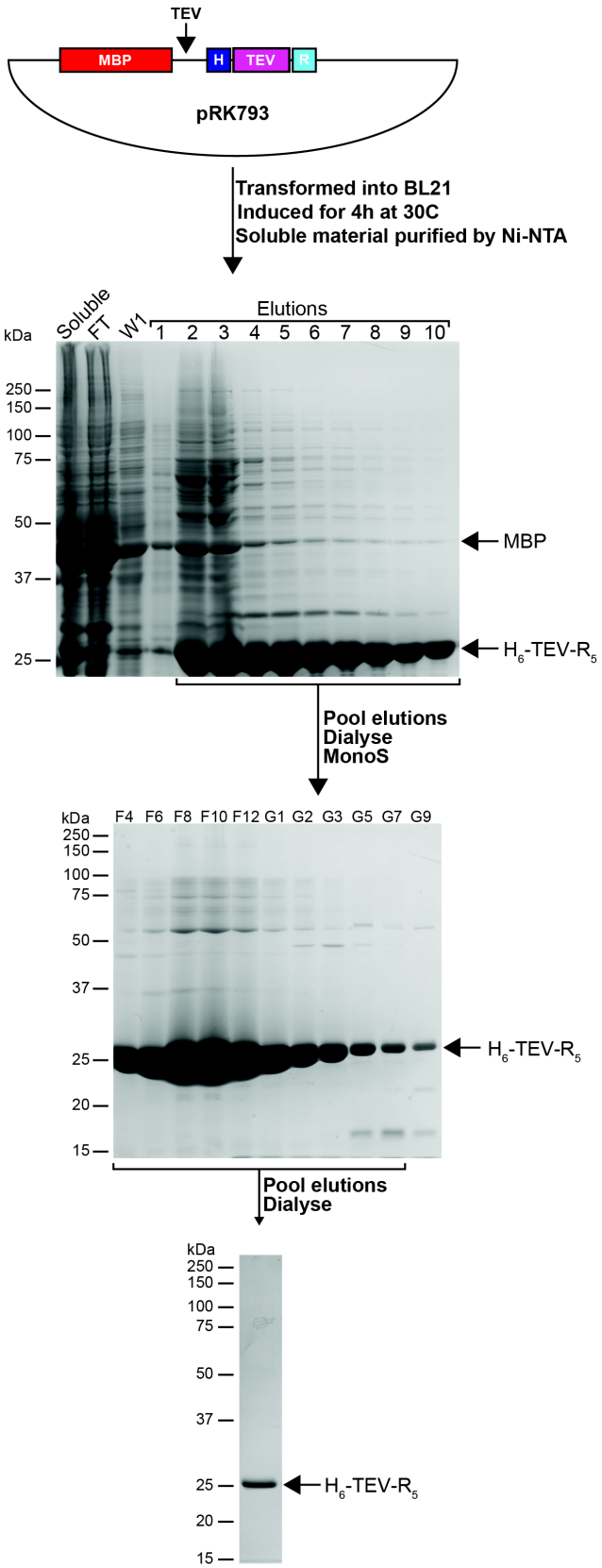



**Fig. 3.3: Expression and purification of tobacco etch virus (TEV) protease.**
TEV protease was expressed in bacteria as a self-cleaving maltose-binding protein (MBP) fusion protein with N-terminal 6xHis (H, dark blue box) and C-terminal 5xArg (R, light blue box) tags. TEV protease was first purified via the 6xHis tag and elutions were analysed by SDS-PAGE followed by Coomassie staining. Pooled elutions were then further purified on a MonoS cation exchanger and purified protein was analysed by SDS-PAGE followed by Coomassie staining.

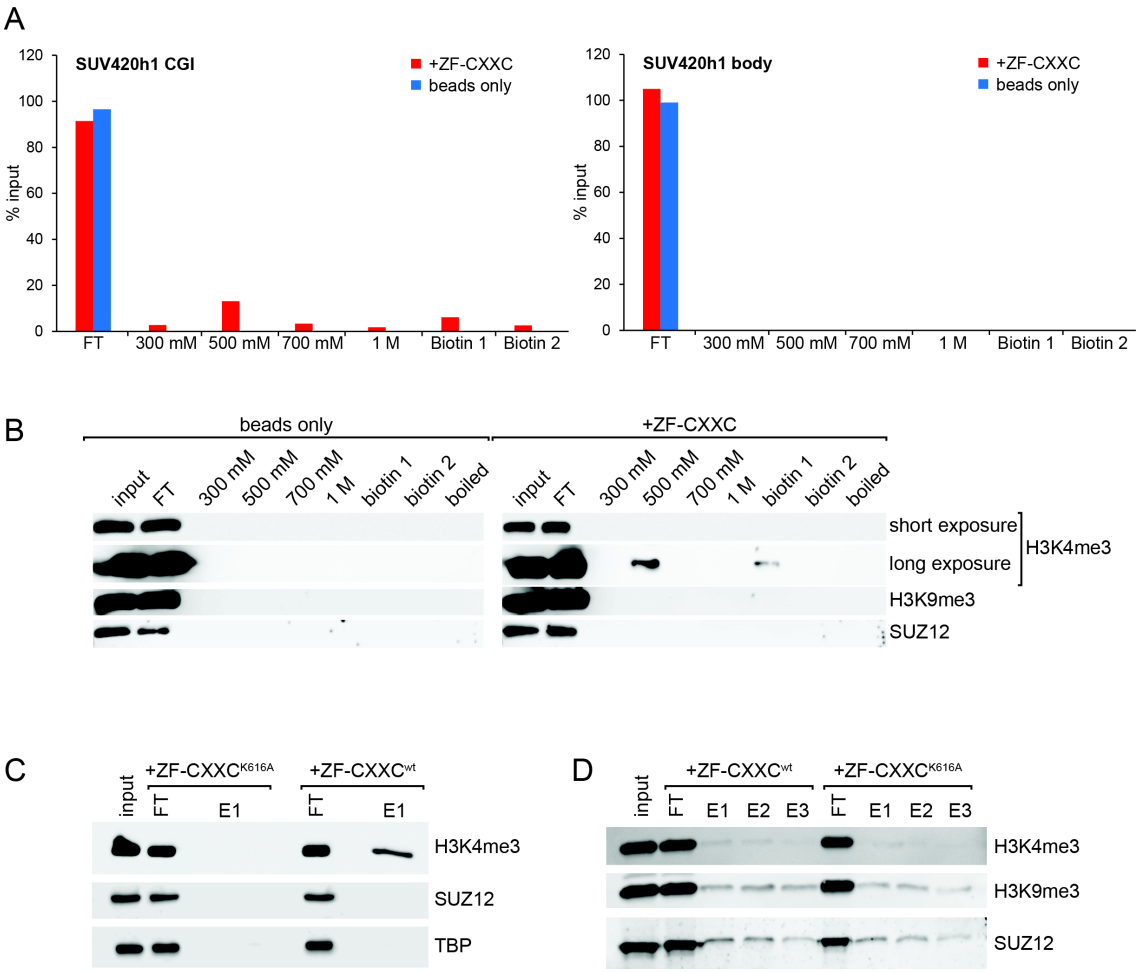



**Fig. 3.4: Optimisation of CXXC affinity purification (CAP).**
**(A)** Flowthrough (FT) and elution fractions from CAP experiments using 25 µg of ZF-CXXC protein or beads only as a negative control were analysed by qPCR at the SUV420h1 CGI region or at the gene body as a negative control.
**(B)** Input, flowthrough (FT), and elution fractions from CAP experiments using 25 µg of ZF-CXXC protein or beads only as a negative control were analysed by western blot using antibodies against H3K4me3, H3K9me3 and SUZ12.
**(C)** Input, flowthrough (FT), and a single step biotin elution fraction from CAP experiments using 250 µg of ZF-CXXC protein were analysed by western blot using antibodies against H3K4me3, SUZ12 and TBP.
**(D)** Input, flowthrough (FT), and biotin elution fractions from CAP experiments using 10 mg of ZF-CXXC protein were analysed by western blot using antibodies against H3K4me3, H3K9me3 and SUZ12.

To optimise the CAP method to increase the yield of CGI chromatin, two modifications were introduced. First, elution of bound chromatin was carried out in a single step through competition with biotin rather than using a salt gradient, to concentrate purified proteins in a single elution. Second, increasing amounts of ZF-CXXC protein were used to increase the proportion of CGI chromatin that is captured. With these changes, CAP was carried out using 250 µg of ZF-CXXC$^{wt}$ and ZF-CXXC$^{K616A}$ proteins (Fig. 3.4C). Western blot analysis shows that H3K4me3 was enriched in the ZF-CXXC$^{wt}$ elution but not in the ZF-CXXC$^{K616A}$, confirming that ZF-CXXC$^{wt}$ could specifically purify chromatin associated with CGIs. However, the levels of H3K4me3 enrichment remained comparatively low and H3K4me3 was not depleted in the flowthrough. Moreover, the CGI- and promoter-associated proteins SUZ12 and TBP could not be detected in the purification, suggesting that the proportion of CGI chromatin captured in this experiment remained relatively low. To attempt to increase the yield further, CAP was carried out using 10 mg of ZF-CXXC protein (Fig. 3.4D). However, these higher amounts of protein led to increased background binding so that H3K4me3, H3K9me3 and SUZ12 were all detectable at similar levels in purifications with both ZF-CXXC$^{wt}$ and ZF-CXXC$^{K616A}$.

The results of these CAP experiments show that although ZF-CXXC can successfully capture chromatin associated with CGIs, the amounts of chromatin that are purified are insufficient to capture a high proportion of the CGI chromatin present in the input sample or to detect known CGI-associated proteins. This is in contrast to previous work using ZF-CXXC to purify free CGI DNA from genomic DNA, in which essentially 100% of input CGI DNA could be captured (Blackledge et al., 2012). One explanation for the much lower efficiency of CGI purification in the present study could be that incorporation of CGIs into chromatin and crosslinking of CGI-bound proteins to the DNA renders the CpG dinucleotides that are recognised by ZF-CXXC inaccessible, especially given that ZF-CXXC domains require free linker DNA to bind efficiently (Zhou et al., 2012). To overcome this limitation, therefore, an alternative approach was developed in which ZF-CXXC could be expressed in ES cells *in vivo*, crosslinked *in situ* to bound CGIs, and used to purify the associated chromatin.

## 3.3 An *in vivo* approach to purify CGI chromatin

To express ZF-CXXC *in vivo*, the domain was cloned into a mammalian expression vector under the control of the CAG promoter, with C-terminal 3xFLAG and 2xStrepII tags (3F2S) (Fig. 3.5A). In addition, vectors were generated to encode a C-terminal Venus fluorescent protein, and all ZF-CXXC proteins were expressed as fusions with the ERT2 domain. ERT2 is a modified form of the estrogen receptor (ER) ligand binding domain, which sequesters proteins in the cytoplasm through interaction with chaperone proteins. Treatment with the small molecule 4-hydroxytamoxifen (4-OHT) relieves this sequestration and permits translocation of the protein into the nucleus. ERT2 was included in the ZF-CXXC constructs to allow inducible binding of the domain to CGIs, and to prevent constitutive ZF-CXXC binding from interfering with the recruitment of endogenous CGI-binding proteins.

Western blot analysis of transiently transfected ES cells showed that the constructs were expressed as expected (Fig. 3.5B), and stable cell lines were subsequently generated. Using the ZF-CXXC$^{wt}$-3F2S-Venus-ERT2 stable cell line, a 4-OHT time course was carried out to verify that treatment of the cells leads to translocation of the expressed protein from the cytoplasm to the nucleus (Fig. 3.5C). Analysis of the time course experiment by western blot showed that the protein is present exclusively in the cytoplasm in the absence of 4-OHT and translocates to the nucleus within 24 h of treatment. Further analysis of 4-OHT treatment by immunofluorescence using the Venus tag shows that translocation to the nucleus occurs within 4 h (Fig. 3.5D), suggesting that this length of treatment would be suitable for future experiments.

To test whether the expressed ZF-CXXC proteins bind appropriately to CGIs upon 4-OHT treatment, ChIP experiments were carried out using FLAG antibody in cell lines stably expressing wild type or mutant ZF-CXXC protein, in the presence or absence of the Venus tag (Fig. 3.5E). The ChIP experiments show that ZF-CXXC$^{wt}$ proteins were recruited to CGIs only after treatment with 4-OHT, but absent from non-CGI regions. By contrast, ZF-CXXC$^{K616A}$ proteins were unable to bind CGIs or non-CGI regions in the presence or absence of 4-OHT. Notably, although ZF-CXXC$^{wt}$ constructs lacking the Venus tag bound to

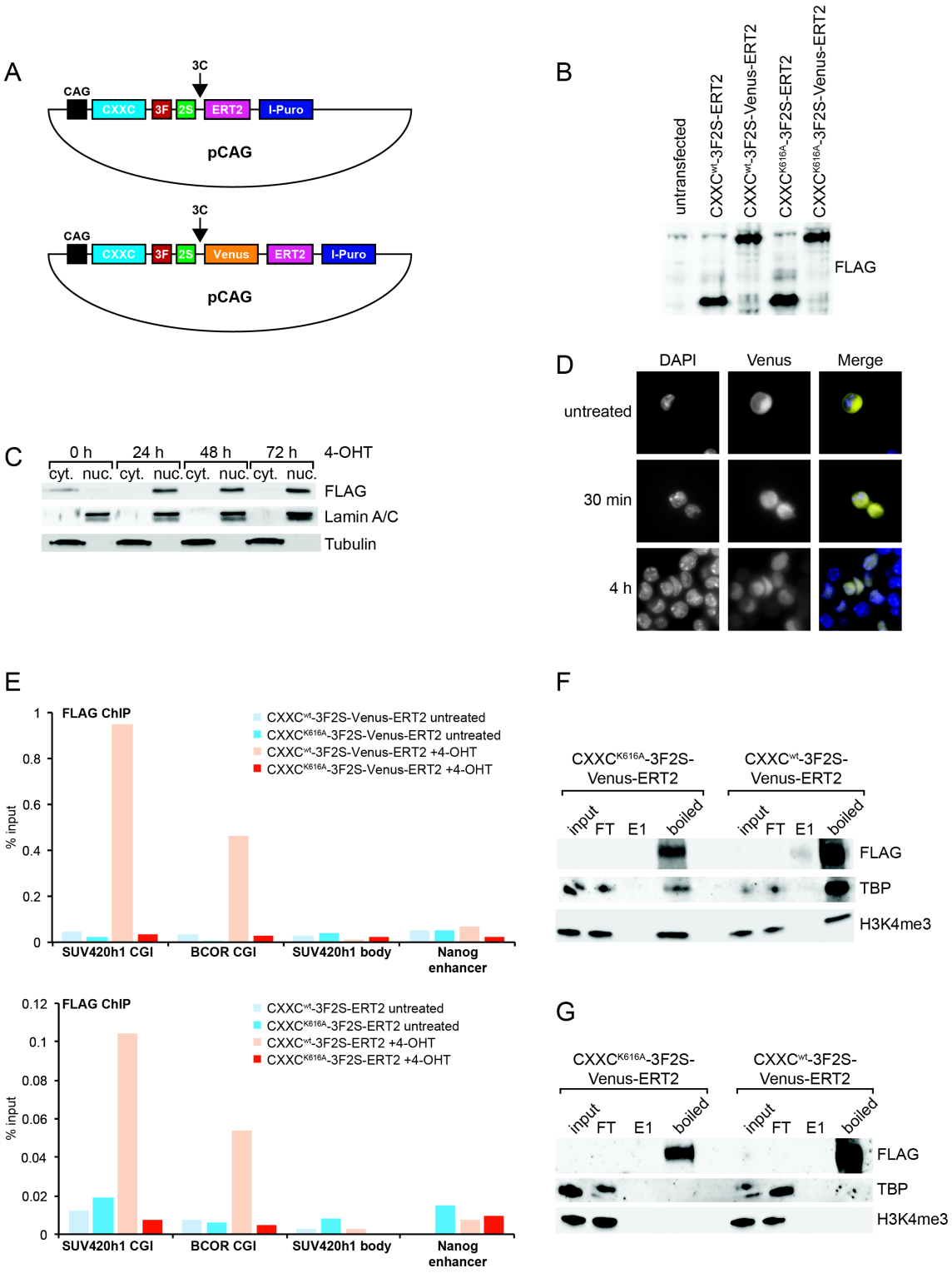



**Fig. 3.5: An *in vivo* CXXC affinity purification approach.**
**(A)** ZF-CXXC protein was expressed under the control of a CAG promoter with C-terminal 3xFLAG (3F, red box) and 2xStrepII (2S, green box) tags, followed by a 3C protease site and without (top) or with (bottom) a Venus fluorescent tag. Nuclear localization of the protein was controlled by inclusion of a ERT2 domain (magenta box), and cell line selection was made possible by an internal ribosome entry site (IRES) followed by a puromycin selectable marker (I-Puro).
**(B)** Whole cell extracts from ES cells expressing the indicated ZF-CXXC constructs were analysed by western blot using FLAG antibody.
**(C)** ES cells stably expressing ZF-CXXC$^{wt}$-3F2S-Venus-ERT2 were treated for 24 h, 48 h or 72 h with 800 nM 4-hydroxytamoxifen (4-OHT) and the cytoplasmic or nuclear

CGIs, constructs containing Venus bound more efficiently and with greater enrichment over background, although the absolute enrichment is only 1% of the input chromatin. The ZF-CXXC-3F2S-Venus-ERT2 cell lines were therefore selected to attempt to purify CGI chromatin.

To purify CGI chromatin using this *in vivo* approach, cells stably expressing wild type or mutant protein were treated for 4 h with 4-OHT, crosslinked and harvested to prepare chromatin. Large scale ChIP experiments were then carried out using antibodies against FLAG (Fig. 3.5F) or using StreptactinXT (3.5G) beads to pull down the ZF-CXXC protein. The protein and any associated chromatin was then eluted using either 3xFLAG peptide or biotin, respectively, and the beads were finally boiled in SDS-PAGE loading buffer to elute any remaining bound material. Input, flowthrough and elutions were then analysed by western blot probing for the ZF-CXXC protein via the FLAG tag, and for the CGI- and promoter-associated protein TBP and histone modification H3K4me3. Whilst the ZF-CXXC proteins were successfully pulled down, they failed to elute efficiently, and could only be eluted from the beads by boiling. For the FLAG purification, elution of the protein upon boiling also led to the elution of TBP and H3K4me3, but these were also found at high background levels in the ZF-CXXC$^{K616A}$ control (Fig. 3.5F). In the StreptactinXT purification, by contrast, neither TBP nor H3K4me3 could be detected in the purified material. These results suggest that, like the *in vitro* approach, this *in vivo* approach was unable to purify CGI chromatin at levels greater than background.

---

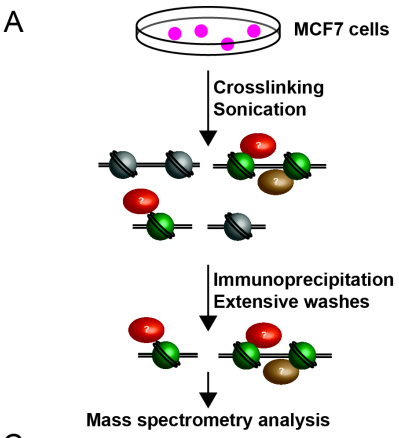
**(Fig. 3.5 cont.)** localization of the protein was tested by subcellular fractionation into cytoplasmic (cyt.) or soluble nuclear (nuc.) extracts followed by western blot using FLAG antibody or antibodies against Lamin A/C or tubulin as controls.
**(D)** ES cells stably expressing ZF-CXXC$^{wt}$-3F2S-Venus-ERT2 were treated for 30 mins or 4 h and localization of the proteins was analysed by immunofluorescence (IF).
**(E)** Chromatin immunoprecipitation (ChIP) experiments were carried out using FLAG antibodies in either untreated cells or cells treated with 4-OHT for 4 h, in cell lines stably expressing constructs with Venus (top) or without Venus (bottom). ChIP experiments were then analysed by qPCR at two CGI regions and two non-CGI regions.
**(F)**, **(G)** Cells stably expressing the indicated ZF-CXXC constructs were treated with 4-OHT for 4 h and chromatin was purified via the FLAG tags (F) or the Strep tags (G). Purified material was eluted (E1) through competition with 3xFLAG peptide (F) or biotin (G) and by boiling the beads in SDS-PAGE loading buffer. Inputs, flowthrough (FT) and elutions were then analysed by western blot using antibodies against FLAG, TBP and H3K4me3.

## 3.4 A ChIP-mass spectrometry approach to purify promoter chromatin

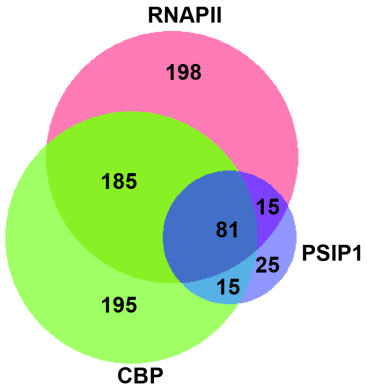
Given that the approaches used to purify CGI- and promoter-associated proteins via the ZF-CXXC domain could not successfully enrich for CGI chromatin, an alternative ChIP-mass spectrometry approach was then used to purify promoter-associated proteins. In collaboration with Dr Anca Farcaș (CRUK Cambridge Institute), and under my supervision, rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) was used to pull down chromatin crosslinked to promoter-bound targets, and associated proteins were identified by mass spectrometry (Fig. 3.6A) (Mohammed et al., 2016). In this method, cells from the MCF7 human breast cancer cell line were crosslinked and chromatin solubilised by sonication. ChIP was then carried out using antibodies against the S5 phosphorylated C-terminal domain of RNA polymerase II (RNAPII), the histone acetyltransferase CBP, the PSIP1 component of the MLL1/2 complex, and the PRC2 subunit SUZ12. After extensive washing of immunocomplexes, the purified chromatin was eluted and analysed by mass spectrometry.

RNAPII is primarily found at active promoters and CBP at both enhancers and promoters (Seila et al., 2008; Yue et al., 2014), whilst PSIP1 is expected to be found at CGI regions as part of the MLL1/2 complex (Denissov et al., 2014). SUZ12, by contrast, is found at the repressed subset of CGI promoters (Ku et al., 2008). RIME experiments were carried out using one RNAPII and one CBP antibody, and two antibodies against each of PSIP1 and SUZ12, as well as negative control experiments using immunoglobulin G (IgG). Overlap of the identified proteins including only proteins that were consistently detected in independent experiments and with multiple antibodies where possible, and that were not detected in the IgG controls, showed a large number of proteins that were co-bound in RIME experiments for RNAPII, CBP and PSIP1 (Fig. 3.6B). This suggests that these proteins could represent a set of core promoter-associated proteins.

A



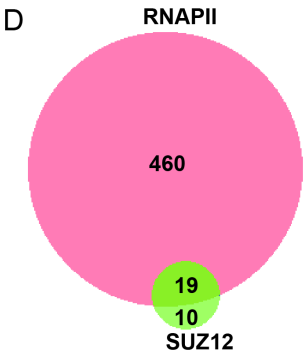

B




C

| | | | RNAPII | | CBP | | PSIP1 | | SUZ12 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Role | Protein ID | Protein | Coverage | Unique peptides | Coverage | Unique peptides | Coverage | Unique peptides | Coverage | Unique peptides |
| DNA damage response | P09874 | PARP1 | 15.68% | 10 | 4.93% | 2 | 25.35% | 19 | 1.87% | 1 |
| | P14866 | HNRNPL | 30.05% | 10 | 53.48% | 16 | 13.75% | 5 | 23.43% | 7 |
| | P27695 | APEX1 | 5.66% | 1 | 10.69% | 2 | 17.61% | 4 | 5.35% | 1 |
| | P13010 | XRCC5 | 26.23% | 11 | 47.13% | 25 | 15.98% | 7 | 14.07% | 5 |
| | P12956 | XRCC6 | 32.35% | 13 | 40.89% | 19 | 22.66% | 8 | 8.54% | 3 |
| | | | | | | | | | | |
| CHTOP complex | Q99873 | PRMT1 | 24.80% | 7 | 26.95% | 7 | 7.01% | 2 | 30.19% | 9 |
| | P84090 | ERH | 14.42% | 1 | 14.42% | 1 | 14.42% | 1 | 22.12% | 3 |
| | Q9Y3Y2 | CHTOP | 6.45% | 1 | 6.45% | 1 | 6.45% | 1 | 6.45% | 1 |
| | | | | | | | | | | |
| FACT complex | Q9Y5B9 | SPT16 | 13.75% | 11 | 2.48% | 1 | 6.69% | 5 | 1.15% | 1 |
| | Q08945 | SSRP1 | 14.81% | 10 | 0% | 0 | 8.32% | 5 | 0% | 0 |
| | | | | | | | | | | |
| Transcriptional regulation | Q13263 | TRIM28 | 44.67% | 23 | 26.23% | 14 | 28.98% | 15 | 1.08% | 1 |
| | Q15424 | SAFB | 13.88% | 4 | 7.43% | 5 | 0.98% | 1 | 7.98% | 6 |
| | Q92769 | HDAC2 | 7.38% | 2 | 7.38% | 2 | 3.48% | 1 | 0% | 0 |

D




E

| | | | RNAPII | | CBP | | PSIP1 | | SUZ12 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Role | Protein ID | Protein | Coverage | Unique peptides | Coverage | Unique peptides | Coverage | Unique peptides | Coverage | Unique peptides |
| PRC2 | Q15022 | SUZ12 | 0% | 0 | 0% | 0 | 0% | 0 | 58.46% | 34 |
| | O75530 | EED | 0% | 0 | 0% | 0 | 0% | 0 | 29.93% | 11 |
| | Q92800 | EZH1 | 0% | 0 | 0% | 0 | 0% | 0 | 1.07% | 1 |
| | Q15910 | EZH2 | 0% | 0 | 0% | 0 | 0% | 0 | 11.66% | 7 |
| | Q09028 | RBBP4 | 0% | 0 | 0% | 0 | 1.88% | 1 | 29.60% | 3 |
| | Q16576 | RBBP7 | 0% | 0 | 18.82% | 4 | 10.82% | 3 | 40.47% | 7 |
| | Q6ZN18 | AEBP2 | 0% | 0 | 0% | 0 | 0% | 0 | 28.05% | 8 |
| | Q92833 | JARID2 | 0% | 0 | 0% | 0 | 0% | 0 | 12.04% | 12 |
| | A6NHQ4 | EPOP | 0% | 0 | 0% | 0 | 0% | 0 | 5.54% | 1 |
| | O43189 | PHF1 | 0% | 0 | 0% | 0 | 0% | 0 | 28.22% | 8 |
| | Q5T6S3 | PHF19 | 0% | 0 | 0% | 0 | 0% | 0 | 13.45% | 5 |
| | Q9Y483 | MTF2 | 0% | 0 | 0% | 0 | 0% | 0 | 7.42% | 3 |
| | | | | | | | | | | |
| PRC1 | P35226 | BMI1 | 18.71% | 5 | 0% | 0 | 0% | 0 | 0% | 0 |
| | O00257 | CBX4 | 1.43% | 1 | 0% | 0 | 0% | 0 | 1.43% | 1 |
| | Q9HC52 | CBX8 | 10.54% | 3 | 0% | 0 | 0% | 0 | 0% | 0 |
| | Q8IXK0 | PHC2 | 1.05% | 1 | 0% | 0 | 0% | 0 | 0% | 0 |
| | Q8NDX5 | PHC3 | 0.92% | 1 | 0% | 0 | 0% | 0 | 0% | 0 |

**Fig. 3.6: Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) to identify regulatory element-associated proteins.**

**(A)** Schematic showing the RIME approach to purifying proteins associated with regulatory elements. Cells from the human breast cancer cell line MCF7 were crosslinked and chromatin solubilised by sonication. Chromatin was immunoprecipitated using antibody against a target protein and immunocomplexes were extensively washed before purified proteins were analysed by mass spectrometry.

**(B)** Venn diagram showing numbers of proteins identified by RIME using antibodies against RNA polymerase II (RNAPII), the MLL1/2 complex component PSIP1 and the histone acetyltransferase CBP. The results represent data from at least three independent experiments, and only proteins that were present in all purifications but

This set of potential promoter-binding proteins includes known regulators of transcription such as the FACT (facilitates chromatin transcription) complex, histone deacetylases (HDACs) and the transcriptional repressor TRIM28 (Fig. 3.6C). However, these RIME experiments also identified additional proteins that could play a role at promoters. These include proteins involved in the DNA damage response, such as the poly(ADP-ribose) polymerase PARP1, and the XRCC5 and XRCC6 proteins (also known as Ku80 and Ku70, respectively) and their interaction partners APEX1 and HNRNPL. The presence of these proteins could point to a general role for DNA damage response proteins at sites of transcription initiation, consistent with observations that sites of transcription initiation are associated with genome instability (Aguilera and Gómez-González, 2008; Helmrich et al., 2013).

A second set of proteins identified as promoter-associated by RIME is the complex associated with the arginine methyltransferase PRMT1, including CHTOP (Chromatin target of PRMT1) and ERH (Enhancer of rudimentary homologue). PRMT1 can associate with CHTOP and ERH to form a complex that monomethylates and asymmetrically dimethylates H4R3 (H4R3me1 and H4R3me2a, respectively) (van Dijk et al., 2010; Takai et al., 2014). This complex has been shown to bind to 5-hydroxymethylated cytosine (5hmC) (Takai et al., 2014), which is found enriched at promoter and enhancer elements (Ficz et al., 2011; Pastor et al., 2011; Wu and Zhang, 2011), and the complex is thought to play an important role in transcriptional activation following gene induction (van Dijk et al., 2010). Binding of this complex to promoters could therefore represent a general mode of transcriptional regulation by the PRMT1 complex.

---

**(Fig. 3.6 cont.)** not in IgG negative control experiments were included in this analysis.
**(C)** Representative results from one set of RIME experiments using antibodies against RNAPII, CBP, PSIP1 and SUZ12. Coverage and unique peptide counts are shown for proteins with roles in DNA damage, chromatin binding and transcriptional regulation.
**(D)** Venn diagram showing numbers of proteins identified by RIME using antibodies against RNAPII and SUZ12. The results represent data from at least three independent experiments, and only proteins that were present in all purifications but not in IgG negative control experiments were included in this analysis.
**(E)** Representative results from one set of RIME experiments using antibodies against RNAPII, CBP, PSIP1 and SUZ12. Coverage and unique peptide counts are shown for PRC2 complex subunits.
All RIME experiments presented in this figure were carried out by Dr Anca Farcaș (CRUK Cambridge Institute) under my supervision.

RIME experiments with SUZ12 antibodies, by contrast, identified comparatively few proteins overall, with 29 proteins consistently detected in multiple experiments using two different antibodies, compared to 479 proteins identified using the RNAPII antibody (Fig. 3.6D). Of the proteins identified in SUZ12 RIME experiments, approximately two-thirds are also detected in association with RNAPII. These proteins include XRCC6, CHTOP and ERH, indicating that these proteins may be important in the regulation of promoters that are repressed by Polycomb group (PcG) proteins as well as active promoters. The remaining proteins that were not identified in RNAPII RIME are all components of the PRC2 complex (Fig. 3.6E). Every subunit of PRC2 was identified in at least one SUZ12 RIME experiment, including the core PRC2 subunits SUZ12, EZH2 (and its paralogue EZH1), EED and RBBP4 and RBBP7, and the accessory subunits AEBP2, JARID2, PHF1, PHF19, MTF2 and EPOP. Of these, only the RBBP4 and RBBP7 subunits, which also form part of other complexes such as the nucleosome remodelling factor (NURF) complex (Barak et al., 2003), are also detected in RNAPII RIME experiments, suggesting that PRC2 is found solely at PcG-repressed loci in MCF7 cells and is not distributed to promoters more generally.

Consistent with previous work using ChIP-mass spectrometry of PRC2 proteins (Alekseyenko et al., 2014), very few proteins were identified by SUZ12 RIME in addition to the PRC2 complex. Indeed, even the PRC1 complex, which generally co-occupies PcG-repressed loci with PRC2, is not detected in SUZ12 RIME, although the complex is present in MCF7 cells, as indicated by identification of PRC1 proteins in RNAPII RIME (Fig. 3.6E). This relative absence of other proteins from SUZ12 RIME could suggest that PRC2-bound sites are not co-occupied by PRC1 or other proteins in this cell type. However, an alternative possibility is that PRC2 forms a biochemically distinct complex that does not enter sufficiently close proximity with other proteins to crosslink to non-PRC2 proteins efficiently. This would mean that even proteins that are present at the same sites as PRC2 are not found in SUZ12 RIME experiments.

Overall, these results suggest that RIME could provide a viable alternative approach to identify proteins that function at regulatory elements. Indeed, these preliminary experiments suggest that DNA damage response proteins and

PRMT1 and its associated complex may play a hitherto largely unexplored role in regulating promoter function.

## 3.5 A candidate approach to study MLL3 and MLL4 function at enhancers

To understand how the H3K4 methyltransferase complexes MLL3 and MLL4 (MLL3/4) contribute to enhancer function, a similar proteomics approach was adopted in which MLL3 and MLL4 were purified from ES cells. The purified complexes could then be analysed to identify the interaction partners of MLL3/4, and the complexes could further be used in enzymatic assays *in vitro* to understand how their catalytic activities are regulated.

To purify MLL3/4 complexes from ES cells, cell lines were generated using CRISPR/Cas9 genome editing in which either MLL3 or MLL4 was tagged with 3xFLAG and 2xStrepII tags (3F2S) at the N-terminus of either protein (Fig. 3.7A, B). PCR-based screening identified possible positive clones for each cell line (Fig. 3.7C, D), and these were confirmed by further PCR analysis (Fig. 3.7E), immunoprecipitation (IP) followed by western blot for the known interaction partner UTX (Fig. 3.7F), and by sequencing. These analyses identified cell lines in which one allele of either MLL3 or MLL4 was successfully tagged. Due to the high GC content of the endogenous MLL3 and MLL4 genes, sequencing of the untagged alleles proved impossible, and the high molecular weight of the MLL3 (540 kDa) and MLL4 (600 kDa) proteins meant that tagged and untagged proteins could not be distinguished by western blot. Therefore, it cannot be stated definitively whether the remaining untagged alleles in these cell lines are either wild type or, more likely, mutated and rendered non-functional. Neither case, however, should interfere with the function of the tagged proteins.

Once the 3F2S-tagged cell lines were established, 3F2S-MLL4 was purified via the FLAG affinity purification (Fig. 3.8). Analysis of the purified complex by SDS-PAGE followed by silver staining revealed the presence of proteins that

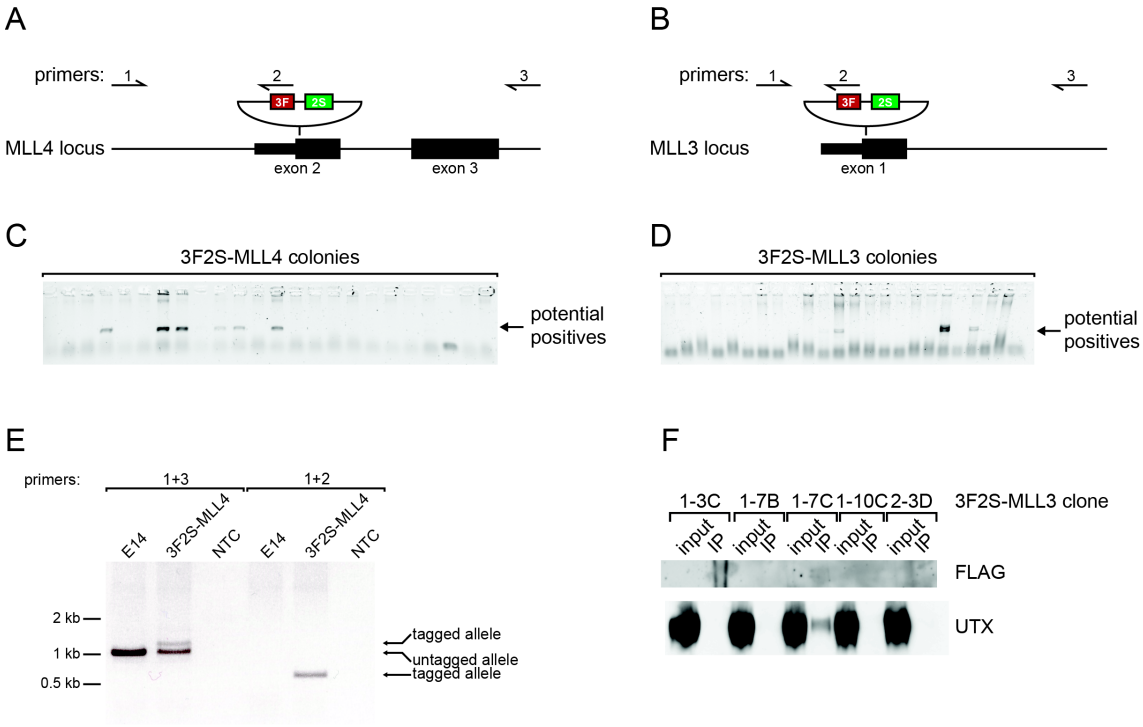
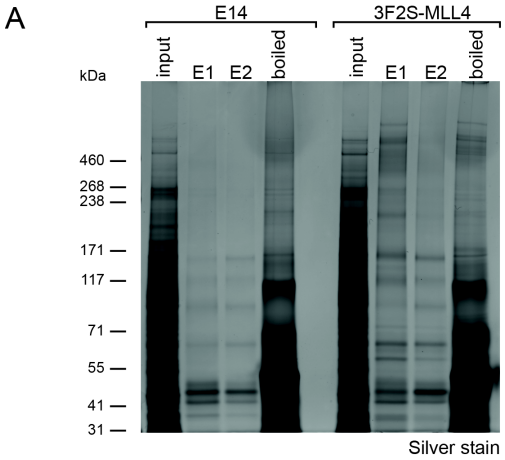


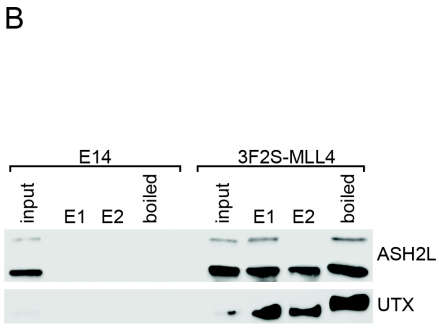

**Fig. 3.7: A CRISPR/Cas9 strategy to endogenously tag MLL3 and MLL4.**
**(A)**, **(B)** Homology directed repair (HDR) templates carrying 3xFLAG (3F, red box) and 2xStrepII (2S, green box) tags were used to insert N-terminal affinity tags at the start of the coding regions of the MLL4 (A) and MLL3 (B) loci using CRISPR/Cas9-mediated genome editing. Introns are represented by lines, 5' untranslated region (UTR) by small boxes and coding region by larger boxes. Primers used for PCR-based screening assays are indicated above the genomic loci.
**(C)**, **(D)** ES cell colonies were screened for insertion of the tag cassette by PCR using primers 1 and 2 indicated in (A) and (B), and potentially positive clones are indicated.
**(E)** Successful tagging of a heterozygous 3F2S-MLL4 cell line was confirmed by PCR using the indicated primers, compared to wild type parental E14 ES cells and a non-template control (NTC).
**(F)** Potential 3F2S-MLL3 clones were screened by immunoprecipitation (IP) of whole cell extracts using FLAG M2 resin, with IPs analysed by western blot using antibodies against FLAG and interaction partner UTX.

---

are not seen in a negative control purification from wild type E14 ES cells, suggesting that a complex of proteins was successfully purified (Fig. 3.8A). The presence of the MLL4 complex in the purified material was confirmed by western blot, which indicated the presence of known MLL4 complex components UTX and ASH2L (Fig. 3.8B).

To determine whether the purified complex contained all of the expected MLL4 complex members and whether any additional proteins co-purified with MLL4, the purified material was analysed by mass spectrometry (Fig. 3.8C, D). This analysis showed that purified 3F2S-MLL4 contains all of the known components




C

| Complex | Name | Score | Peptides | Unique peptides |
|---|---|---|---|---|
| MLL4 | MLL4 | 323 | 194 | 176 |
| | UTX | 323 | 73 | 65 |
| | PTIP | 323 | 63 | 63 |
| | ASH2L | 323 | 35 | 2 |
| | ASH2L | 10 | 34 | 1 |
| | RBBP5 | 323 | 30 | 30 |
| | WDR5 | 323 | 21 | 21 |
| | MLL4 | 6 | 18 | 1 |
| | NCOA6 | 135 | 15 | 8 |
| | NCOA6 | 23 | 8 | 1 |
| | PA1 | 82 | 6 | 6 |
| | DPY30 | 143 | 6 | 6 |
| | | | | |
| MLL4-related | UTY | 323 | 37 | 2 |
| | UTY | 7 | 34 | 0 |
| | MLL3 | 14 | 4 | 2 |

D

| Complex | Name | Score | Peptides | Unique peptides |
|---|---|---|---|---|
| SETD1A | SETD1A | 120 | 15 | 15 |
| | HCFC1 | 68 | 10 | 10 |
| | WDR82 | 31 | 5 | 5 |
| | CFP1 | 12 | 2 | 2 |
| | | | | |
| | OGT | 39 | 6 | 6 |
| | | | | |
| TIP60/p400 | RUVBL2 | 72 | 12 | 12 |
| | RUVBL1 | 77 | 10 | 10 |
| | | | | |
| Cohesin/ | SMC3 | 106 | 18 | 18 |
| condensin | SMC2 | 95 | 14 | 14 |
| | SMC1A | 18 | 3 | 3 |
| | PDS5A | 16 | 2 | 2 |
| | SMC4 | 11 | 2 | 2 |
| | | | | |
| BAF | SMARCC1 (BAF155) | 85 | 10 | 10 |
| | BRG1 (BAF190A) | 13 | 2 | 2 |
| | SMARCD1 (BAF60A) | 12 | 2 | 2 |
| | SMARCE1 (BAF57) | 6 | 1 | 1 |
| | ARID1A (BAF250A) | 18 | 3 | 3 |
| | ACTL6A (BAF53A) | 31 | 5 | 5 |
| | | | | |
| NuRD | RBBP4 | 35 | 5 | 5 |
| | MTA3 | 33 | 5 | 5 |
| | CHD4 | 26 | 4 | 4 |
| | HDAC2 | 30 | 3 | 3 |
| | SALL4 | 12 | 1 | 1 |
| | MBD3 | 6 | 1 | 1 |
| | CTBP2 | 21 | 3 | 3 |
| | GATAD2A | 11 | 2 | 2 |
| | GATAD2B | 11 | 2 | 2 |
| | | | | |
| | BCL6 | 8 | 1 | 1 |
| | | | | |
| CoREST | RBBP4 | 35 | 5 | 5 |
| | HDAC2 | 30 | 3 | 3 |
| | LSD1 | 11 | 2 | 2 |
| | RCOR2 | 6 | 1 | 1 |

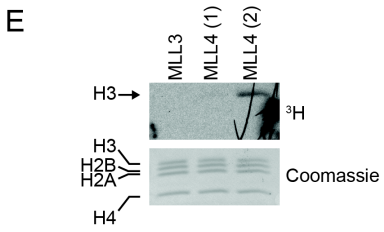



**Fig. 3.8: Purification of endogenous MLL4 complex from ES cells.**
**(A)** 3F2S-MLL4 was purified via the FLAG tags, with purification from wild type E14 ES cells as a negative control. Purified material was eluted by competition with 3xFLAG peptide (E1, E2) and by boiling in SDS-PAGE loading buffer. Purified material was analysed by SDS-PAGE followed by silver staining.
**(B)** As in (A), except purified material was analysed by western blot using antibodies against MLL4 complex components ASH2L and UTX.
**(C), (D)** Proteins purified from 3F2S-MLL4 cells was analysed by mass spectrometry, showing the presence of known MLL4 complex subunits (C) and proteins that have not previously been identified as MLL4 interaction partners (D).
**(E)** Purified MLL3 complex and two independent MLL4 complex purifications were used as enzymes in histone methyltransferase (HMT) assays with recombinant nucleosome arrays and $^3$H-labelled S-adenosyl methionine (SAM). Reactions were analysed by SDS-PAGE, transferred onto a PVDF membrane, stained with Coomassie to reveal equal loading of nucleosomes between reactions, and exposed to film to generate an autoradiograph ($^3$H).

of the MLL4 complex (Fig. 3.8C), namely the WDR5, RBBP5, ASH2L and DPY30 subunits that are common to all mammalian H3K4 methyltransferases, and the UTX, PTIP, PA1 and NCOA6 subunits that are distinct to the MLL3/4 complexes. In addition to these proteins, UTY, a paralogue of UTX that has not previously been identified as part of the MLL4 complex, was detected, and small amounts of the MLL3 protein. UTX is an H3K27 demethylase enzyme encoded on the X chromosome, whilst UTY is encoded on the Y chromosome in males and lacks catalytic activity *in vitro* and *in vivo* (Hong et al., 2007; Lan et al., 2007; Shpargel et al., 2012). This could suggest that UTX and UTY have functional roles as part of the MLL4 complex that are independent of their catalytic activity, consistent with results suggesting a functional redundancy between UTX and UTY in development and in tumour suppressor function that does not depend on H3K27 demethylase activity (Gozdecka et al., 2018; Shpargel et al., 2012)

Further analysis of the mass spectrometry data shows that 3F2S-MLL4 co-purified with several other known chromatin-associated proteins (Fig. 3.8D). These include components of the SETD1A H3K4 methyltransferase complex, including CFP1, WDR82, HCFC1 and the enzymatic subunit SETD1A (Lee and Skalnik, 2005). Additional co-purified proteins include components of the TIP60/p400 histone acetyltransferase complex, the cohesin and condensin complexes involved in chromatin looping, the BAF chromatin remodelling complex, the NuRD nucleosome remodelling and histone deacetylase complex, and the CoREST histone deacetylase complex. The most probable explanation for the detection of these proteins in the 3F2S-MLL4 purification is that many of these complexes play a role at the same genomic locations as the MLL4 complex, with TIP60/p400, cohesin, BAF, NuRD and CoREST all known to bind to enhancers (Kagey et al., 2010; Laurent et al., 2015; Morris et al., 2014; Phillips-Cremins et al., 2013; Wang et al., 2009). These proteins may therefore come into contact with MLL4 without forming a stable complex, consistent with the lower number of peptides identified for these proteins compared to the core MLL4 complex subunits.

The presence of components of the SETD1A complex in the 3F2S-MLL4 purification is more surprising than the presence of known enhancer-binding

proteins, given that the SETD1A complex is thought to bind primarily at active promoters rather than enhancers, through its CFP1 and WDR82 subunits (Brown et al., 2017; Clouaire et al., 2012; Lee and Skalnik, 2008; Thomson et al., 2010). The presence of SETD1A subunits in association with MLL4 could suggest that SETD1A has an additional role at distal regulatory elements. This possibility would be consistent with the reduced levels of DNA methylation found at enhancers (Lister et al., 2009; Schmidl et al., 2009; Stadler et al., 2011; Thurman et al., 2012), which could lead to the recruitment of the SETD1A complex through the ZF-CXXC domain of the subunit CFP1 (Voo et al., 2000). An alternative, but not mutually exclusive, possibility is that MLL4 plays a greater role at promoter elements than previous reports would suggest (Dorighi et al., 2017; Hu et al., 2013c). Consistent with this hypothesis, examination of MLL4 ChIP-seq data shows that approximately one-quarter of MLL4 binding sites are at gene promoters in the human colon cancer cell line HCT116 (Hu et al., 2013c). Moreover, the observation that MLL3/4 shows little occupancy at promoter regions in ES cells was carried out in 2i+LIF culture conditions (Dorighi et al., 2017), which can have drastically altered chromatin profiles compared to ES cells cultured only in the presence of LIF (Leitch et al., 2013). One possibility is therefore that MLL3/4 could bind to a broader range of promoter elements in the LIF-only conditions used in this study. To test whether MLL4 binds to promoters in addition to enhancers, ChIP experiments were carried out using antibodies against MLL4 in wild type ES cells and against the FLAG and StrepII tags in the 3F2S-MLL4 cell line. However, MLL4 signal could not be detected by qPCR even at known MLL4 binding sites, suggesting that the antibodies used were not able to immunoprecipitate MLL4 in ChIP conditions.

Finally, to understand how MLL3/MLL4 enzymatic activity is regulated, purified complexes were used in histone methyltransferase (HMT) assays with recombinant nucleosome arrays and $^{3}$H-labelled S-adenosyl methionine (SAM) (Fig. 3.8E). Were MLL3/4 complexes to have sufficient activity towards nucleosomes, MLL3/4 enzymatic activity could then be tested in the presence of specific histone modifications or of other chromatin proteins, to determine whether these influence MLL3/4 activity. However, although activity could be detected with MLL4 purifications, the activity was relatively weak, and no

activity was detectable with reactions using the MLL3 complex. This weak activity meant that it was not possible to test the activity of the complexes in other settings, as changes in activity would be difficult to detect and potentially unreliable.

Overall, these results show that tagging of endogenous alleles of MLL3/4 proteins facilitates the purification of their associated complexes and of other additional proteins. This could potentially indicate that H3K4 methyltransferases have a broader role at regulatory regions in ES cells than has previously been appreciated, with greater overlap between the SETD1A and MLL3/4 complexes at target sites.

## 3.6 Summary and discussion

Promoter and enhancer regions are important for the regulation of gene expression in mammalian cells. Although it is clear that the chromatin states at these regions play a key role in mediating their function, how these states are established and interpreted remains incompletely understood. Work in this chapter attempted to further our understanding of the mechanisms by which regulatory regions function by using proteomics approaches to reveal the proteins that bind to promoters and enhancers.

### *3.6.1 Purification of CGIs and promoters*

Previous work has successfully isolated the chromatin associated with large genomic regions such as telomeres, pericentromeric heterochromatin and rDNA, and identified proteins associated with these regions by mass spectrometry using the technique called PICh (proteomics of isolated chromatin) (Déjardin and Kingston, 2009; Ide and Dejardin, 2015; Saksouk et al., 2014). CGIs and telomeres constitute similar proportions of the genome, with both comprising at most 1% of the mouse genome (Antequera and Bird, 1993; Déjardin and Kingston, 2009). Therefore, it was plausible that an approach that extended the capacity of the ZF-CXXC domain to purify CGI DNA (Blackledge et al., 2012; Illingworth et al., 2008) to purify CGI chromatin could

yield sufficient material to permit identification of CGI-associated proteins. However, differences in the methodologies applied in these studies meant that a CAP-based approach was less successful than purifications of other genomic compartments.

The major difference between CAP and the PICh approach applied previously is the identity of the affinity reagent used to bind to the chromatin region of interest. Whilst CAP makes use of the ZF-CXXC protein domain that binds to unmethylated CpG, PICh uses locked nucleic acids (LNAs) that hybridize to repetitive DNA elements. One consequence of this is that whilst CpG dinucleotides might be occluded by binding of other proteins, LNAs are likely to be able to hybridize to at least partial complementary sequences at a given target site, and each target site will have multiple regions that can be bound due to their highly repetitive nature. Therefore, the efficiency of binding between the affinity reagent and the target site is likely to be greater in PICh than in CAP. This could explain why only a small proportion of CGI chromatin was successfully enriched with CAP.

An additional limitation of the CAP approach compared to PICh is that the ZF-CXXC domain is also likely to have lower affinity for its target site than the LNAs used in PICh. ZF-CXXC has an affinity for unmethylated CpG of approximately 0.6 μM (Blackledge et al., 2010), whereas LNAs can have picomolar affinity in hybridization reactions *in vitro* (Möhrle et al., 2005). This again suggests that LNAs would bind target sites with greater efficiency than ZF-CXXC. It would be possible to overcome this binding inefficiency by increasing the amount of ZF-CXXC affinity reagent used in the CAP experiment. However, because the low affinity of ZF-CXXC for CpG prohibits stringent washes, using large quantities of ZF-CXXC leads to greatly increased background binding of non-CGI chromatin. Together, these issues meant that a CAP approach to purify CGI chromatin was not viable. Similarly, although the *in vivo* approach of expressing ZF-CXXC in cells allowed the domain to access its target sites, the approach was limited by low pulldown efficiency, with only around 1% of any given CGI being purified in ChIP experiments.

By contrast, RIME experiments using antibodies against endogenous proteins that bind to promoter elements with high affinity, combined with a stringent bioinformatics analysis to rule out false positives, appears to provide a more profitable avenue for identification of novel regulatory element binding proteins. Indeed, the experiments described here suggest that DNA damage response proteins and the PRMT1-CHTOP complex may play a role at promoters in MCF7 cells. Further work, including ChIP experiments for the identified proteins, will be required to confirm that these complexes are truly present at promoters, but these proteins have potential for future mechanistic study of promoter function.

However, RIME experiments also have their limitations, as exemplified by the SUZ12 RIME results presented here. Whilst these successfully purified the known PRC2 complex, they failed to identify any novel factors that might be unique to PcG-repressed promoters and that might be required to maintain the repressed state. One possible explanation for this is that promoters bound and repressed by PcG proteins have a more compact chromatin structure that prevents binding of other proteins, and that this may be one mechanism by which gene repression is mediated. However, the finding that even PRC1 proteins were not detected in SUZ12 RIME, despite the observation that PRC1 and PRC2 binding overlaps in MCF7 cells (ENCODE Project Consortium, 2012), suggests that a technical explanation is more likely. This could be that some nucleoplasmic, rather than chromatin-bound, PRC2 is present in the chromatin preparations used for RIME experiments, and that binding of this PRC2 fraction to the antibody competes with the chromatin-associated PRC2. Alternatively, the chromatin-bound PRC2 complex could be relatively self-contained and fail to crosslink efficiently to other proteins at the same genomic locations, preventing identification of proteins that associate with PcG-repressed promoters. This suggests that whilst RIME is a valuable protoeomics tool, it may not be equally useful for all possible target sites.

### *3.6.2 Purification of MLL3/4 complexes*

Endogenous tagging of the MLL3 and MLL4 loci permitted successful purification of the MLL4 complex from ES cells and identification of associated proteins by mass spectrometry. However, identified proteins that are not part of the known MLL4 complex were only purified at comparatively low levels and comprised mostly known regulatory element-binding proteins, suggesting that these proteins interact only transiently with MLL4 rather than representing novel MLL4 complex components. The technical difficulties of ChIP experiments with these proteins meant that it was not possible to use RIME to identify novel enhancer-bound proteins at MLL3/4 target sites. Moreover, *in vitro* approaches to studying regulation of MLL3/4 enzymatic activity were not possible because of the poor activity displayed by the complexes purified from ES cells. The reason for this is most likely that only small amounts of complex could be purified due to the low abundance of endogenous MLL3/4 proteins *in vivo*. One way of overcoming this problem could be to express the enzymatic subunit exogenously at higher levels to purify a larger amount of enzyme. However, because the enzymatic subunit would be overexpressed compared to the other components of the complex, purification would yield a complex that may not exhibit the appropriate stoichiometry of the endogenous MLL3/4 complex, and may therefore have different enzymatic properties (Zhang et al., 2015). An alternative possibility would be to express and purify the MLL3/4 complexes recombinantly to yield large amounts of protein. However, such an approach is again limited by the need to express multiple complex components at the correct stoichiometry, and also by the large size of the enzymatic subunits, which renders cloning the full length *Mll3* and *Mll4* cDNAs into recombinant expression vectors highly problematic.

However, this suggests an alternative approach to study regulatory element function using purified enzymes. In addition to MLL3/4 complexes, enhancers and promoters are bound by other chromatin-modifying proteins, including the HATs CBP and p300. Although these proteins associate with many other proteins *in vivo*, they do not require complex formation for activity, and they have been successfully purified from insect cells using the baculovirus

expression vector system to yield large amounts of pure protein for enzymatic experiments. Therefore, an alternative strategy to study regulatory element function would be to make use of purified CBP/p300 to understand how these proteins acetylate chromatin to regulate gene expression.

# 4. The CBP TAZ2 domain directs H3K27 acetylation in chromatin

CBP and its paralogue p300 are histone acetyltransferase (HAT) enzymes that are conserved in metazoans, plants and some unicellular eukaryotes (Bordoli et al., 2001; Sebé-Pedrós et al., 2011; Yuan and Giordano, 2002). Homozygous knockouts of either CBP or p300 are embryonic lethal in mice, with lethality recapitulated by mutants lacking enzymatic activity (Oike et al., 1999; Tanaka et al., 1997; Yao et al., 1998). Germline mutation of one allele of CBP or p300 in humans leads to a rare developmental disorder called Rubinstein-Taybi Syndrome (RSTS) (Hennekam, 2006; Petrij et al., 1995), and CBP/p300 is emerging as a target for cancer treatment (Attar and Kurdistani, 2017; Jin et al., 2017). Together, these observations suggest that the HAT activity of CBP/p300 plays a key role in development and disease.

Histone acetylation is thought to function through regulation of gene expression, with acetylation contributing to activation of transcription. Initial work identifying CBP/p300 as HATs showed that these enzymes can acetylate lysine residues within all four core histone proteins (Bannister and Kouzarides, 1996; Ogryzko et al., 1996), and further studies have identified a large number of sites that can be acetylated by CBP/p300 *in vitro*, both in histone N-terminal tails and in the globular domain of histone H3 (An et al., 2002; Di Cerbo et al., 2014; Ogryzko et al., 1996; Tropberger et al., 2013). However, *in vivo* studies have shown that loss of CBP/p300 activity, either through genetic ablation or through drug-mediated enzymatic inhibition, leads to loss of specific histone acetylation marks (Jin et al., 2011; Pasini et al., 2010; Tie et al., 2009; Weinert et al., 2018), with mass spectrometry analysis indicating that inactivation of CBP/p300 leads to almost total depletion of H3K27ac, H3K18ac, and acetylation of the H2B N-terminal tail in mammalian cells, while acetylation of other histone residues remains largely unchanged (Weinert et al., 2018). H3K27ac is considered to be a key histone modification in the regulation of gene expression, specifically marking active gene regulatory elements. This modification frequently overlaps with CBP/p300 binding genome-wide, suggesting that H3K27 may be a key

target for CBP/p300 HAT activity (Creyghton et al., 2010; Heintzman et al., 2009; Rada-Iglesias et al., 2011).

There are several possible explanations for the apparent contradiction between *in vivo* and *in vitro* analysis of CBP/p300 substrate specificity. One possibility is that multiple HAT enzymes exist in the cell and that whilst one or more of these are able to compensate for loss of CBP/p300 activity towards many histone lysine residues, they are unable to do so for H3K27ac and certain other residues. However, kinetic studies using CBP/p300 inhibitors show that loss of CBP/p300 activity leads to rapid loss of H3K27ac whilst acetylation of residues such as H3K9 is maintained, even before compensatory mechanisms are likely to have taken effect (Weinert et al., 2018). This implies that the observed substrate specificity is unlikely to be attributable solely to compensatory acetylation by other HATs, but is rather an inherent property of the CBP/p300 enzymes.

An alternative explanation for the discrepancy between the promiscuity of histone acetylation by CBP/p300 *in vitro* and its specificity *in vivo* is the difference between substrates in these experiments. Many *in vitro* HAT experiments demonstrating acetylation of histone residues that are apparently not key targets of CBP/p300 *in vivo* were carried out using non-physiological substrates for histone acetylation, such as free core histones. By contrast, histone proteins that are acetylated by CBP/p300 *in vivo* are assembled into histone octamers and incorporated into DNA to form polynucleosomal arrays. This physiological substrate could render lysine residues that are accessible to CBP/p300 in free histones inaccessible in chromatin *in vivo*, with residues potentially occluded by the presence of other histone proteins or DNA. Indeed, several reports indicate that CBP/p300 acetylates free histone proteins more effectively and more promiscuously than nucleosomal histones, and that nucleosome assembly also leads to alterations in substrate specificity (An and Roeder, 2003; Kraus et al., 1999). However, how mechanistically CBP/p300 might interact with nucleosome substrates to generate the specificity observed *in vivo* remains unclear.

Therefore, to examine how CBP/p300 substrate specificity is achieved, an *in vitro* biochemical approach was adopted using purified recombinant CBP proteins together with either reconstituted histone octamer or nucleosome array substrates to map which domains of CBP contribute to specificity towards H3K27ac. Understanding how mechanistically CBP/p300 specifically targets H3K27 for acetylation could lead to a deeper understanding of how CBP/p300 functions to direct H3K27ac *in vivo* to regulate the fundamental process of gene expression.

## 4.1 Expression and purification of full length CBP protein

To understand how CBP regulates acetylation of histones in chromatin, the protein was first expressed and purified for use in enzymatic assays. CBP is a large protein of approximately 265 kDa with a complex domain structure (Fig. 4.4A). Therefore, to express intact full length mouse CBP at high levels the protein was expressed in Sf9 insect cells using the baculovirus expression vector system (BEVS) with N-terminal 6xHis and FLAG tags (Fig. 4.1A). CBP was then purified from Sf9 cell lysate through binding of the FLAG tag to FLAG M2 resin, and eluted under native conditions by competition with FLAG peptide (Fig. 4.1B) to yield a relatively pure, concentrated and undegraded sample of CBP, as assessed by SDS-PAGE and Coomassie staining (Fig. 4.1C).

## 4.2 Reconstitution of nucleosome array

To test the activity of purified CBP towards chromatin, nucleosome arrays were reconstituted *in vitro*. To generate nucleosome arrays, histone octamers were incorporated into a DNA template comprising 12 repeats of a 177 bp sequence containing the 147 bp 601 strong nucleosome positioning sequence (Lowary and Widom, 1998) flanked by 15 bp linker DNA overhangs (Fig. 4.2A). Incorporation of histone octamer was confirmed by the observation of a gel shift upon reconstitution of the arrays with increasing amounts of histone octamer (Fig. 4.2B). The quality of the array was analysed using a restriction digest-based assay, adapted from (Yuan et al., 2012), that makes use of a *ScaI* restriction enzyme site located in the linker DNA between nucleosome positioning sequences (Fig. 4.3C). Digestion of arrays leads to release of either

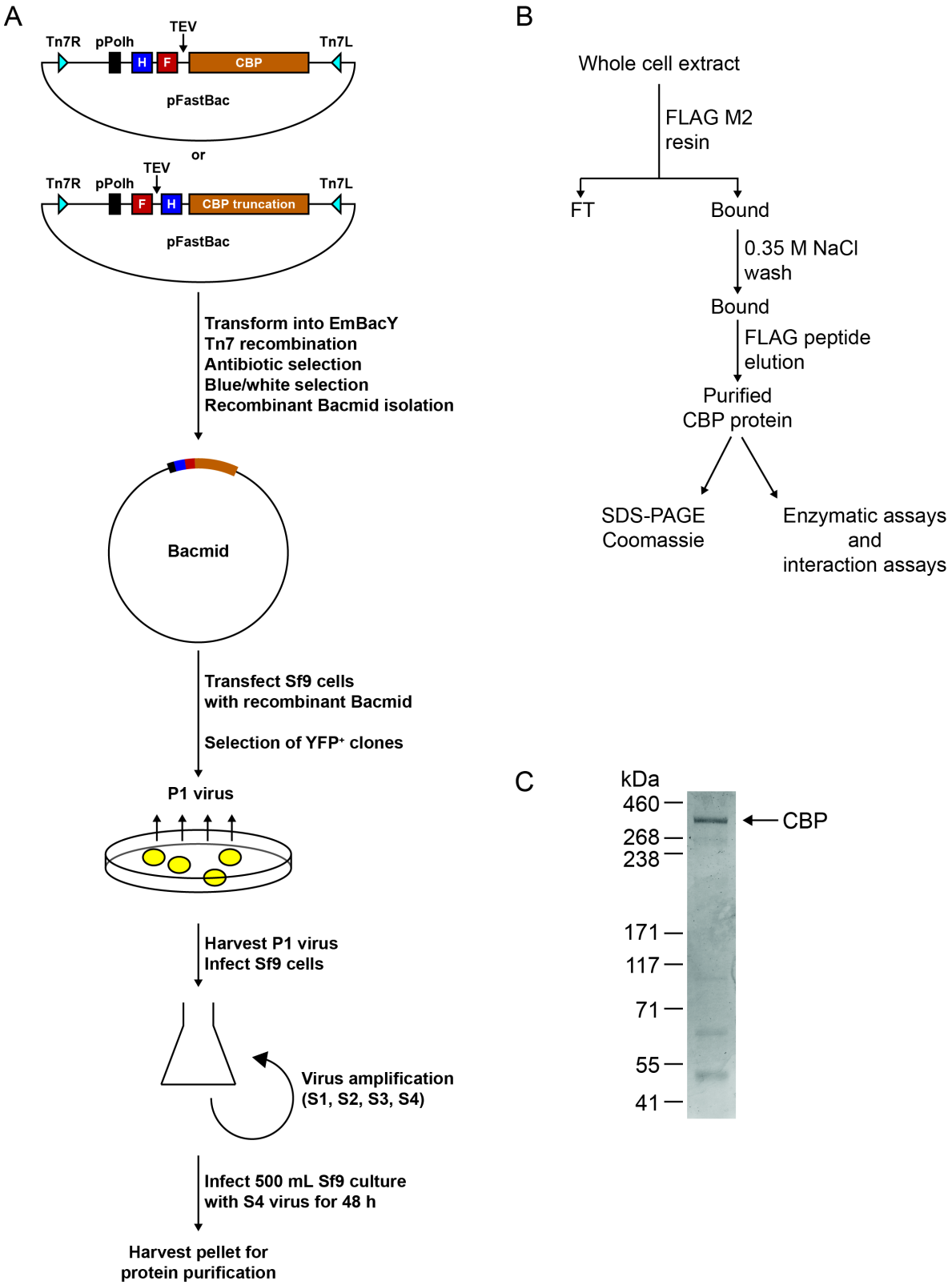



**Fig 4.1: Expression and purification of CBP protein from Sf9 cells.**
**(A)** CBP or CBP truncation proteins were cloned with 6xHis (H, blue box) and FLAG (F, red box) tags with TEV cleavage site (TEV) in the pFastBac vector, under the control of the Polyhedrin promoter (pPolh, black box) between Tn7 attachment sites (light blue triangles). The donor plasmids were used to transform EmBacY competent bacteria, recombinants were selected and bacmid DNA was isolated. Recombinant bacmid was then transfected into Sf9 cells and P1 virus was harvested and used to amplify the viral stock for protein expression.

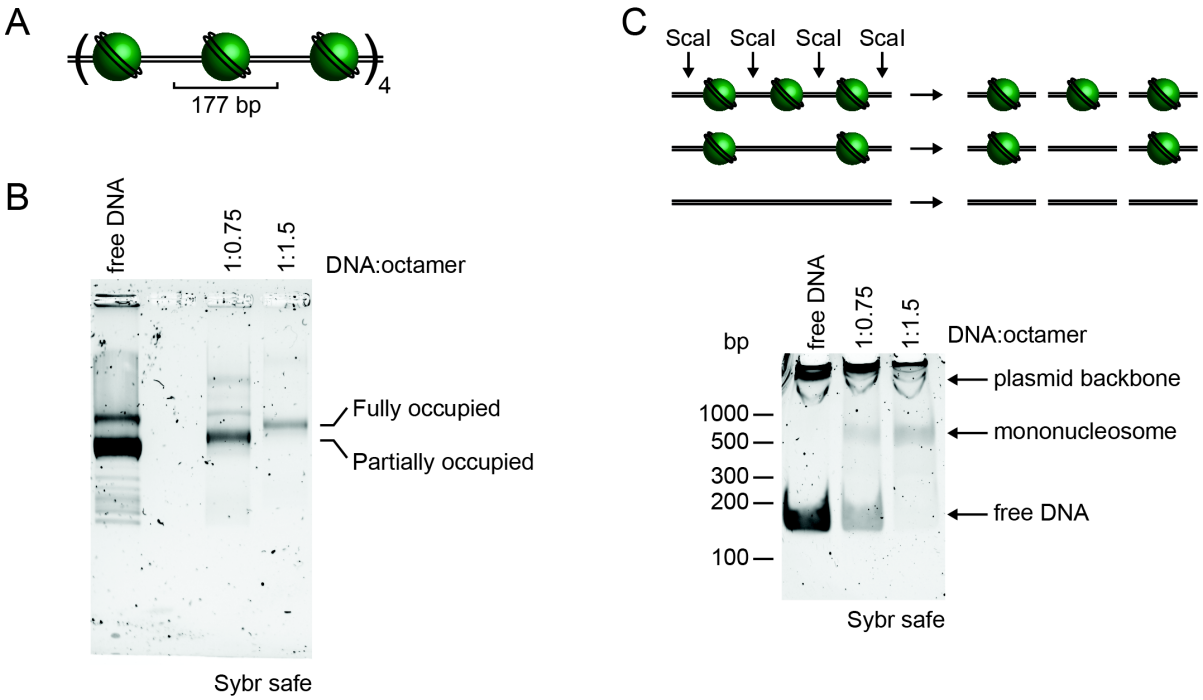



**Fig 4.2: Preparation of recombinant nucleosome arrays.**
**(A)** Arrays comprise 12 repeats of a 177 bp sequence containing the 601 nucleosome positioning sequence.
**(B)** Nucleosome arrays were reconstituted using different DNA:histone octamer mass ratios and analysed for gel shift by separating on a 0.5% native agarose gel and staining with Sybr safe.
**(C)** The quality of nucleosome arrays reconstituted *in vitro* was tested by digestion with ScaI (shown schematically, top) to release either free DNA from unoccupied positioning sequences or mononucleosomes from nucleosome-occupied sequences. Reactions were analysed by native PAGE followed by staining with Sybr safe (bottom), to determine optimal nucleosome assembly.

---

**(Fig. 4.1 cont.)**
**(B)** Following protein expression, harvested cells were lysed by sonication to prepare whole cell extract, which was applied to FLAG M2 resin. Protein-bound beads were washed and pure protein was eluted by competition with FLAG peptide. The purity and quality of protein was then analysed by SDS-PAGE followed by Coomassie staining, and protein was used for downstream applications including histone acetyltransferase (HAT) assays and interaction assays.
**(C)** Purified full length CBP protein was separated by SDS-PAGE on a 3-8% Tris-Acetate gel and stained with Coomassie.

free 177 bp DNA if a nucleosome positioning sequence is unoccupied, or to release of a higher molecular weight mononucleosome if the sequence is occupied by a histone octamer.

The results of this of this assay showed that reconstitution of arrays with a DNA:histone octamer mass ratio of 1:1.5 leads to full occupancy of the DNA but does not lead to oversaturation of the array, which would result in blocking of some *ScaI* sites and therefore to the release of dinucleosomes and polynucleosomes upon *ScaI* digestion (Fig.4.2C). These reconstituted nucleosome arrays were therefore a suitable substrate to test the enzymatic activity of CBP towards chromatin.

## 4.3 Establishing a histone acetyltransferase (HAT) assay

To test how CBP acetylates chromatin substrates, suitable HAT assays were first established and optimised. To this end, two types of HAT assay were developed, first using a $^{3}$H label to allow the overall acetylation of individual histones to be examined, and second using unlabelled substrate and antibodies to probe for acetylation of individual histone lysine residues by western blot (Fig. 4.3). In both assays, CBP enzyme was mixed together with nucleosome substrate, and the HAT reaction was initiated by addition of either $^{3}$H-labelled or unlabelled acetyl-CoA cofactor. The reactions were then quenched and analysed by SDS-PAGE followed by autoradiography for $^{3}$H-labelled HAT assays to allow simultaneous detection of total acetylation of the four histone proteins, or followed by western blot to examine individual histone acetylation marks for unlabelled reactions. Together, these assays provide powerful tools to assess the activity and specificity of CBP enzymes *in vitro*.

## 4.4 The C-terminus of CBP is required for acetylation of histone H3 in chromatin

CBP is a large, multi-domain protein (Fig. 4.4A) comprising three main parts. The enzymatic activity of CBP is located in the centre of the protein and contains the HAT domain itself, but also requires the bromodomain, PHD finger

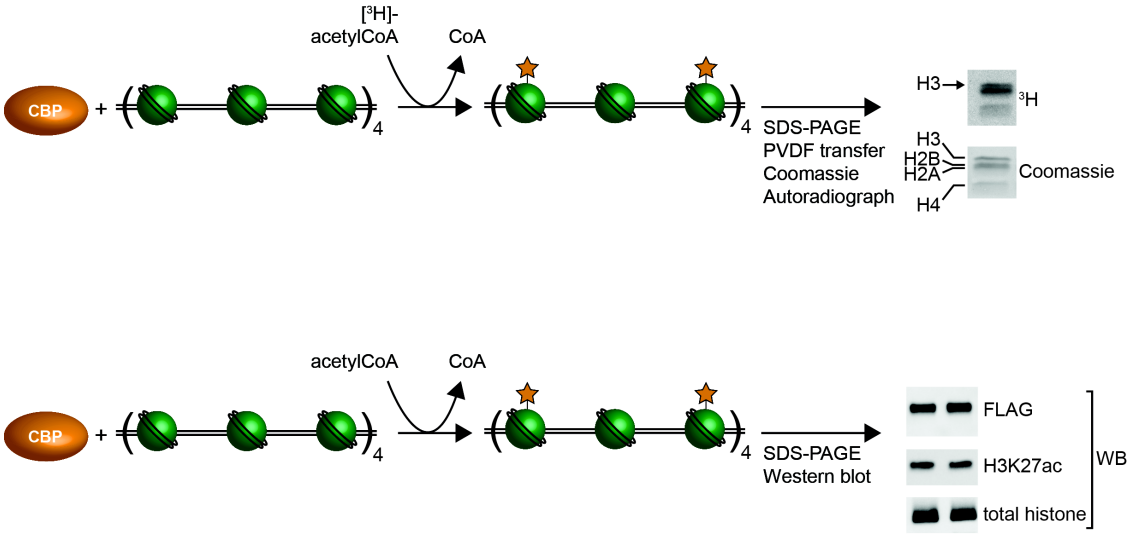



**Fig. 4.3: A histone acetyltransferase (HAT) assay to analyse CBP function.**
CBP enzyme is incubated with nucleosome arrays and an acetyltransferase reaction is started by addition of either $^{3}$H-labelled (top) or unlabelled (bottom) acetyl-CoA cofactor, leading to acetylation of histone residues (green stars). $^{3}$H-labelled acetylation reactions (top) are then analysed by SDS-PAGE followed by transfer to a PVDF membrane, Coomassie staining and exposure to film to generate an autoradiograph ($^{3}$H). Unlabelled acetylation reactions (bottom) are analysed by SDS-PAGE followed by western blot using antibodies against specific histone acetylation modifications, against total histones present in each reaction, and against FLAG to show levels of enzyme.

and RING finger to be fully functional (Delvecchio et al., 2013; Kalkhoven et al., 2002; Manning et al., 2001). Together these domains make up the CBP catalytic core. The N-terminal portion of the protein contains a transcriptional adaptor zinc finger (TAZ) domain called TAZ1, and a kinase inducible domain (KID)-interacting domain (KIX). Both of these domains mediate interactions with other proteins, including p53 for the TAZ1 domain (Krois et al., 2016) and a large number of KID-containing proteins such as CREB and c-Myb for the KIX domain (reviewed in Thakur et al., 2014). The C-terminus of CBP contains a ZZ-type zinc finger immediately downstream of the HAT domain, which has recently been shown to bind to the N-terminus of histone H3 (Zhang et al., 2018), a second TAZ-type zinc finger called TAZ2 and a nuclear coactivator binding domain (NCBD) which interacts with proteins including p160 coactivators such as SRC-1 (Waters et al., 2006).

To understand the contribution of these domains to the acetylation of histone H3 in a physiologically relevant chromatin substrate, CBP proteins containing CBP core together with either the N-terminus (Nter-core) or C-terminus (core-

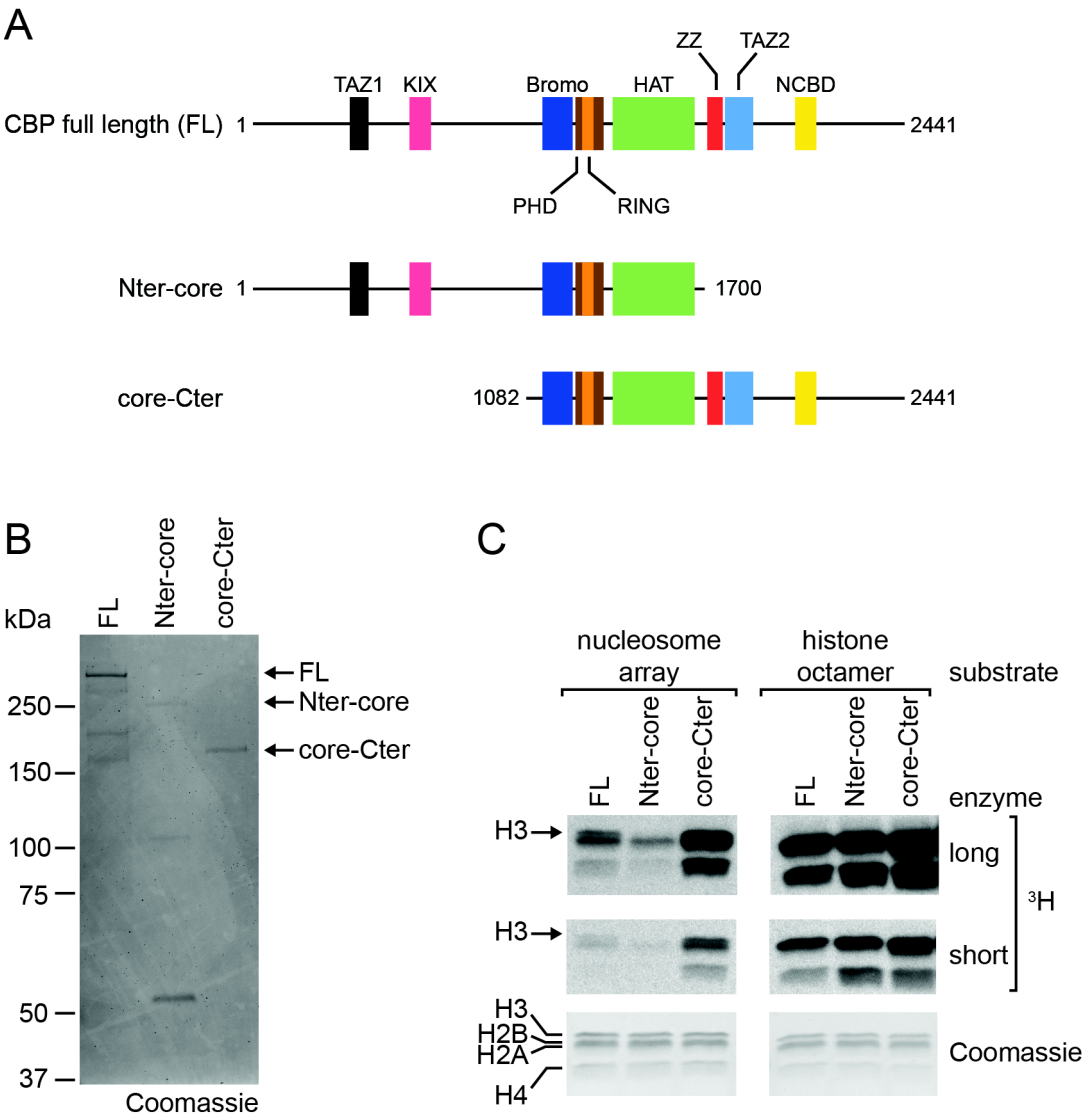



**Figure 4.4: CBP C-terminus mediates histone H3 acetylation in nucleosome substrates.**
**(A)** Schematic representation of the domain architecture of full length (FL) mouse CBP (top), Nter-core (middle) and core-Cter (bottom) proteins, showing unstructured regions (line) and structured domains (boxes), with numbers representing amino acid positions. TAZ1, transcription adaptor zinc finger domain 1; KIX, kinase-inducible domain (KID)-interacting domain; Bromo, bromodomain; PHD, plant homeodomain zinc finger; RING, really interesting new gene zinc finger; HAT, histone acetyltransferase domain; ZZ, ZZ-type zinc finger; TAZ2, transcription adaptor zinc finger domain 2; NCBD, nuclear coactivator binding domain.
**(B)** Proteins expressed in Sf9 cells were purified via the FLAG tag and analysed by SDS-PAGE followed by Coomassie staining.
**(C)** HAT assays using CBP FL, Nter-core or core-Cter enzymes with nucleosome array or histone octamer substrates as indicated. Coomassie staining of membrane shows loading of histone proteins, with individual histone proteins indicated, and $^{3}H$ shows long and short autoradiograph exposures with the position of histone H3 indicated.

Cter) were expressed and purified from Sf9 cells (Fig. 4.4A, B). The activities of full length CBP (FL), CBP Nter-core and CBP core-Cter were then tested using either nucleosome arrays or histone octamers as substrates in a $^{3}$H-labelled HAT assay (Fig. 4.4C). All three enzymes had comparable activity towards histone octamers, which in the conditions of this experiment likely dissociate to form H2A-H2B dimers and H3-H4 tetramers, including similar activity towards histone H3. Moreover, consistent with previous observations, the activity of CBP was greater towards histone octamer than towards nucleosomes and appears to be primarily directed against histones H3 and H4 (An and Roeder, 2003; Kraus et al., 1999).

Importantly, using nucleosome array substrates, only enzymes containing the C-terminus of CBP were active towards histone H3, confirming previous studies showing that the CBP core is not sufficient for H3 acetylation in nucleosomes (Bannister and Kouzarides, 1996; Zhang et al., 2018), whilst the N-terminus of CBP was dispensable for H3 acetylation. In addition, in nucleosomes specificity appeared to be shifted towards acetylation of H2A and/or H2B, in agreement with previous work (An and Roeder, 2003; Kraus et al., 1999), and this shift was enhanced for enzymes containing the CBP C-terminus. Together, these results suggest that CBP has altered substrate specificity towards nucleosome substrates compared to histone octamers alone, and that an activity is contained within the C-terminus of CBP that is required for efficient acetylation of histone H3 in the context of nucleosomes.

## 4.5 The CBP TAZ2 domain is required for H3K27ac

To determine which domains within the CBP C-terminus are required to direct H3 acetylation, further constructs were generated containing the two most promising candidate domains, namely the ZZ and TAZ2 domains immediately adjacent to the HAT domain (Fig. 4.5A). CBP core (C), core-ZZ (CZ) and core-ZZ-TAZ2 (CZT) proteins were expressed and purified from Sf9 cells (Fig. 4.5B) and their enzymatic activities tested towards nucleosome arrays in $^{3}$H-based HAT assays (Fig. 4.5C). This experiment showed that while CBP core is active towards nucleosomes, it lacks detectable activity towards histone H3, similar to the Nter-core construct (Fig. 4.4C). While showing similar levels of activity

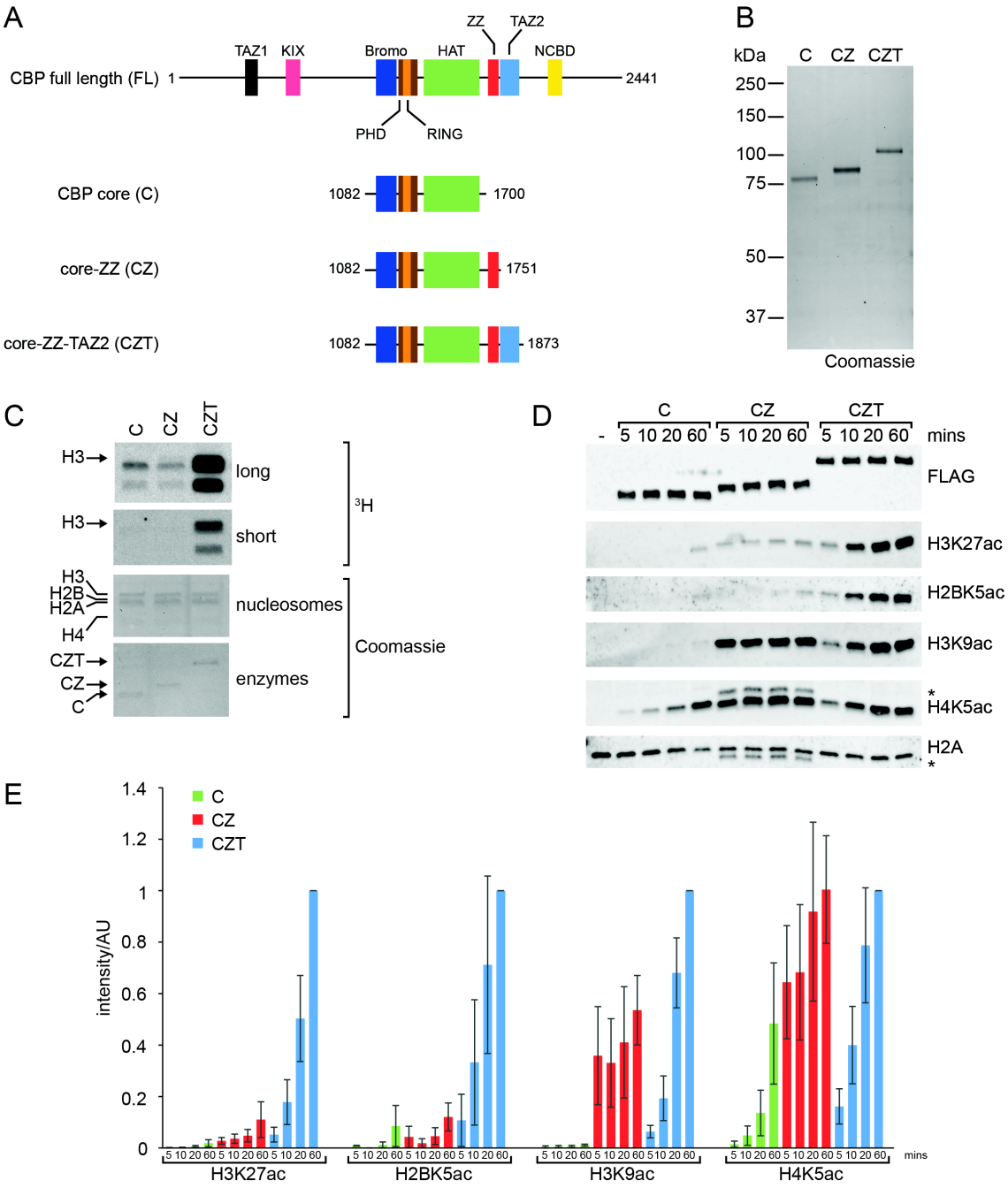



**Figure 4.5: The CBP TAZ2 domain is required for efficient H3K27ac.**
**(A)** Schematic representation of the domain architecture of CBP FL, CBP enzymatic core (C), core-ZZ (CZ) and core-ZZ-TAZ2 (CZT).
**(B)** Proteins expressed in *Sf9* cells were purified via the FLAG tag and analysed by SDS-PAGE followed by Coomassie staining.
**(C)** HAT assays showing reactions with C, CZ and CZT enzymes with nucleosome array. Long and short autoradiograph exposures, and Coomassie staining to show enzymes and histone proteins are shown.
**(D)** Western blot analysis of unlabelled HAT assay timecourse experiments using no enzyme (-), C, CZ or CZT with nucleosome array, for reaction times of 5, 10, 20 or 60 mins, analysed using antibodies against FLAG, H3K27ac, H2BK5ac, H3K9ac, H4K5ac or total H2A. Asterisks indicate non-specific bands present in CZ reactions (see Appendix, Fig. S1).
**(E)** Quantification of HAT assay shown in (D). Signal for each antibody is normalized to CZT reaction at 60 mins and represents the mean of three independent experiments, with error bars indicating SEM.

towards histones H2A, H2B and H4 as CBP core, CZ exhibited a small but clear increase in acetylation towards H3, consistent with previous work suggesting that the ZZ domain influences CBP activity towards nucleosomes and increases H3 acetylation (Bannister and Kouzarides, 1996; Zhang et al., 2018). CZT, by contrast, showed higher levels of acetylation of all four core histone proteins andacetylates H3 to far higher levels than CZ, indicating that the TAZ2 domain of CBP plays an important role in directing H3 acetylation in nucleosomes.

To extend this analysis, unlabelled HAT assays were carried out followed by western blot for H3K9ac, H3K27ac, H2BK5ac and H4K5ac to determine which specific histone lysine residues are acetylated by CBP in the presence of the ZZ and TAZ2 domains (Fig. 4.5D, E). Consistent with the $^{3}$H-labelled HAT assays, this time course experiment showed that CBP core only weakly acetylates any of the residues tested. CZ shows markedly greater activity towards H3K9 and H4K5 than CBP core alone, but shows only slightly increased activity towards H3K27 and H2BK5. CZT, by contrast, is able to acetylate H3K9 and H4K5 to the same final levels as CZ, albeit with a decreased rate of reaction, but, importantly, CZT generates ten-fold higher levels of H3K27ac and H2BK5ac. This suggests that the higher levels of H3 acetylation generated by CZT are attributable primarily to increased H3K27ac and that the TAZ2 domain of CBP is therefore essential for efficient acetylation of H3K27 in nucleosome substrates.

## 4.6 Summary and discussion

CBP/p300 regulates transcription through acetylating histones, with the major products of CBP/p300-mediated histone acetylation *in vivo* being H3K18ac, H2B acetylation, and the key marker of active enhancers, H3K27ac (Jin et al., 2011; Weinert et al., 2018). However, *in vitro* studies have suggested a far more promiscuous HAT activity for CBP/p300, and it has been unclear how to reconcile these two sets of observations.

Work in this chapter attempted to resolve this discrepancy by examining CBP enzymatic activity *in vitro* using purified proteins to avoid the possibility of

compensatory effects in the cell, while making use of reconstituted nucleosome array substrates to better reflect the substrate encountered by CBP/p300 *in vivo*. These experiments showed that CBP has altered specificity in the context of nucleosomes compared to non-nucleosomal histone substrates. Moreover, acetylation of key targets of CBP/p300 *in vivo*, such as H3K27 and H2BK5, requires a zinc finger domain called TAZ2 located in the C-terminus of CBP in addition to the enzymatic core.

These results suggest therefore that enzymatic assays using nucleosome array as substrate can recapitulate CBP activity towards H3K27ac observed *in vivo*. It further suggests that packaging histones into nucleosomes is inherently suppressive to the catalytic activity of the CBP enzymatic core towards H3K27. This suppression can be partially overcome by the presence of the ZZ domain of CBP, in agreement with previous work (Bannister and Kouzarides, 1996; Zhang et al., 2018), presumably through binding of the ZZ domain to the N-terminal tail of histone H3 (Zhang et al., 2018). However, these reports suggested that the ZZ domain is sufficient to generate high levels of H3K27ac, whereas this study indicates that the TAZ2 domain is also required. One explanation for this is that while these previous studies use nucleosomes as substrates, mononucleosomes were used rather than nucleosome arrays. This suggests that the ZZ domain of CBP/p300 is sufficient to overcome the suppressive effect towards H3K27ac of assembly into nucleosomes but not of packaging into the more physiological substrate of nucleosome arrays.

The TAZ2 domain therefore plays an important role in specifying CBP/p300 activity towards H3K27, and could explain how CBP/p300 targets H3K27 *in vivo*. Understanding the mechanism by which the TAZ2 domain achieves targeted H3K27ac in chromatin could therefore provide key insights into how H3K27ac is generated by CBP/p300 genome-wide and how these chromatin-modifying enzymes function to regulate gene expression.

# 5. The CBP TAZ2 domain binds DNA to drive interactions with chromatin

The TAZ2 domain of CBP directs the acetylation of H3K27 in chromatin substrates whilst reducing promiscuous activity towards other histone residues (see Chapter 4). TAZ2 is a highly conserved domain (Ponting et al., 1996; Sebé-Pedrós et al., 2011) and mutations in the domain in humans are associated with disease (Menke et al., 2016, 2018), consistent with TAZ2 playing an important role in mediating CBP/p300 function. However, how mechanistically TAZ2 might contribute to histone acetylation specificity is not clear.

Previous work has suggested that TAZ2 is involved in mediating protein-protein interactions with transcription factors that facilitate CBP/p300 recruitment to target sites, with structures solved for TAZ2 interactions with the intrinsically disordered activation domains of p53, STAT1 and C/EBP (Bhaumik et al., 2014; Feng et al., 2009; Jenkins et al., 2009, 2015; Wojciak et al., 2009). Whilst for the best studied interaction with p53 there have been attempts to address functional relevance through the design of mutations that disrupt interactions *in vitro* (Jenkins et al., 2009, 2015), for other interactions this has remained unexplored. Moreover, little work has been carried out to test the importance of these interactions *in vivo*. Furthermore, the validity of these structures is unclear given that in a crystal structure of free TAZ2, a normally unstructured region at the C-terminus of the domain has been found to fold into a helical structure and interact with a neighbouring TAZ2 molecule (Dyson and Wright, 2016; Miller et al., 2009). This suggests that under the conditions of crystallisation, TAZ2 could readily form non-physiological interactions with unstructured protein sequences, meaning that the use of high concentrations of interacting peptide or chimeric TAZ2 fusion proteins may result in structures that are difficult to interpret.

Importantly, in HAT assays using purified CBP proteins, none of the proposed interaction partners should be present to influence TAZ2 function. Moreover, if trace amounts of transcription factor were present in reactions through

interaction of CBP with endogenous Sf9 proteins, there are no binding sites for known CBP interaction partners within the DNA template used in these assays. This suggests that TAZ2 might influence histone acetylation specificity through a novel mechanism unrelated to interaction with other proteins.

In HAT assays, TAZ2 influences CBP function in the presence only of chromatin, acetyl-CoA and the enzyme itself. One mechanism through which TAZ2 could function is through autoregulation of the enzyme. CBP/p300 is regulated through an unstructured loop region within the HAT domain that occludes the substrate binding pocket of the enzyme. Autoacetylation of this loop leads to its eviction from the catalytic site and to relief of CBP/p300 inhibition (Liu et al., 2008; Ortega et al., 2018; Thompson et al., 2004). One possible mechanism, therefore, through which TAZ2 could regulate CBP/p300 function is through influencing acetylation of the autoinhibitory loop. However, the TAZ2 domain does not affect CBP activity towards non-nucleosome substrates, such as histone octamers, *in vitro.* Moreover, previous studies have shown that deletion of TAZ2 does not affect CBP/p300 autoacetylation (Kraus et al., 1999; Ortega et al., 2018) and that TAZ2 does not form direct interactions with the CBP/p300 core (Aguilar-Gurrieri, 2013). Together, these observations suggest that TAZ2 does not function to influence CBP/p300 activity by modulating autoregulation.

An intriguing alternative possibility, therefore, is that TAZ2 might affect CBP/p300 substrate specificity by interacting with chromatin. To test this possibility, a combination of sequence and structural analyses together with *in vitro* interaction and enzymatic assays were used to determine how mechanistically the TAZ2 domain contributes to CBP/p300 function.

## 5.1 TAZ2 is a highly conserved and positively charged domain

To address how TAZ2 functions mechanistically, the sequence of the domain was first examined by alignment with CBP/p300 proteins from humans to the early multicellular animal *Trichoplax adhaerens* and the unicellular protist *Capsaspora owczarzaki* (Fig. 5.1A). This analysis shows that the TAZ2 domain is highly conserved even among distantly related species, and that this conservation extends not only to key structural residues, such as the zinc-coordinating residues in this zinc finger domain, but also to other non-structural amino acids, including multiple lysine and arginine residues. Crystal structures of the domain show that TAZ2 is organised into four alpha helices, with three zinc-binding clusters in the loops and helix ends (Fig. 5.1B). Calculation of the surface charge across the domain shows that the TAZ2 surface is strongly positively charged (Fig. 5.1C), consistent with the conservation of basic amino acids, yielding an isoelectric point (pI) of 10.1. By comparison, the pI of human histone proteins ranges from 10.3 for H2B to 11.9 for H3. Together, these observations suggest the intriguing possibility that, in the context of HAT assays, TAZ2 could mediate interactions with negatively charged molecules, such as the DNA onto which nucleosomes are assembled.

## 5.2 TAZ2 is a sequence-independent DNA-binding domain

To test whether the TAZ2 domain interacts with DNA, TAZ2 was first expressed and purified from bacteria, with the ZZ domain purified as a negative control domain (Fig. 5.2A). DNA pulldown experiments were then carried out by incubating ZZ or TAZ2 with either streptavidin beads alone or with streptavidin beads coated in biotinylated 147 bp DNA. These assays showed that whilst ZZ binds in neither condition, TAZ2 binds only in the presence of DNA (Fig. 5.2B). This TAZ2 DNA-binding activity was confirmed by electrophoretic mobility shift assay (EMSA) experiments, which show that whilst increasing concentrations of TAZ2 are able to generate a band shift, ZZ is unable to bind DNA even at relatively high concentrations (Fig. 5.2C).

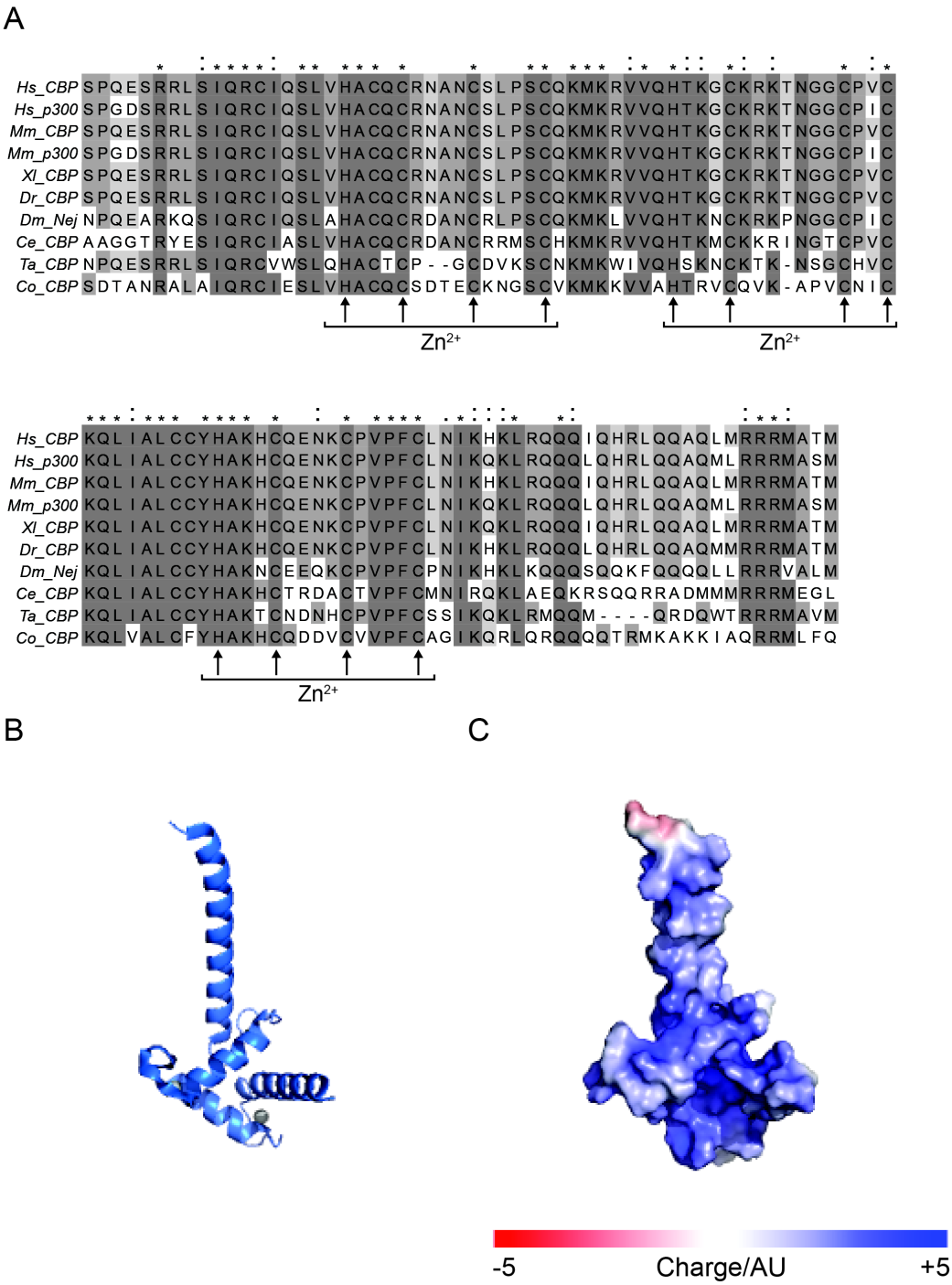



**Figure 5.1: The CBP TAZ2 domain is highly conserved with a positive surface charge.**
**(A)** Alignment of CBP and p300 protein sequences in humans (*Hs*), mouse (*Mm*), *Xenopus laevis* (*Xl*), *Danio rerio* (*Dr*), *Drosophila melanogaster* (*Dm*), *Caenorhabditis elegans* (*Ce*), *Trichoplax adhaerens* (*Ta*) and *Capsaspora owczarzarki* (*Co*); greyscale shading shows the level of conservation, asterisks (*) indicate identical residues, colons (:) indicate similar residues, and arrows beneath the sequences indicate zinc-coordinating residues.
**(B)** Cartoon depicting the crystal structure of human p300 TAZ2 domain (PDB ID: 3IO2) showing protein in blue and coordinated zinc ions as grey spheres.
**(C)** Surface depiction of the structure in (B) with colour-coded surface charge.

To understand whether TAZ2 exhibits sequence specificity or preferences in its DNA binding, further EMSA experiments were carried out using four 29-bp DNA probes of either 40% or 60% GC content with otherwise randomly generated DNA sequences (Fig. 5.2D). These show that TAZ2 binds to these sequences with similar affinity, suggesting that TAZ2 binds to DNA independent of its sequence.

To determine whether this DNA binding of TAZ2 also facilitates the interaction of enzymatically active constructs with DNA, and could therefore explain how TAZ2 affects CBP histone acetylation activity, EMSA experiments were carried out using CBP core-ZZ (CZ) and CBP core-ZZ-TAZ2 (CZT) proteins together with 147 bp DNA (Fig. 5.2E). Whilst CZ was unable to bind DNA, inclusion of the TAZ2 domain in the construct was sufficient to achieve DNA binding. Together, these results show that TAZ2 binds DNA in a sequence-independent manner and can mediate interactions between an active CBP protein and free DNA.

## 5.3 TAZ2 drives CBP interactions with nucleosomes

The observation that TAZ2 binds to free DNA raised the possibility that TAZ2 might function to mediate interactions between CBP and nucleosomes. To test whether TAZ2 can interact with nucleosomal DNA or whether it would require free linker DNA to interact with nucleosomes, EMSA experiments were carried out with TAZ2 and ZZ domains and nucleosome core particles (NCPs) assembled with either the minimal 147 bp 601 nucleosome positioning DNA sequence or a longer 209 bp sequence with 31 bp linker DNA overhangs (Fig. 5.3A, B). These experiments showed that whilst the ZZ domain is unable to bind to NCPs at the concentrations tested, despite the previously identified interaction with histone H3 (Zhang et al., 2018), TAZ2 is able to bind to both 147 bp and 209 bp NCPs with similar affinity, suggesting that TAZ2 can bind to nucleosomal DNA as well as free DNA.

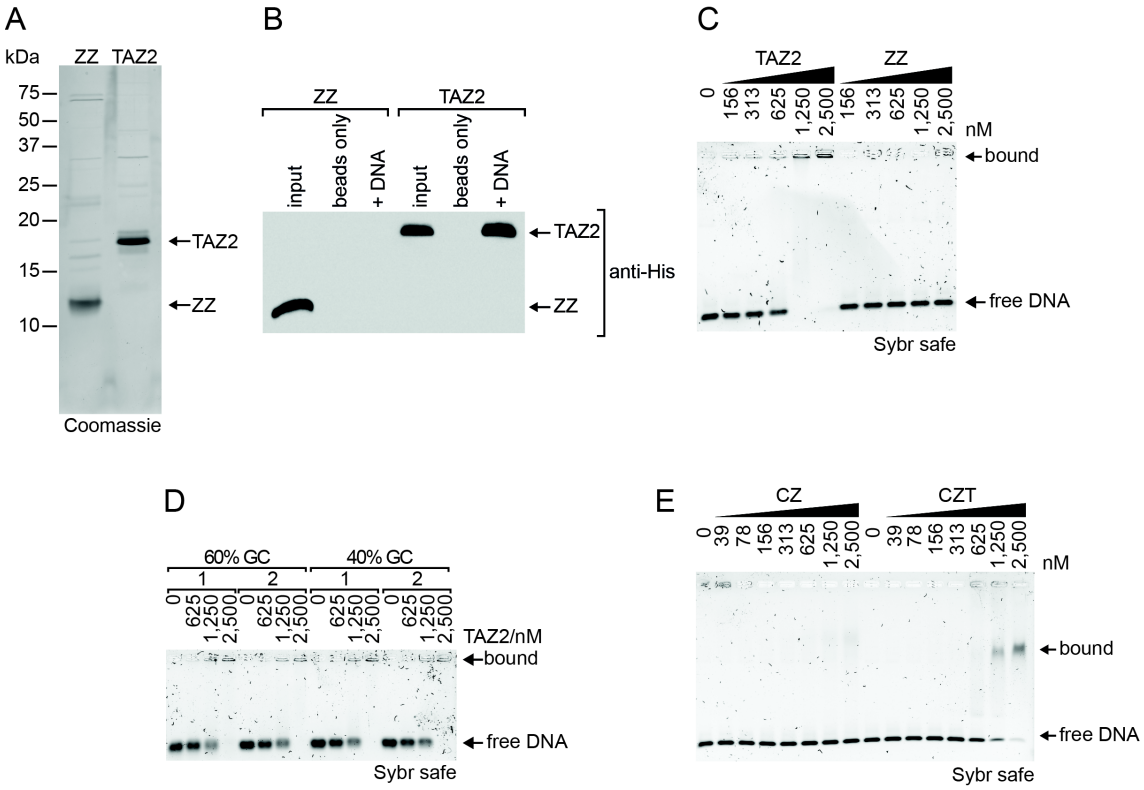



**Fig. 5.2: CBP TAZ2 is a sequence-independent DNA-binding domain.**
**(A)** CBP ZZ and TAZ2 domains were expressed in bacteria, purified via a 6xHis tag and analysed by SDS-PAGE followed by Coomassie staining.
**(B)** Pulldown assay with ZZ or TAZ2 domain and either streptavidin beads alone or streptavidin beads bound by biotinylated 147 bp DNA.
**(C)** Binding reactions using increasing concentrations of TAZ2 or ZZ domain with 147 bp DNA analysed by EMSA.
**(D)** Binding reactions using increasing concentrations of TAZ2 domain with 29 bp duplex DNAs of differing GC content and sequence analysed by EMSA.
**(E)** Binding reactions with increasing concentrations of CBP core-ZZ (CZ) and core-ZZ-TAZ2 (CZT) with 147 bp DNA analysed by EMSA.

To determine whether this DNA binding activity can mediate interactions between an enzymatically competent construct and nucleosomes, EMSAs were carried out using CZ and CZT proteins together with 147 bp and 209 bp NCPs (Fig. 5.3C, D). These experiments show that whilst CZ can generate some binding to NCPs resulting in a small band shift, CZT generates far more efficient binding, as measured by depletion of the free NCP band, for both 147 bp and 209 bp NCPs. Quantification of these EMSAs shows that CZT binds to 147 bp NCPs with a dissociation constant ($K_d$) of 1.0 µM and to 209 bp NCPs with a $K_d$ of 540 nM, indicating that whilst TAZ2 can bind to nucleosomal DNA, the presence of linker DNA aids the association of CZT with nucleosomes.

Finally, to test whether TAZ2 is important for mediating interactions with the more physiological substrate of chromatin, EMSAs were carried out with CBP core, CZ and CZT proteins with 12-mer nucleosome arrays as substrates (Fig. 5.3E). This shows that whilst only weak binding could be observed between CBP core or CZ and nucleosome arrays, CZT binds with a $K_d$ of 309 nM, indicating that CZT binds to nucleosome arrays with approximately three-fold higher affinity than to 147 bp NCPs.

These results show that the TAZ2 domain of CBP can bind to nucleosomal DNA, and that the presence of this domain is necessary and sufficient for stable binding of CBP constructs to nucleosomes. This suggests that mediating interactions with nucleosomes may be a key function of the TAZ2 domain, and that this may explain how TAZ2 determines CBP substrate specificity.

## 5.4 TAZ2 DNA binding determines CBP activity towards nucleosomes

To test whether the DNA binding activity of TAZ2 determines how CBP interacts with its substrate in enzymatic assays, HAT assays were performed using CBP core, CZ, core-ΔZZ-TAZ2 (CΔZT) and CZT. These were carried out using nucleosome array substrates assembled with two different ratios of DNA:histone octamer to generate chromatin with different numbers of nucleosomes per DNA molecule (Fig. 5.4A). Using these substrates shows that CBP core and CZ are similarly active towards the two different substrates, whilst CΔZT and CZT show higher activity towards substrates with higher numbers of nucleosomes per DNA at all tested acetylation sites. This suggests that in the presence of the TAZ2 domain, TAZ2-mediated DNA binding drives chromatin association leading to greater levels of acetylation when more nucleosomes are found on the same DNA molecule. By contrast, in the absence of TAZ2, the location of nucleosomes on DNA is no longer important and equal levels of acetylation are seen regardless of the number of nucleosomes found on each DNA. Moreover, the effect seen with the TAZ2 domain requires only this domain as the effect can be observed in reactions with CΔZT constructs that lack the ZZ domain.

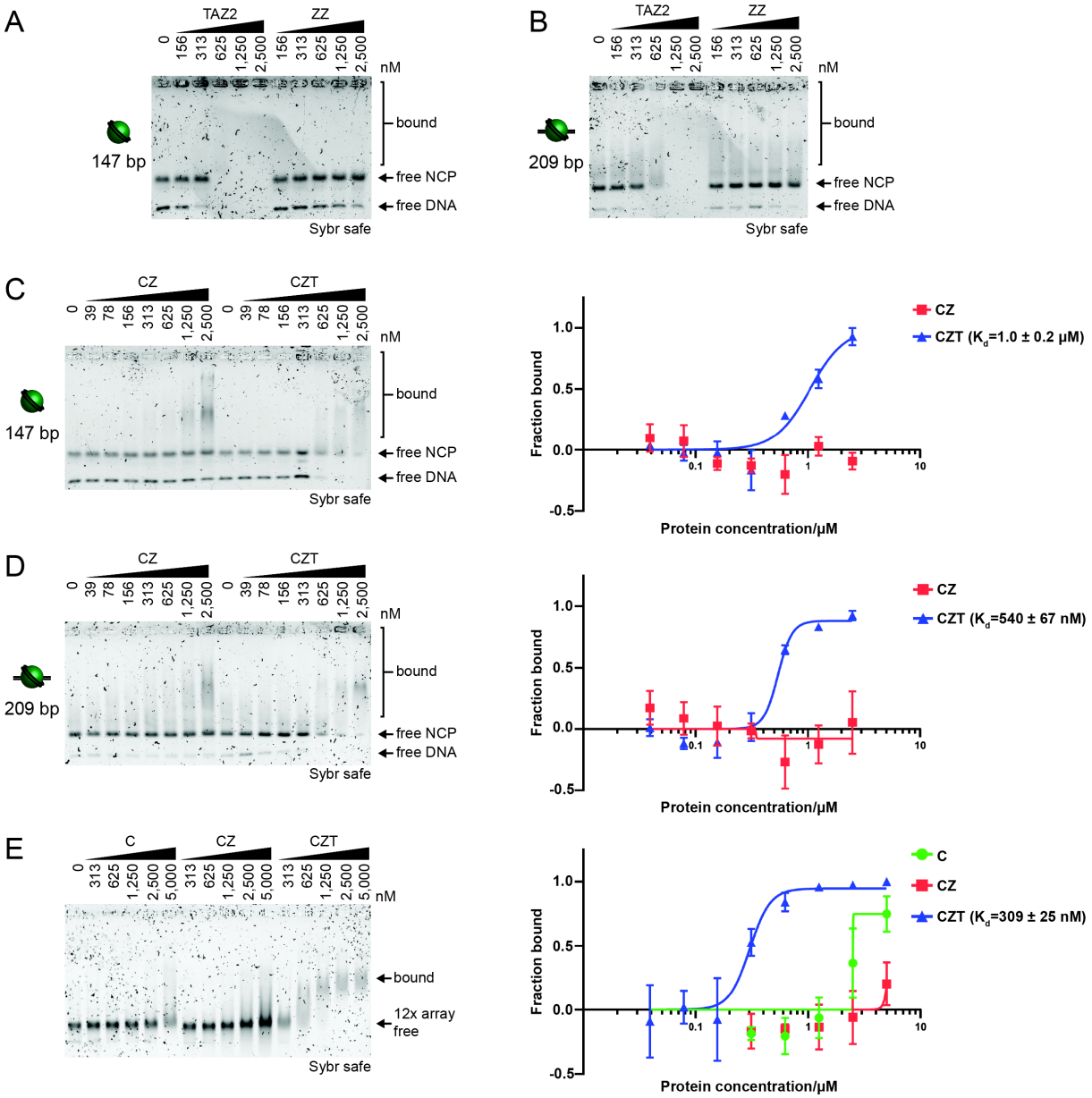



**Fig. 5.3: CBP TAZ2 domain mediates interaction with nucleosomes.**
**(A), (B)** Binding reactions using increasing concentrations of TAZ2 or ZZ domain with 147 bp (A) or 209 bp (B) nucleosome core particle (NCP), analysed by EMSA.
**(C), (D)** Left: binding reactions using increasing concentrations of CZ and CZT with 147 bp (A) or 209 bp (B) NCPs analysed by EMSA. Right: quantitative analysis of EMSAs from three independent experiments, measured by depletion of free NCP band. Error bars represent SEM.
**(E)** Left: binding reactions using increasing concentrations of C, CZ or CZT with 12x nucleosome array analysed by EMSA. Right: quantitative analysis of EMSA from three independent experiments, measured by depletion of free nucleosome array band. Error bars represent SEM.

To understand further how TAZ2-mediated nucleosome interactions influence CBP enzymatic activity, HAT assays were carried out using CBP core, CZ and CZT enzymes with either 147 bp NCP or nucleosome array as substrates (Fig. 5.4B, C). The results show that whilst CBP core has little enzymatic activity toward either substrate, CZ is highly active towards all tested residues in NCP substrates. The increase in activity between CBP core and CZ towards these residues in NCPs shows that the ZZ domain is sufficient to increase acetylation towards multiple histone tail lysines in NCPs, not only towards H3K27 as reported previously (Zhang et al., 2018). The reason for this discrepancy is most likely that whilst the present study uses a two-fold molar excess of nucleosome compared to enzyme, previous work has used a three-fold molar excess of enzyme (Zhang et al., 2018). This could mean that in this previous report, acetylation of residues such as H3K9 was saturated and appeared to remain unchanged by the presence of the ZZ domain, whilst less readily modified residues such as H3K27 continued to show increased acetylation.

The addition of the TAZ2 domain, however, led to decreased levels of acetylation in NCP substrates compared to CZ, in spite of the observation that CZT has higher affinity for NCPs than proteins lacking TAZ2 (Fig. 5.3C, D). One explanation for this is that the increased affinity of CZT for NCPs leads to a stable association between CZT and nucleosomes that effectively sequesters the enzyme. This would therefore prevent CZT from detaching from one NCP substrate and re-binding to a second, resulting in an overall reduction in acetylation. This implies that whilst the previously reported binding of the ZZ domain to the N-terminal tail of histone H3 (Zhang et al., 2018) is sufficient to bring about a transient interaction between CZ and nucleosomes which increases acetylation compared to CBP core alone, this interaction does not lead to stable binding of CZ to NCPs, and therefore facilitates greater levels of acetylation of this non-physiological substrate.

Importantly, incorporation of nucleosomes into chromatin arrays led to a decrease in CZ activity. The extent of this decrease, however, differs between different histone lysine residues, so that whilst there is an approximately ten-fold loss in H3K27ac and H2BK5ac, there is only a two-fold loss of H3K9ac and

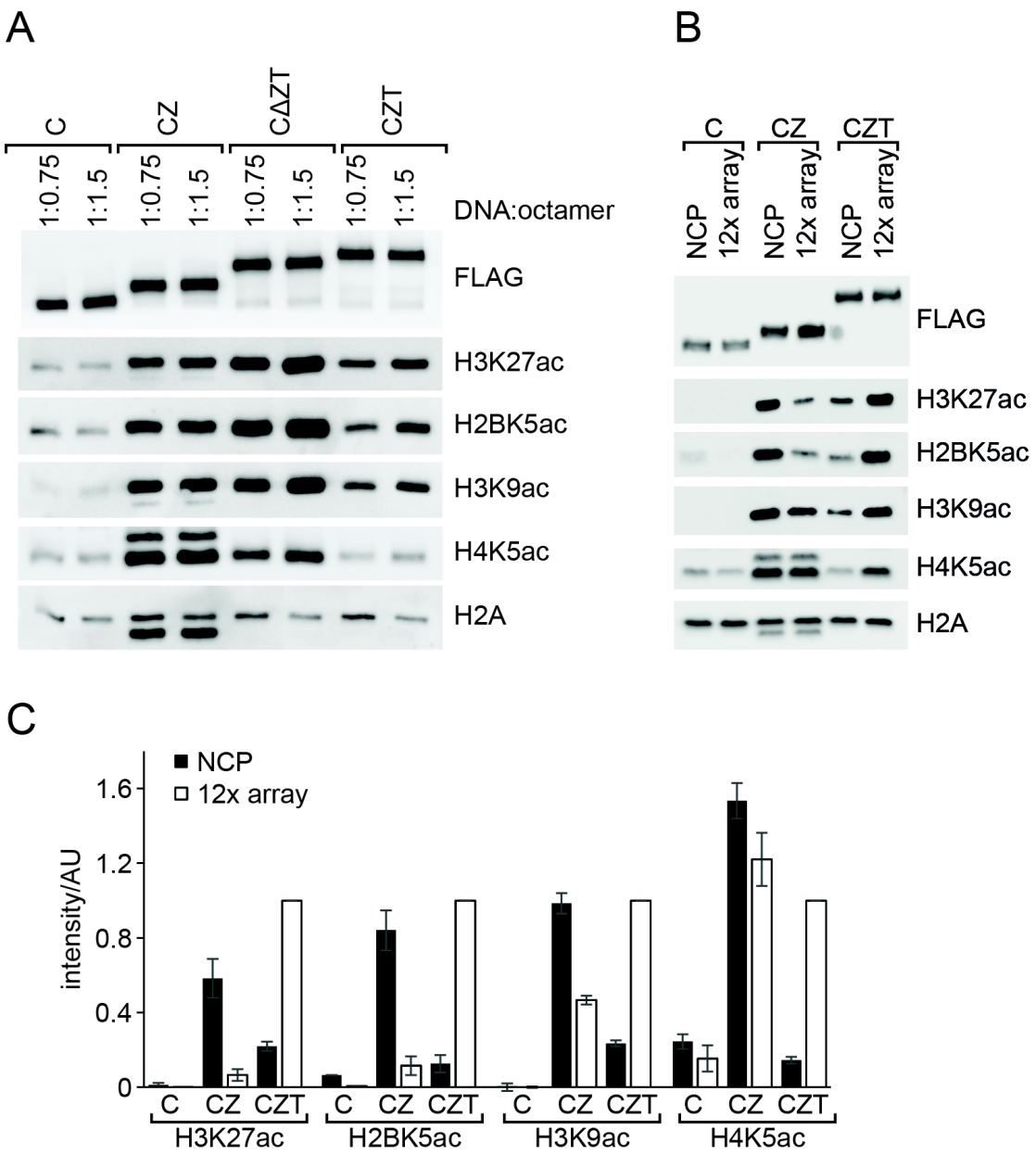



**Fig 5.4: CBP TAZ2-mediated DNA binding defines histone acetylation in chromatin.**
**(A)** Western blot analysis of HAT assays using C, CZ, CΔZT or CZT enzymes with nucleosome array substrates of differing nucleosome occupancy, analysed using antibodies against FLAG, H3K27ac, H2BK5ac, H3K9ac, H4K5ac, or total H2A.
**(B)** Western blot analysis of HAT assays using C, CZ or CZT enzymes with 147 bp NCP or nucleosome array substrates, analysed using antibodies against FLAG, H3K27ac, H2BK5ac, H3K9ac, H4K5ac or total H2A.
**(C)** Quantification of HAT assay shown in (B). Signal for each antibody is normalized to CZT reaction with nucleosome array and represents the mean of three independent experiments, with error bars indicating SEM.

little change in H4K5ac. This suggests that whilst the transient interaction with nucleosome substrates mediated by ZZ is sufficient to achieve acetylation of H3K9 and H4K5, it is insufficient to generate high levels of H3K27ac and H2BK5ac in more physiological chromatin substrates. This suggests that H3K9 and H4K5 residues might be more accessible in chromatin than H3K27 and H2BK5, and that acetylation of these latter residues might be hindered because of the additional DNA present in nucleosome arrays compared to NCPs.

With CZT enzyme, by contrast, assembly of nucleosomes into arrays led to an increase in activity compared to NCPs, with a five- to ten-fold increase in acetylation of all tested residues. Importantly, however, whilst there is an approximately ten-fold increase in H3K27ac and H2BK5ac in nucleosome arrays between CZ and CZT, the addition of the TAZ2 domain results in only a two-fold increase in H3K9ac and little change in H4K5ac. This shows that the activity of the TAZ2 domain, which binds DNA, can overcome the suppressive effect that incorporation into nucleosome arrays has on acetylation of H3K27 and H2BK5, but is largely dispensable for the acetylation of H3K9 and H4K5.

One mechanism by which TAZ2 might achieve this modulation of activity is through increasing the residency time of the enzyme on nucleosomes, allowing it to direct its activity towards less accessible residues over time. However, if increasing residency time on chromatin were the only function of TAZ2, it would be expected that acetylation of all residues would be increased to a similar extent, whereas H3K27ac and H2BK5ac show specifically greater increases in acetylation. This suggests that an additional mechanism may exist, such as that TAZ2 DNA binding promotes interaction with the nucleosome in a conformation that enables efficient acetylation of these physiologically important residues.

## 5.5 CBP interacts with the nucleosome via the enzymatic core and the TAZ2 domain

To address how TAZ2 interacts with the nucleosome and whether this could explain how mechanistically TAZ2 directs H3K27ac, a crosslinking mass

spectrometry approach using CZT protein and 147 bp NCPs was adopted to map to the level of individual amino acids how domains within CBP interact with one another and how CBP interacts with the nucleosome. This technique (reviewed in Rappsilber, 2011) takes advantage of the zero-length chemical crosslinker EDC ((1-ethyl-3-(3-dimethylaminopropyl) carbodiimide) to covalently link the carboxyl groups of aspartate or glutamate side chains with either primary amines found in lysine side chains and N-terminal residues or the hydroxyl groups of serine, threonine and tyrosine residues, before analysing the resulting products by mass spectrometry to identify amino acids that are found in close spatial proximity.

First, confirming previous work suggesting that TAZ2 does not form interactions within CBP that might influence its activity (Aguilar-Gurrieri, 2013), no crosslinks were detected between TAZ2 and other regions of CBP. By contrast, extensive crosslinks were observed between other domains of the protein, fitting well with the compact structure previously observed for the CBP/p300 enzymatic core (Delvecchio et al., 2013; Park et al., 2017).

Second, multiple contacts were detected between CBP and the histone proteins of the nucleosome (Fig. 5.5A, B). Three crosslinks were detected between the CBP bromodomain and the nucleosome. The first connects K1177 of CBP, within the fourth alpha helix of the bromodomain and located near the acetyl-lysine binding site, with E52 of histone H4, a residue on the outer surface of the nucleosome near to the site where the H3 N-terminal tail exits the nucleosome core. Two further residues from the same alpha helix within the bromodomain, E1184 and E1189, located distal to the acetyl-lysine binding pocket and closer to the exit to the substrate-binding site of the HAT domain, were crosslinked in proximity to the key substrate site of H3K27. Moreover, two residues located near the opening to the substrate-binding pocket within the HAT domain, E1354 and E1357, form contacts with residues near H3K27, namely H3T22 and H3S28, and E1357 forms an additional contact with H2BP1. These interactions between the CBP core and both the surface of histone H4 and the N-terminal tails of histone H3 and H2B suggest that the enzyme is positioned for acetylation of H3K27 and the H2B tail, and that, consistent with previous work, contacts formed by the bromodomain are key to interaction with the

nucleosome and therefore to enzymatic activity (Manning et al., 2001; Zhang et al., 2018).

In addition to these interactions between histone proteins and the CBP core, one further contact was detected between the TAZ2 domain and the nucleosome. One explanation as to why so few crosslinks were observed between TAZ2 and histone proteins despite the key role played by the domain in nucleosome interactions is that many TAZ2 residues, particularly positively charged and crosslink-forming lysine residues, are likely to be in contact with DNA. Multiple amino acids that are involved in interacting with DNA are therefore likely to be shielded from coming into sufficiently close proximity with histone proteins for the formation of crosslinks, despite such residues playing an important role in the interaction with nucleosomes. Nevertheless, residue K1807 of TAZ2, located at the C-terminal end of the second alpha helix, crosslinks to H3E59, a residue found in the globular domain of histone H3 on the lateral surface of the histone octamer, close to nucleosomal DNA and to the site at which the H3 N-terminal tail exits the nucleosome between the gyres of DNA. This suggests a mechanism by which TAZ2 might bind to the DNA proximal to this region, orienting the CBP core and bringing it into stable contact with the tail of histone H3 to allow efficient acetylation of H3K27.

## 5.6 Summary and discussion

The TAZ2 domain of CBP is required for efficient acetylation of the physiologically important residue H3K27 in chromatin substrates whilst reducing promiscuity towards other substrate residues. Work in this chapter attempted to address the mechanism by which TAZ2 directs the activity of CBP. Sequence and structural analysis revealed that the surface of TAZ2 has overall positive charge, and DNA-binding experiments show that TAZ2 interacts with DNA in a sequence-independent manner. This DNA binding activity mediates interactions with nucleosomes and determines how CBP acetylates nucleosomal histone proteins.

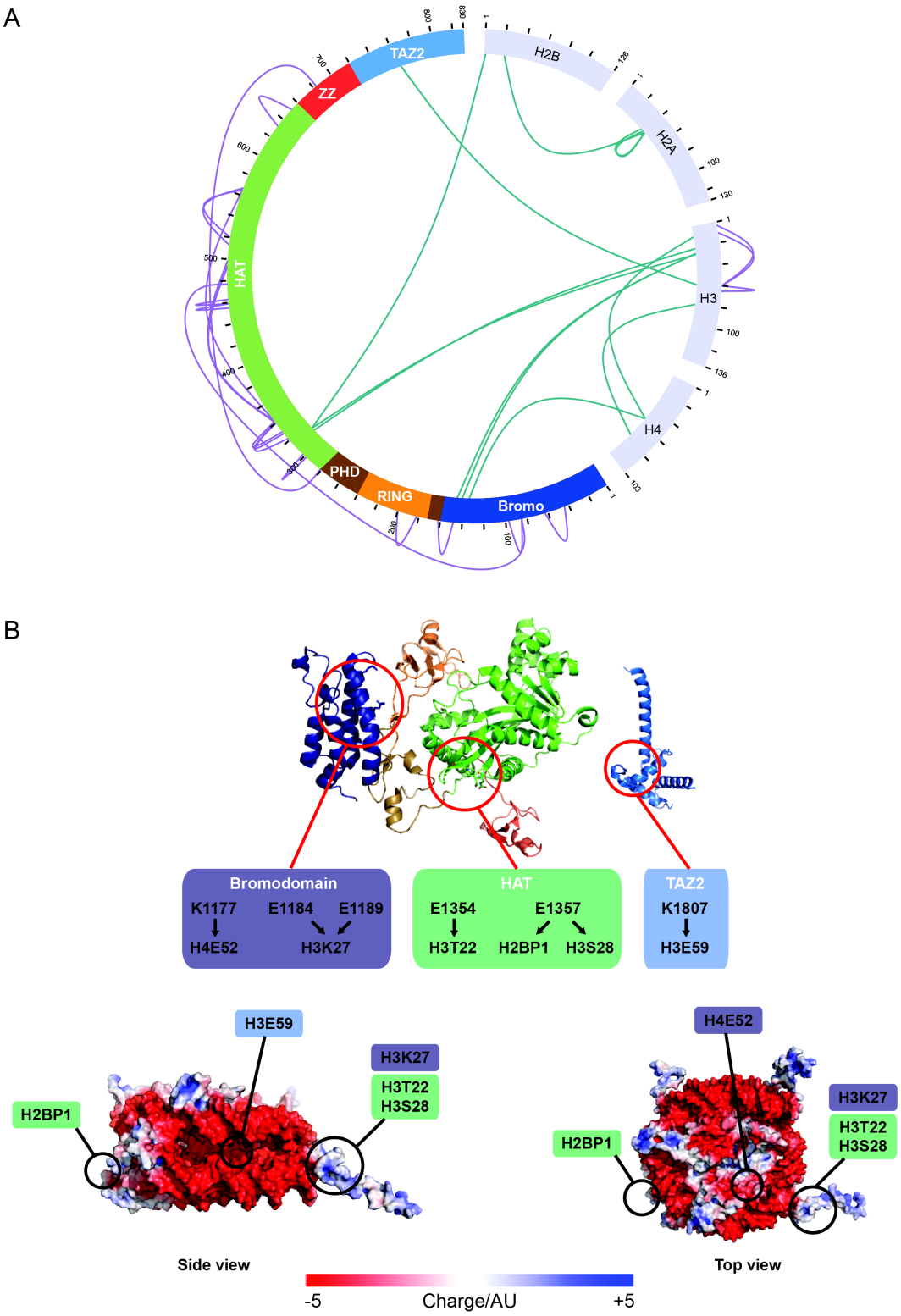



**Fig. 5.5: CBP interacts with the nucleosome via the enzymatic core and the TAZ2 domain.**
**(A)** Crosslinking mass spectrometry analysis of CZT in complex with 147 bp NCP. Green lines within the circle represent inter-protein crosslinks and purple lines outside the circle represent intra-protein crosslinks.
**(B)** Top: Cartoon representation of structures of CZT domains shaded as follows: bromodomain, dark blue; PHD finger, brown; RING finger, orange; HAT domain, green; ZZ domain, red; TAZ2 domain, light blue. (PDB IDs of structures: p300 core, 4BHW; CBP CZ, 5U7G; TAZ2, 3IO2.)

Previous work has suggested that CBP/p300 might contain a DNA binding activity (Rikitake and Moran, 1992; Song et al., 2002; Suhara et al., 2002), although the evidence was somewhat contradictory (Manning et al., 2001). Moreover, it was unclear whether any observed DNA binding was dependent on the presence of other, sequence-specific DNA-binding proteins (Rikitake and Moran, 1992; Suhara et al., 2002). In addition, problematic biochemical analysis meant that the observed DNA binding could be dependent on denaturation of the protein (Song et al., 2002), either through generation of mutations that would be expected to perturb the structural integrity of domains, or though purification of zinc-rich proteins in the presence of metal ion-chelating agents such as EDTA and EGTA, which has previously been shown to lead to non-physiological interactions of CBP/p300 (Matt et al., 2004; Ortega et al., 2018). The present work attempted to avoid these difficulties to show unambiguously that TAZ2 binds directly to DNA in a sequence-independent manner, suggesting that this activity is important for the previously observed role for TAZ2 in regulating acetylation of nucleosome substrates (Kraus et al., 1999).

These results provide a mechanistic explanation as to why TAZ2 is required for acetylation of H3K27 specifically in the context of chromatin, but not in other substrates such as histone octamers, suggesting that TAZ2 is required to overcome an inherent suppression of acetylation by the additional DNA present in nucleosome arrays. This further suggests that this key physiological substrate of CBP/p300 is rendered inaccessible to the enzyme in chromatin, but not in non-physiological substrates devoid of DNA, such as histone octamers, or with limited DNA, such as NCPs.

One reason H3K27 might be less accessible in the context of chromatin is its proximity to nucleosomal DNA, which could, for steric or electrostatic reasons, prevent the H3K27 residue from efficiently entering the catalytic site of the HAT

---

**(Fig. 5.5 cont.)**
Middle: CBP residues involved in inter-protein crosslinks, separated by domain, with arrows to corresponding crosslinked histone residues.
Bottom: Surface representation of NCP structure coloured by charge, showing side view (left) or top view (right). Circles indicate the positions of histone residues involved in crosslinks to CBP, with the domain of CBP to which a residue is crosslinked indicated by the colour of the shaded box: bromodomain, dark blue; HAT domain, green; TAZ2 domain, light blue. (PDB ID: 1KX5.)

domain. This possibility is supported by the observation that the N-terminal tail of histone H2B, which also emerges from the nucleosome core between the gyres of DNA and stays in close proximity to DNA, is similarly dependent on TAZ2 for efficient acetylation specifically in the context of nucleosome arrays. By contrast, acetylation of H3K9, which is also present on the H3 tail but located distal to the DNA compared to H3K27, is less dependent on TAZ2. Moreover, acetylation of H4K5, present on the H4 tail that exits the nucleosome from the upper surface of the nucleosome and away from DNA, is not dependent on TAZ2 for acetylation even in chromatin substrates. The crosslinking mass spectrometry results suggest that TAZ2 overcomes this suppressive effect on H3K27 and H2B acetylation and mediates substrate specificity by facilitating binding of CBP to DNA at the lateral surface of the nucleosome and directing acetylation towards proximal histone tail residues.

Such a key role for sequence non-specific DNA binding in chromatin interactions is not unique to CBP/p300, but appears to be a common mechanism in chromatin-binding proteins (reviewed in Afek and Lukatsky, 2012). Numerous examples of sequence-independent DNA binding have been identified in chromatin modifying proteins and complexes, encompassing both active and repressive histone modifying enzymes and nucleosome remodellers, including Polycomb repressive complexes 1 and 2 (PRC1 and PRC2), the H3K9 methyltransferase SUV39h1, the H2B ubiquitylase complex DOT1L, the histone acetyltrasferase MOZ, the H3K4 methyltransferase complex MLL1 and the SWI/SNF chromatin remodelling complex BAF (Bentley et al., 2011; Holbert et al., 2007; Morrison et al., 2017; Park et al., 2019; Shirai et al., 2017; Wang et al., 2017b; Worden et al., 2019). This suggests that sequence-independent DNA binding is a major biological mechanism for generating interactions with chromatin.

An important role for the TAZ2 domain in determining CBP/p300 chromatin binding is consistent with previous findings *in vivo* that deletion of a region encompassing TAZ2 leads to widespread re-localisation of p300 in the nucleus and to increased acetylation of non-histone targets (Ortega et al., 2018). It may further explain the importance of TAZ2 to CBP/p300 function, given that TAZ2 was found to be essential in CRISPR-Cas9 domain knockout screens (He et al.,

2019), and that mutations that affect the structure of the domain cause Rubinstein-Taybi syndrome (RSTS) with symptoms comparable in severity to patients harbouring mutations that disrupt CBP/p300 catalytic activity (Menke et al., 2016, 2018).

Together, these results suggest that TAZ2 is conserved as a core part of the catalytic machinery of CBP/p300, and that sequence-independent DNA binding to nucleosomes at target sites could play a key role in directing accurate H3K27ac placement for the proper regulation of gene expression.

# 6. The CBP TAZ2 domain binds DNA to direct specific H3K27ac *in vivo and in vitro*

The TAZ2 domain of CBP directs substrate specificity in chromatin substrates, binds to DNA, and is required for CBP interactions with nucleosomes (see Chapters 4 and 5). However, it remains possible that the role of TAZ2 in determining acetylation specificity *in vitro* is independent of its DNA binding activity, and moreover it is unclear whether these functions are important for CBP/p300 function *in vivo*.

In cells, CBP/p300 is thought to be recruited to target loci through combinatorial interactions with the activation domains of multiple transcription factors bound simultaneously to gene regulatory elements (Goodman and Smolik, 2000). TAZ2-mediated DNA binding would then be expected to direct CBP/p300 activity towards substrate residues in nearby nucleosomes, resulting in the establishment of histone acetylation and gene transcription. However, experiments in which the p300 enzymatic core, which lacks the TAZ2 domain, is tethered to a locus *de novo* via a catalytically inactive Cas9 (dCas9) protein show that this minimal region of p300 is sufficient to generate H3K27ac and to activate transcription (Hilton et al., 2015). This suggests that when CBP/p300 is targeted to a gene regulatory element in a manner resembling transcription factor-mediated recruitment, the TAZ2 domain is not required for CBP/p300 function.

Nevertheless, the importance of CBP/p300 in stabilising transcription factors on DNA has been shown *in vitro* (Suhara et al., 2002), whilst deletion of a region of p300 that includes the TAZ2 domain leads to relocalisation of the protein within the nucleus (Ortega et al., 2018). This suggests that this region of CBP/p300 is important for interactions with chromatin even when other domains could facilitate binding at target sites through interactions with sequence-specific DNA-binding proteins. This implies that the activity of CBP/p300 catalytic core alone is not sufficient to mediate the full function of these enzymes *in vivo*. However, whilst previous work has suggested that the bromodomain and ZZ

domain of CBP/p300 are important in generating histone acetylation in cells (Zhang et al., 2018), it is unclear whether chromatin interactions mediated by TAZ2 are also important for histone acetylation *in vivo*.

To test whether TAZ2-mediated DNA binding is required for histone acetylation specificity both *in vitro* and *in vivo*, sequence analysis together with structural modelling was used to identify mutations that could specifically reduce DNA binding activity. These proteins were then expressed in cells and targeted to a locus *de novo* to determine what role TAZ2 DNA binding has in regulating histone acetylation and gene expression *in vivo*.

## 6.1 Conserved basic residues are required for TAZ2 DNA binding

To identify residues within the TAZ2 domain that might be involved in DNA binding, the TAZ2 protein sequence was aligned to 150 different TAZ2 sequences with a range of identity from 35% to 95% using the ConSurf algorithm (Ashkenazy et al., 2016). The level of conservation of individual residues was then mapped onto the structure of the TAZ2 domain to identify conserved surfaces that could be involved in interactions with DNA (Fig. 6.1B). Inspection of this structure identified three amino acids with positively charged side chains that are highly conserved even amongst evolutionarily distant species and which would not be expected to be important for the structural integrity of the domain (Fig. 6.1A, B). These three residues (R1769, K1832 and K1850, numbered according to mouse CBP) were then selected to drive modelling of a TAZ2-DNA interaction using the program HADDOCK (High Ambiguity-Driven Docking), which uses *ab initio* thermodynamic modelling together with user-defined interaction constraints to generate a structural model for the interaction (Van Zundert et al., 2016). This model suggests that these amino acids could indeed mediate an interaction with DNA (Fig. 6.1B). Moreover, the model fits with previous observations that TAZ2 interacts with DNA independent of sequence (see Chapter 5), as these residues form ionic interactions with the negatively charged phosphate backbone of DNA without interrogating the DNA sequence through interactions with nucleotide bases.

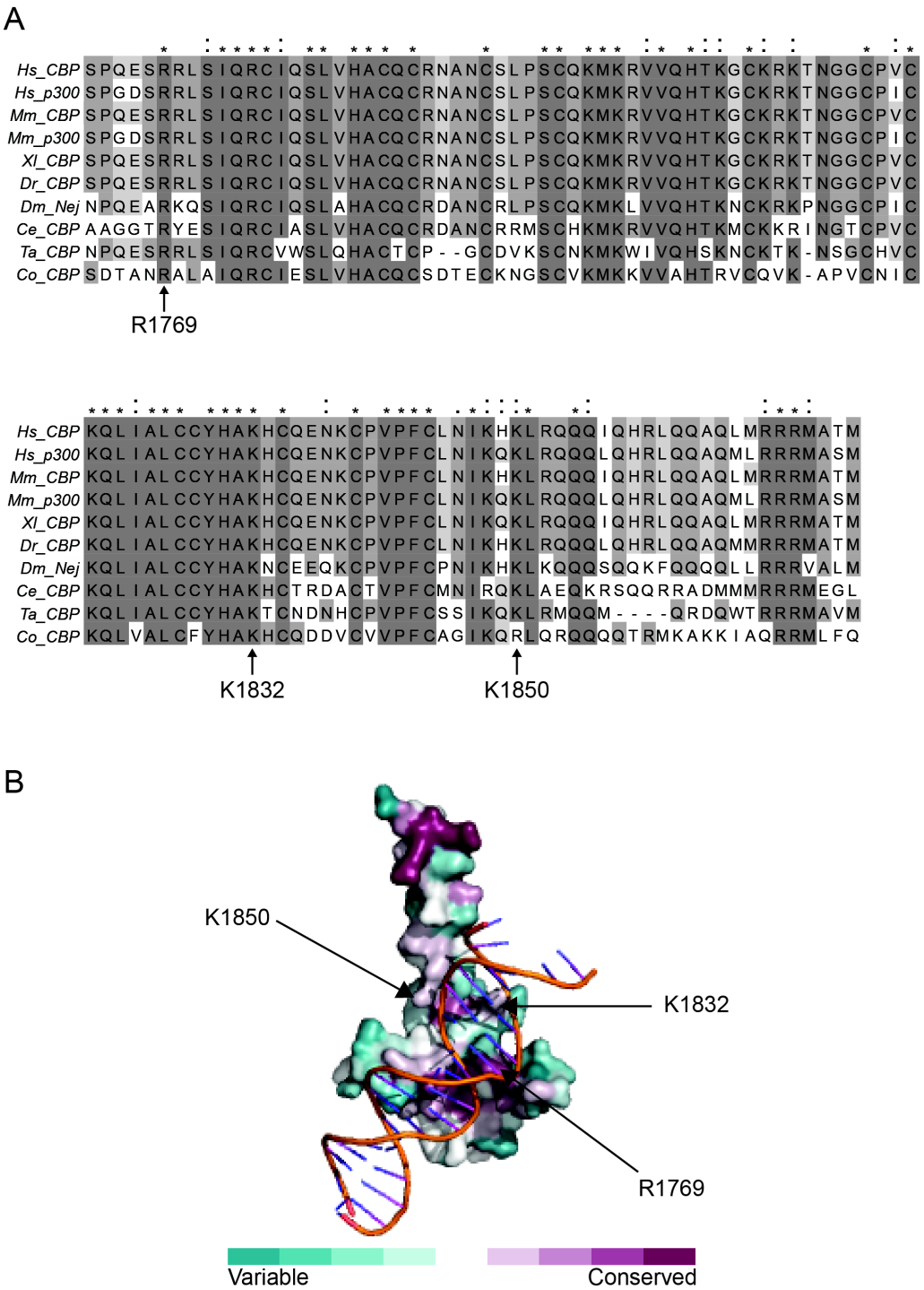



**Fig. 6.1: CBP TAZ2 contains highly conserved basic residues that could mediate DNA-binding.**

**(A)** Alignment of CBP and p300 protein sequences in humans (*Hs*), mouse (*Mm*), *Xenopus laevis* (*Xl*), *Danio rerio* (*Dr*), *Drosophila melanogaster* (*Dm*), *Caenorhabditis elegans* (*Ce*), *Trichoplax adhaerens* (*Ta*) and *Capsaspora owczarzarki* (*Co*); greyscale shading shows the level of conservation, asterisks (*) indicate identical residues, colons (:) indicate similar residues and arrows beneath sequence indicate residues that were mutated in this study.

**(B)** Surface depiction of human p300 TAZ2 structure (PDB ID: 3IO2) coloured according to level of conservation of individual amino acids as determined by ConSurf and modelled to bind a DNA molecule using HADDOCK. Arrows indicate the positions of residues that were mutated in this study.

To test whether these three amino acids could be involved in binding DNA, all three residues were mutated to glutamate to reverse the charge of their side chains, and the resulting mutated TAZ2 domain probed for DNA binding activity. The intact TAZ2 domain ($TAZ2^{wt)}$ and TAZ2 carrying the mutations R1769E/K1832E/K1850E ($TAZ2^{mut}$) were then expressed and purified from bacteria (Fig. 6.2A). DNA binding activity was first tested using a DNA pulldown experiment (Fig. 6.2B). This shows that whilst $TAZ2^{wt}$ binds to streptavidin beads coated with biotinylated 147 bp DNA, binding of $TAZ2^{mut}$ is almost undetectable. Furthermore, EMSA experiments using a 147 bp DNA probe showed that binding of $TAZ2^{mut}$ to DNA, although not completely abolished, is severely reduced compared to $TAZ2^{wt}$ (Fig. 6.2C), and that binding of $TAZ2^{mut}$ to 147 bp nucleosome core particles (NCPs) is greatly diminished (Fig. 6.2D).

Together these results suggest that TAZ2 contains a DNA-binding surface that includes at least three positively charged amino acids that bind DNA independent of sequence, and that mutation of these residues to reverse their charge compromises the capacity of TAZ2 to bind to DNA.

## 6.2 TAZ2 DNA binding mediates histone acetylation specificity *in vitro*

To test whether the observed requirement for the TAZ2 domain for acetylation specificity towards H3K27 is dependent on the DNA binding activity of the domain, the R1769E/K1832E/K1850E mutations that reduce DNA binding were introduced into the CBP core-ZZ-TAZ2 (CZT) construct to generate an enzyme with compromised DNA binding ($CZT^{mut}$). $CZT^{mut}$ was expressed and purified from Sf9 cells (Fig. 6.3A) and its activity was tested towards nucleosome arrays in $^{3}$H-labelled histone acetyltransferase (HAT) assays alongside CBP core, core-ZZ (CZ) and intact CZT enzyme ($CZT^{wt}$) (Fig. 6.3B). Consistent with previous results (see Chapter 4), inclusion of an intact TAZ2 domain resulted in greater overall activity towards nucleosomes and a specifically greater activity towards histone H3 than was observed with CBP core or CZ. However, mutation of TAZ2 led to a decrease in overall activity, with the greatest loss of activity

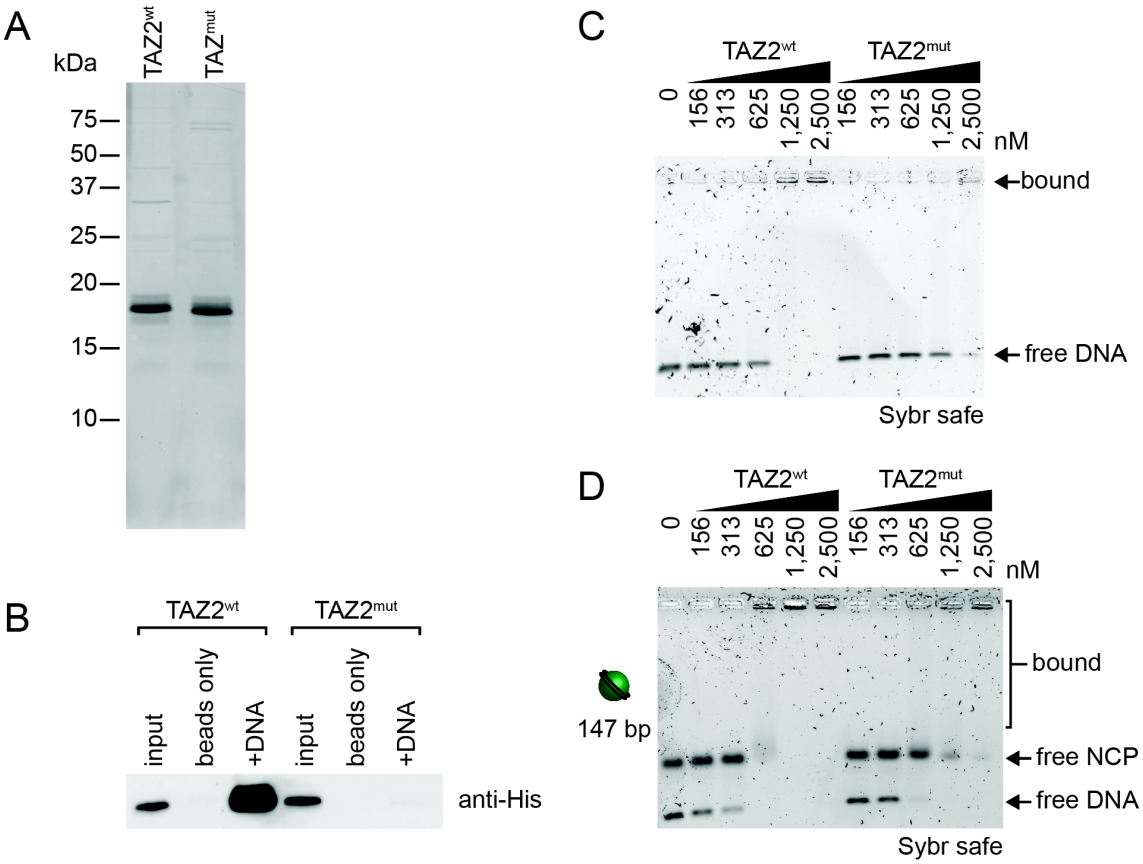



**Fig. 6.2: TAZ2 mutations greatly reduce DNA-binding.**
**(A)** CBP TAZ2$^{wt}$ and TAZ2$^{mut}$ domains were expressed in bacteria, purified via a 6xHis tag and analysed by SDS-PAGE followed by Coomassie staining.
**(B)** Pull down assay with TAZ2$^{wt}$ or TAZ2$^{mut}$ domain and either streptavidin beads or streptavidin beads bound by biotinylated DNA.
**(C)** Binding reactions using increasing concentrations of TAZ2$^{wt}$ or TAZ2$^{mut}$ domain with 147 bp DNA analysed by EMSA.
**(D)** Binding reactions using increasing concentrations of TAZ2$^{wt}$ or TAZ2$^{mut}$ domain with 147 bp NCP analysed by EMSA.

seen towards histone H3. This suggests that TAZ2-mediated DNA binding is indeed required for CBP substrate specificity.

To further examine this effect of TAZ2 DNA binding on substrate specificity, unlabelled HAT assays were carried out using CZ, CZT$^{wt}$ and CZT$^{mut}$ enzymes in time course experiments, which were then analysed by western blot for specific histone modifications (Fig. 6.3C, D). These experiments confirmed that CZT$^{wt}$ has specifically greater activity towards H3K27 and H2BK5 than CZ but a lower rate of acetylation towards H3K9 and H4K5. CZT$^{mut}$ showed a lower rate of acetylation than CZT$^{wt}$ towards all tested residues, but the greatest loss of acetylation is towards H3K27 and H2BK5, the residues that are most dependent on the presence of the TAZ2 domain.

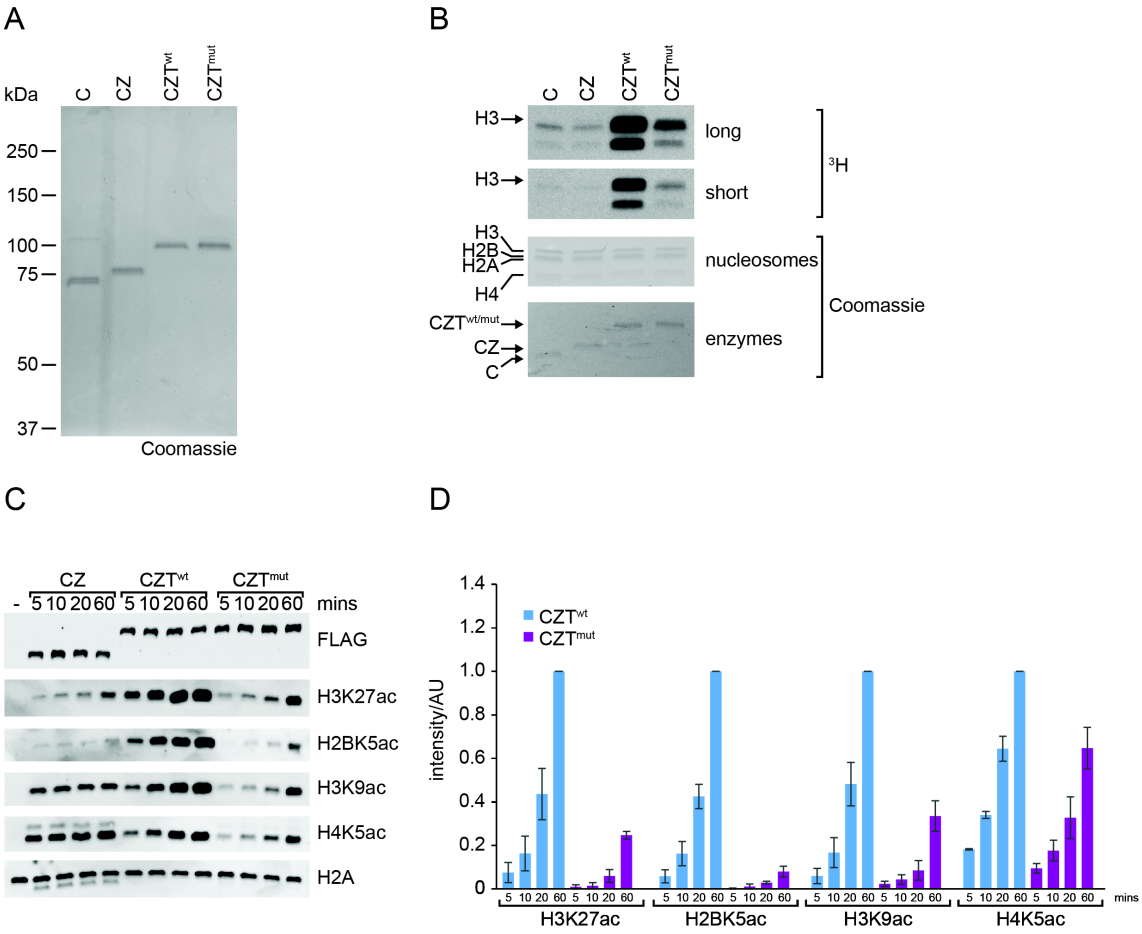



**Fig. 6.3: TAZ2 DNA binding directs H2K27ac in chromatin.**
**(A)** CZ, CZT$^{wt}$ and CZT$^{mut}$ proteins were expressed in Sf9 cells, purified via the FLAG tag and analysed by SDS-PAGE followed by Coomassie staining.
**(B)** HAT assays showing reactions with CZ, CZT$^{wt}$ and CZT$^{mut}$ enzymes with nucleosome array. Long and short autoradiograph exposures, and Coomassie staining to show enzymes and histone proteins are shown.
**(C)** Western blot analysis of unlabelled HAT assay timecourse experiments using no enzyme (-), CZ, CZT$^{wt}$ or CZT$^{mut}$ with nucleosome array, for reaction times of 5, 10, 20 or 60 mins, analysed using antibodies against FLAG, H3K27ac, H2BK5ac, H3K9ac, H4K5ac or total H2A.
**(D)** Quantification of HAT assay shown in (D). Signal for each antibody is normalized to CZT reaction at 60 mins and represents the mean of three independent experiments, with error bars indicating SEM.

These results suggest that the DNA binding activity of the TAZ2 domain of CBP contributes to overall levels of histone acetylation, but that DNA binding is specifically required to generate high levels of acetylation towards H3K27 and the N-terminal tail of H2B, the most important *in vivo* targets of CBP/p300.

## 6.3 TAZ2 drives CBP association with chromatin *in vivo*

To test whether TAZ2 DNA binding is important to CBP function *in vivo* as well as *in vitro*, CBP core, CZ, $CZT^{wt}$ and $CZT^{mut}$ proteins were expressed transiently in mouse embryonic stem (ES) cells, with the four proteins expressed to similar levels in whole cell extracts (Fig. 6.4A). To examine whether TAZ2 mediates interactions with chromatin, nuclei from transfected cells were isolated and nuclear proteins extracted with increasing salt concentrations (Fig. 6.4B). In this experiment, proteins that are weakly bound to chromatin are likely to be extracted at lower salt concentrations, whereas proteins that are strongly bound to chromatin will be solubilised only at higher salt concentrations. The results show that CBP core is primarily extracted in the 150 mM and 300 mM salt fractions, and CZ peaks in the 300 mM salt fraction, suggesting that the previously reported interaction of the ZZ domain with histone H3 contributes to CBP chromatin binding (Zhang et al., 2018). $CZT^{wt}$ protein, by contrast, elutes primarily in the 300 mM and 450 mM salt fractions, suggesting that addition of the TAZ2 domain leads to stronger interactions of CBP with chromatin *in vivo*, consistent with the results observed *in vitro*. Surprisingly, mutation of the DNA binding domain in $CZT^{mut}$ does not affect the nuclear fractionation profile of the protein. This is most likely because $CZT^{mut}$ retains some DNA binding activity and any subtle changes in occupancy cannot be detected in this assay examining global chromatin binding.

To test whether expression of these CBP truncations and mutants leads to changes in histone acetylation, acid extracted histones from transfected cells were probed for histone acetylation marks by western blot (Fig. 6.4C). The results suggest that there are no overall changes in the levels of H3K27ac, H2BK5ac, H3K9ac or H4K5ac between transfected and untransfected cells, or between the different CBP constructs. Although no global changes in histone

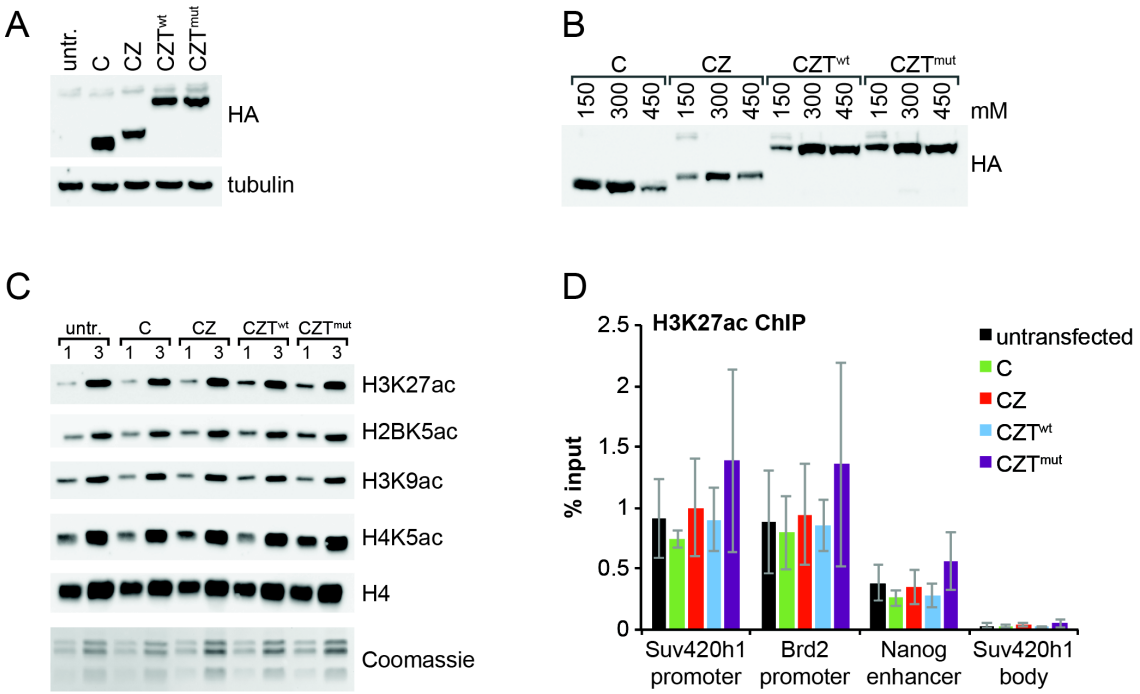



**Fig. 6.4: TAZ2 drives CBP binding to chromatin *in vivo*.**
**(A)** ES cells were transiently transfected for 48 h with CBP core (C), CZ, CZT$^{wt}$ or CZT$^{mut}$, and whole cell extract from transfected and untransfected cells was analysed by western blot for HA tag and tubulin as loading control.
**(B)** Nuclei were isolated from cells transiently expressing CBP core, CZ CZT$^{wt}$ or CZT$^{mut}$, washed sequentially with nuclear extract buffer containing the indicated salt concentrations, and nuclear extracts were analysed by western blot for HA tag.
**(C)** Histone acid extracts were prepared from untransfected and transiently transfected cells and analysed by western blot for H3K27ac, H2BK5ac, H3K9ac, H4K5ac and histone H4, and by Coomassie staining.
**(D)** ChIP for H3K27ac was carried out in untransfected and transiently transfected cells and analysed by qPCR. Signal represents the mean of three independent transfections and error bars indicate SEM.

acetylation were observed, to determine whether there were changes in H3K27ac at the level of individual loci, chromatin immunoprecipitation (ChIP) was carried out followed by quantitative PCR (qPCR) at known CBP/p300-bound gene regulatory elements (Fig. 6.4D). This shows that, consistent with the observations for global histone acetylation levels, there are no changes in H3K27ac at the level of individual genes.

These findings that expression of CBP proteins do not lead to changes in histone acetylation are in contrast to recent work suggesting that expression of a similar p300 CZT construct leads to widespread changes in histone acetylation in the human lung cancer cell line H1299, both at a global level and

at the level of individual genes (Zhang et al., 2018). One explanation for this difference is that the previous study made use of lentiviral transduction followed by antibiotic selection for 4-10 days to generate stably overexpressing cell lines. This means that the efficiency of transfection and the level of proteins expressed is likely higher than in the present study, which could account for the observed differences in acetylation.

However, examination of data from ChIP followed by massively parallel sequencing (ChIP-seq) carried out in (Zhang et al., 2018) shows that only 679 peaks of p300 CZT protein were identified genome-wide, whereas more than 25,000 peaks of H3K27ac have previously been mapped in H1299 cells (Mi et al., 2017), suggesting that p300 CZT binding is not taking place at the vast majority of endogenous CBP/p300 binding sites. Moreover, closer inspection of the ChIP-seq data shows that sites that were identified as p300 CZT binding sites, and validated by ChIP-qPCR, closely mirror peaks in input signal and fluctuations in background signal that are also observed in the vector only control ChIP (see Appendic, Fig. S2). In addition, background signal in the p300 CZT ChIP data is greatly increased overall compared to the vector only control, so that p300 CZT peaks could in part be explained by differences in background levels and fluctuations in input across the genome. Moreover, the increases in H3K27ac that are also detected by ChIP-qPCR at these sites are not reflected in the ChIP-seq data, in which H3K27ac remains at background levels. It is unclear whether the p300 CZT binding seen across the genome is increased background due to variation between samples, or whether it reflects true binding of p300 CZT throughout the genome. However, the latter would suggest that the protein is overexpressed to a non-physiological level, as such coating of the genome is not observed for endogenous p300 protein (Yue et al., 2014), and therefore that the observed changes may not represent true CBP/p300 function *in vivo*.

The results of the present study, therefore, suggest that TAZ2 is sufficient to mediate higher affinity binding of CBP to chromatin *in vivo*, but does not induce the stable, targeted binding of CBP required to generate changes in histone acetylation. This is apparently in conflict with previous work (Zhang et al., 2018), although detailed examination of this report shows that the observed

changes in histone acetylation are likely to represent either differences in background levels between samples or effects that result from expression of non-physiological levels of protein. Indeed, given that the CZT construct used here lacks major protein-protein interaction domains such as TAZ1, KIX and NRID that are important for interaction with transcription factors (see Chapter 4 and Fig. 4.4), it is unsurprising that CZT is not bound sufficiently stably at target sites for changes in histone acetylation to be observed. This supports a model in which the major targeting mechanism for CBP/p300 *in vivo* is through interaction with transcription factors, and that TAZ2 plays a role in stabilising the protein on chromatin and directing histone acetylation following recruitment.

## 6.4 TAZ2 stabilises CBP bound to chromatin and directs substrate specificity to regulate gene expression

To test this model of transcription factor-directed CBP/p300 function, a system was established in which CBP enzymes could be recruited *de novo* to a target gene as a dCas9 fusion protein, mimicking the recruitment of CBP at regulatory elements by transcription factors (Fig. 6.5A). In this system, dCas9 was fused to CBP core, CZ, $CZT^{wt}$, $CZT^{mut}$ or to a CZT construct carrying a mutation (D1436Y) that renders the enzyme catalytically inactive ($CZT^{ci}$) as a negative control, or to the VP160 transcriptional activator as a positive control (Cheng et al., 2013). These constructs were transfected transiently into ES cells together with four plasmids expressing different single guide RNAs (sgRNAs) to target the dCas9 fusion proteins to the promoter of the *Ascl1* gene, and the effect of tethering different fusion proteins on histone acetylation and gene expression was measured by ChIP and reverse transcriptase-qPCR (RT-qPCR).

Following expression for 48 h, the six dCas9 fusion proteins were expressed at similar levels, as shown by western blot of whole cell extracts (Fig. 6.5B). To test whether the proteins were bound at similar levels to the *Ascl1* target locus, ChIP was carried out using an antibody against an HA tag present at the N-terminus of the dCas9 fusion proteins, and binding was measured by qPCR using primers located near the site of dCas9 targeting (-260 bp compared to the transcription start site of *Ascl1*) or at two regions 300-400 bp either side of the

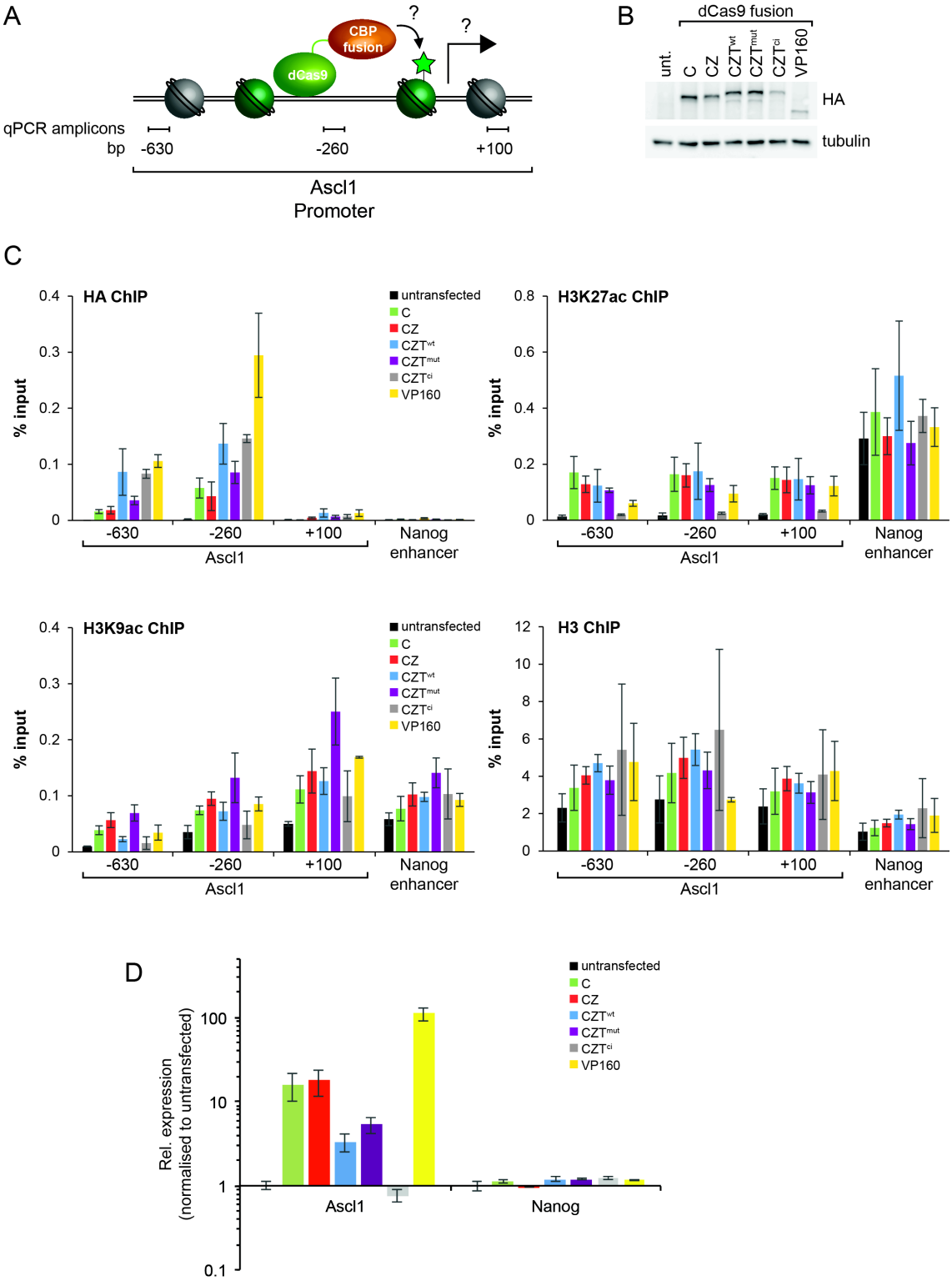



**Fig. 6.5: TAZ2 DNA binding stabilises CBP on chromatin and promotes specific H3K27ac *in vivo* to regulate gene expression.**

**(A)** Schematic showing targeting of CBP fusion proteins to the promoter of the *Ascl1* gene by dCas9, with potential outcomes such as histone acetylation (green star) and gene transcription. Positions of bars represent the locations of amplicons used for ChIP-qPCR relative to the transcription start site of *Ascl1*.

**(B)** ES cells were transfected for 48 h with constructs expressing four gRNAs targeting Ascl1 and dCas9 fused to CBP core (C), CZ CZT$^{wt}$, CZT$^{mut}$ (CZT with DNA-binding mutations), CZT$^{ci}$ (CZT with catalytically inactivating mutations) or VP160. Whole cell

target site (-630 bp and +100 bp) (Fig. 6.5A, C). Whilst background HA signal in untransfected cells and at the endogenous *Nanog* enhancer control region was very low, there was a clear increase in binding of all fusion proteins at the -630 bp and -260 bp regions, with little signal detected at +100 bp. The positive control VP160 fusion is the most highly bound protein, possibly as a result of stabilising interactions with the transcriptional machinery (Hall and Struhl, 2002). Importantly, although binding of all fusion proteins was robustly detected, the greatest levels of binding by CBP fusion proteins were achieved by $CZT^{wt}$ and $CZT^{ci}$, with lower levels of binding detected either in the absence of the entire TAZ2 domain, for CBP core and CZ, or in the absence of TAZ2 DNA binding activity for $CZT^{mut}$. This shows that when a sequence-directed DNA binding protein tethers CBP proteins to a target site, TAZ2 plays a role in stabilising CBP on chromatin in a manner that depends on the DNA binding activity of TAZ2 but is independent of the catalytic activity of the enzyme.

To test whether these differences in fusion protein binding result in differences in histone acetylation, ChIP was carried out using antibodies against H3K27ac, a physiological product of CBP/p300, and against H3K9ac, a modification that is not dependent on CBP/p300 *in vivo* (Jin et al., 2011; Weinert et al., 2018) (Fig. 6.5C). These experiments showed that dCas9-tethered CBP core, CZ, $CZT^{wt}$, $CZT^{mut}$ and VP160 were all able to generate clear increases in H3K27ac compared to the background levels that are present in untransfected cells, leading to H3K27ac at levels only slightly lower than those found at the

---

**(Fig. 6.5 cont.)** extracts from untransfected and transiently transfected cells were analysed by western blot for HA tag and tubulin as loading control.
**(C)** ChIP qPCR analysis was carried out in untransfected and transiently transfected cells, using antibodies against HA to measure binding of the dCas9 fusion protein, H3K27ac, H3K9ac and histone H3. qPCR was carried out at three regions within the Ascl1 promoter and at the active Nanog enhancer as a positive control region. Signal represents the mean of three independent transfections and error bars indicate SEM.
**(D)** Gene expression analysis was performed by RT-qPCR in untransfected and transiently transfected cells, using primers for Ascl1 and for Nanog as a gene that should remain unchanged. Gene expression was measured relative to Gapdh and normalised to the untransfected cells, and is show on a logarithmic scale. Signal represents the mean of three independent transfections and error bars indicate SEM.

endogenous *Nanog* enhancer. Importantly, these changes in H3K27acgenerated by CBP enzymes do not simply reflect changes in overall nucleosome occupancy, as measured by total histone H3 ChIP, and are dependent on the catalytic activity of the tethered proteins, as $CZT^{ci}$ is unable to generate H3K27ac. These results show that the TAZ2 domain of CBP is not required for acetylation of H3K27 *in vivo* when CBP is stably tethered to a target sequence. They further show that the level of H3K27ac generated at the target site is not directly related to the level of enzyme bound, as H3K27ac levels are similar between the different active CBP fusion proteins despite $CZT^{wt}$ binding at higher levels.

However, by contrast with H3K27ac, differences were observed in the levels of H3K9ac generated by the different CBP enzymes. Cells expressing $CZT^{wt}$ show H3K9ac at the same levels that are observed in cells expressing $CZT^{ci}$. By contrast, CBP core, CZ and $CZT^{mut}$ generate higher levels of H3K9ac, especially at the -630 bp and -260 bp regions of *Ascl1*. This suggests that all of the tethered CBP enzymes are able to generate H3K27ac when recruited to this region, but that the role of TAZ2 is to maintain the specificity of CBP-mediated acetylation *in vivo* by facilitating H3K27ac whilst preventing acetylation of off-target sites such as H3K9. Moreover, the greater levels of H3K9ac observed with $CZT^{mut}$ show that this maintenance of specificity is dependent on the DNA binding activity of TAZ2.

To determine whether these changes in histone acetylation in turn affect the regulation of transcription, RT-qPCR was carried out to examine gene expression changes at the *Ascl1* gene and the control *Nanog* gene (Fig. 6.5D). This analysis shows that the greatest change in *Ascl1* expression is in cells transfected with VP160 fusion protein, which shows a greater than 100-fold increase in expression compared to untransfected cells, presumably as a result of direct recruitment of components of the transcriptional machinery (Hall and Struhl, 2002). All CBP fusion constructs elicit an increase in transcription, with the exception of $CZT^{ci}$, showing that these gene expression changes are dependent on the acetyltransferase activities of the enzymes. Importantly, whilst targeting of CBP core and CZ led to an approximately 15-fold increase in transcription compared to untransfected cells, targeting of $CZT^{wt}$ only leads to a

three-fold increase in transcription. This suggests that the additional promiscuous acetylation of substrates other than H3K27 by CBP core and CZ, seen both *in vitro* and *in vivo*, leads to greater transcriptional activation than acetylation of H3K27 alone by $CZT^{wt}$. This interpretation is supported by the observation that $CZT^{mut}$, which lacks the DNA binding activity of $CZT^{wt}$ and generates higher levels of H3K9ac, clearly and reproducibly generates higher levels of transcription than $CZT^{wt}$, although not to the same level as CBP core and CZ. One explanation for the observation that $CZT^{mut}$ does not reach the same levels of transcriptional activation as CBP core and CZ could be that a degree of DNA binding activity is retained in $CZT^{mut}$. One possibility is that whilst this small amount of DNA binding does not limit the acetylation of H3K9, it could decrease the acetylation of other off-target substrates such as H4K5, the acetylation of which is reduced *in vitro* for $CZT^{mut}$ (Fig. 6.3C).

These results suggest that the major role of TAZ2 DNA binding *in vivo* is to allow CBP to acetylate target residues such as H3K27, whilst preventing promiscuous acetylation of residues such as H3K9. The function of this could be to limit the extent to which the inherently promiscuous CBP/p300 enzymes can acetylate histones at target sites *in vivo*, to prevent undesirable stochastic activation of gene expression and to allow CBP/p300-mediated histone acetylation to fine tune the expression of target genes.

## 6.5 Summary and discussion

The TAZ2 domain of CBP binds DNA and is required *in vitro* to acetylate H3K27 whilst limiting promiscuity towards other potential substrate sites. Work in this chapter used mutations that diminish TAZ2 DNA binding to show that the influence of TAZ2 on CBP substrate specificity is dependent on its interaction with DNA. Moreover, further work showed that TAZ2 stabilises CBP bound to chromatin *in vivo* and directs CBP substrate selection at target sites to limit transcriptional activation.

However, the dCas9-based tethering experiments presented here have limitations in the extent to which they model endogenous CBP/p300 recruitment by transcription factors. First, Cas9 proteins have greatly increased residency

time on chromatin compared to typical mammalian DNA binding proteins. For example, the glucocorticoid receptor (GR) and oestrogen receptor (ER), the latter of which can recruit CBP/p300 to target sites (Zwart et al., 2011), have a residency time on the scale of seconds (Presman et al., 2017; Swinstead et al., 2016). By contrast, Cas9 proteins have been measured to reside on chromatin for in excess of 3 h (Ma et al., 2016). Second, *in* vivo the levels of CBP/p300 protein are limiting and transcription factors compete for interaction with these coactivators (Hottiger et al., 1998). This means that CBP/p300 will interact dynamically with multiple DNA binding proteins, potentially cycling on and off chromatin rather than remaining stably bound. In the dCas9 system, however, CBP is fused directly to the sequence-specific recruitment module and is unable to be released from chromatin independently of dCas9 through interaction with other proteins.

These considerations suggest that, whilst a dCas9-based tethering system is a valuable tool to probe the function of CBP *in vivo*, it does not reflect true CBP/p300 function in cells entirely faithfully. Indeed, this may explain some discrepancies between the results of this experiment and the properties of CBP observed *in vitro*. For example, the static tethering of CBP core on chromatin may allow this protein to overcome its comparative lack of activity seen *in vitro*, facilitating a similar level of activity to that observed with the CZ construct, which would be expected to have greater activity than CBP core alone. This may also account for the lack of observable differences in H3K27ac between $CZT^{wt}$, which *in vitro* has far greater activity towards H3K27, and the constructs that lack an intact TAZ2 domain. The non-physiological tethering of CBP to the target site by dCas9 may, therefore, be sufficient to diminish the importance of interactions between TAZ2 and chromatin in directing H3K27ac.

Nevertheless, the results of this experiment point to an important role of the TAZ2 domain *in vivo* for stabilising CBP binding at target sites, and in both targeting CBP activity towards H3K27 and reducing its activity towards non-target histone lysine residues. These results further suggest that CBP/p300-mediated acetylation of H3K27 alone is not sufficient to generate high levels of transcriptional activity, but that transcription is enhanced by further acetylation of residues that are not targets by CBP/p300, such as H3K9. This implies that

other HAT enzymes, such as the H3K9 acetyltransferases GCN5 and PCAF (Feller et al., 2015; Gates et al., 2017; Jin et al., 2011), may be required for full activation of gene expression at endogenous target sites, consistent with the overlapping binding pattern of these proteins with CBP/p300 in mammalian cells (Krebs et al., 2011). One reason why this cooperation between different HATs to generate an active chromatin state might be important is that it could allow both for precise control of quantitative transcriptional responses to nuclear signalling and for robustness in switches of gene activity.

This model in which multiple HATs independently acetylate different histone lysine residues to integrate signalling pathways at gene regulatory elements suggests that maintaining the specificity of individual HATs towards their target residues would be highly important in preventing undesirable stochastic activation of gene expression by a single HAT enzyme. This importance may explain why the TAZ2 domain, which both directs H3K27ac and limits acetylation of other histone lysine residues, is so highly conserved amongst CBP/p300 proteins, and why TAZ2 mutations are associated with disease in humans (Menke et al., 2016, 2018). This would mark TAZ2 DNA binding as an important novel mode of regulation for CBP/p300 that facilitates precise control of gene expression.

# 7. Conclusions and implications for future work

Regulation of gene expression is a fundamental process for living organisms, allowing response to external signals and facilitating the development of highly specialised cell types in multicellular organisms. In eukaryotes, gene expression is controlled through binding of proteins to gene regulatory elements, primarily gene promoters and enhancers. One mechanism through which gene regulatory elements function is by generating a chromatin environment that modulates transcription from associated genes. However, the mechanisms by which the chromatin architecture is established at these elements, and how this leads to a functional transcriptional output, remain unclear. Work in this thesis attempted to address these questions, first by using proteomics approaches to generate an inventory of proteins that bind to regulatory regions, and second by adopting a candidate approach and probing how mechanistically the histone acetyltransferase CBP/p300 acetylates chromatin to regulate gene expression. The results of the proteomics work suggested future avenues to explore novel mechanisms by which gene expression is regulated, which were addressed in Chapter 3. This discussion will therefore focus exclusively on the results from Chapters 4, 5, and 6. The results of the CBP/p300 candidate approach in these chapters uncovered a novel DNA binding activity in CBP that regulates histone acetylation *in vitro* and *in vivo*. In this chapter, the main findings of this work will be summarized and the implications of these findings discussed in more detail, together with the questions that remain to be addressed in the future.

## 7.1 Sequence-independent DNA binding: a universal mechanism in chromatin interactions?

Work in Chapter 4 of this thesis showed that the TAZ2 domain of CBP is required for histone acetylation specificity towards H3K27 in nucleosome substrates. Further work in Chapters 5 and 6 showed that TAZ2-dependent

acetylation specificity is mediated by a sequence-independent DNA binding activity. This DNA binding activity functions to direct CBP specificity through two mechanisms. First, binding to DNA increases the efficiency of CBP interaction with nucleosomes, leading to increased overall acetylation of histone substrates. Second, TAZ2 DNA binding leads to specifically increased acetylation of H3K27 compared to alternative substrate lysine residues, potentially by binding to the lateral surface of the nucleosome to direct the CBP catalytic domain towards the H3K27 substrate.

A major focus of previous work was on the role of sequence-specific DNA binding proteins in the recruitment of CBP/p300 and other chromatin-modifying proteins to target loci. Indeed, transcription factor-mediated recruitment likely represents the major mode of CBP/p300 tethering at target loci, as shown by induction of CBP/p300 binding genome-wide by transcription factors such as the estrogen receptor (ER), and CBP-dependent induction of p53 target genes (Gu et al., 1997; Zwart et al., 2011). However, recent work has brought to light the importance of sequence-independent DNA binding in chromatin interactions (Bentley et al., 2011; Holbert et al., 2007; Morrison et al., 2017; Park et al., 2019; Shirai et al., 2017; Worden et al., 2019). A clear example of this is in the recently identified DNA binding activity of Polycomb repressive complex 2 (PRC2), which binds to DNA in a sequence-independent manner both through its core subunits and through the accessory subunit Pcl (PHF1 in mammals) (Choi et al., 2017; Wang et al., 2017b). This DNA binding is thought to stabilise the PRC2 complex on nucleosome substrates, increasing its residency time and allowing it to generate H3K27me3 at target sites. This is consistent with observations that loss of Pcl *in vivo* leads to reduced, but not abolished, PRC2 occupancy at target genes, and diminished H3K27me3 (Hunkapiller et al., 2012; Nekrasov et al., 2007; Walker et al., 2010). This suggests that sequence-independent DNA binding may represent a commonly employed mechanism for stabilising chromatin-modifying activities on chromatin, allowing proteins and complexes such as CBP/p300 and PRC2 to bind stably to a wide variety of target sequences following recruitment by other mechanisms.

Indeed, sequence-independent DNA binding appears to be employed not only by CBP/p300 but by all branches of the histone acetyltransferase (HAT) family.

In the MYST family of HATs, the HAT domain of MOZ is closely linked to a zinc finger that binds DNA (Holbert et al., 2007). In addition, MOZ and MORF possess a NEMM (N-terminal part of Enok, MOZ or MORF) domain that is rich in basic amino acids and contains a histone H1/H5-like fold that is likely to bind to DNA and is required for nuclear localization (Kitabayashi et al., 2002). Moreover, the double PHD finger (DPF) domain of MORF binds DNA independent of sequence, and this activity is important for interactions with nucleosomes (Klein et al., 2019). Although GCN5 family HATs have not been found to bind DNA directly, the ADA2 subunit that interacts with GCN5 as part of the SAGA and ATAC complexes contains a SANT domain with structural similarity to the DNA binding domain of c-Myb (Sun et al., 2018). MYST family members similarly interact with partner proteins that contain DNA binding domains, including the BRPF1, 2 and 3 subunits, which interact with the MOZ, MORF and HBO1 enzymes, and which bind DNA sequence-independently through their PHD-Zn knuckle-PHD (PZP) module (Klein et al., 2016; Lalonde et al., 2013; Liu et al., 2012).

The observation of sequence-independent DNA binding in such a wide variety of chromatin modifiers through structurally distinct domains suggests that evolution of DNA binding is highly favoured in chromatin interactions. A clear possibility is that DNA binding would increase the affinity of chromatin modifiers for their target sites on chromatin. Whilst such sequence-independent interactions would not direct recruitment to specific target sites in the genome, these interactions could cooperate with binding to transcription factors or histone modifications to generate multivalent, high affinity interactions with chromatin.

However, the importance of sequence-independent DNA binding activities has remained relatively unexplored. Multiple reports have shown that mutations that affect the structure of the TAZ2 domain of CBP, or would be expected to affect sequence-independent DNA binding by the PZP domain of BRPF1 and the BRG1 subunit of the BAF chromatin remodelling complex, are associated with disease (Menke et al., 2016, 2018; Morrison et al., 2017; Yan et al., 2017). This suggests that DNA binding is likely to represent an important function of these proteins. Moreover, work in this thesis showed that DNA binding by the TAZ2

domain of CBP stabilises the protein on chromatin in an artificial tethering system. Further work would test whether DNA binding by TAZ2 is also important for CBP binding at endogenous targets and whether loss of this activity would prevent efficient H3K27ac at target sites.

## 7.2 Sequence-independent DNA binding: a mechanism for substrate specificity?

Work in the first decade or so following identification of HAT enzymes suggested that HATs were relatively promiscuous enzymes, each acetylating many different histone lysine residues (reviewed in Kouzarides, 2007; Lee and Workman, 2007; Sterner and Berger, 2000). However, more recent work in both mammals and *Drosophila* has shown that, although individual HATs may acetylate multiple substrates *in vitro*, they have a relatively narrow substrate specificity *in vivo* (Feller et al., 2015; Jin et al., 2011; Weinert et al., 2018).

Work in this thesis showed that TAZ2 DNA binding is required for CBP to acetylate H3K27 specifically. This is in contrast to many other sequence-independent DNA binding domains, where there is evidence of a function in determining affinity for chromatin but not for enzymatic specificity. Surprisingly, however, the MYST family HAT enzyme HBO1 shows a similar dependency on a sequence-independent DNA binding factor for its substrate specificity (Lalonde et al., 2013).

HBO1 can interact in a mutually exclusive fashion with multiple scaffold proteins, including the two homologous proteins BRPF1 and JADE1 (Doyon et al., 2006; Ullah et al., 2008). The HBO1-BRPF1 complex can acetylate free histone H3 and H4 promiscuously *in vitro* (Lalonde et al., 2013). By contrast, in nucleosomal substrates HBO1-BRPF1 specifically acetylates H3, particularly H3K14 and H3K23, which are major physiological substrates of HBO1 (Feng et al., 2016; Lalonde et al., 2013; MacPherson et al., 2019; Yan et al., 2017). This H3 specificity depends on critical residues within the PZP sequence-independent DNA binding domain of BRPF1. In contrast to the HBO1-BRPF1 complex, the HBO1-JADE1 complex is largely unable to acetylate H3, and

instead specifically acetylates histone H4 in a nucleosomal context. Importantly, sequence analysis of the PZP region of JADE1 suggests that this protein lacks residues that are present in BRPF1 and are important for its DNA binding function (Yan et al., 2017). This suggests that, similarly to TAZ2-mediated DNA binding in CBP, BRPF1 DNA binding facilitates specific acetylation of target substrates.

One explanation for this apparent similarity between sequence-independent DNA binding in CBP and the HBO1-BRPF1 complex is that, given the proximity of lysine residues such as H3K23 and H3K27 to nucleosomal DNA, post-translational modification of H3 tail lysine residues proximal to the nucleosome core may require sequence-independent interaction with DNA for efficient acetylation. Mechanistically, the proximity of these H3 tail lysines to nucleosomal DNA may sterically or otherwise hinder access of the HAT domain to these residues, preventing their efficient acetylation. Binding to DNA may provide additional interaction energy, allowing interactions between the HAT domain and the H3K27 residue to become more favourable. A consequence of this hypothesis would be that chromatin-modifying enzymes including CBP/p300 and HBO1, and also complexes such as PRC2, might require sequence-independent DNA binding not only to increase residency time on chromatin but also to allow access to the substrate H3K27 residue, potentially by orienting the enzymes in a topology that favours interaction with H3K27.

Testing this hypothesis would benefit from structural studies to determine, potentially in atomic detail, how domains such as TAZ2 interact with DNA and drive nucleosome binding. The observation that the CBP core-ZZ-TAZ2 (CZT) construct readily forms a complex with the nucleosome core particle (NCP), and that this complex is amenable to crosslinking, suggests that a CZT-NCP complex structure could be suitable for elucidation by techniques such as electron cryo-microscopy. Structural detail of the interaction between TAZ2 and nucleosomal DNA could then permit the generation of a set of mutations in TAZ2 that would more completely ablate DNA binding than those designed here and could be used to study TAZ2 function more effectively *in vivo*.

## 7.3 HAT specificity: a mechanism for robust transcriptional regulation?

Multiple reports show that HAT enzymes possess relatively narrow specificity towards histone substrates *in vivo*, although the mechanisms for directing this specificity remain to be identified for many enzymes. GCN5 and its paralogue PCAF (GCN5/PCAF) primarily acetylate H3K9, MOZ/MORF acetylates H3K23 and MOF acetylates H4K16 (Feller et al., 2015; Jin et al., 2011; Simó-Riudalbas et al., 2015). However, these different acetylation marks are strongly associated with the same set of active promoters and enhancers genome-wide (Ernst et al., 2011; Karmodiya et al., 2012; Wang et al., 2009), and regions that are acetylated are typically co-occupied by multiple HATs, including CBP/p300, GCN5/PCAF and MOF (Krebs et al., 2011; Wang et al., 2009). The question therefore arises, why does the cell utilize multiple HATs targeted to the same region to acetylate multiple different target residues, rather than recruiting a single HAT to a given locus and allowing it to acetylate promiscuously to drive gene expression?

One explanation for the apparent redundancy in HAT recruitment is that it allows for robust regulation of gene expression. In such a model, gene expression would be regulated by the independent recruitment of multiple HATs to the same locus to generate acetylation of multiple different lysine residues, with maximal gene expression being achieved only upon recruitment of multiple HAT enzymes. Consistent with this model, previous work has shown that induction of acetylation through inhibition of histone deacetylases (HDACs) and stimulation of transcription through treatment with epidermal growth factor (EGF) generates H3 molecules that are acetylated not only at a single site but at multiple sites (Clayton et al., 2000). This suggests that multiple HATs are acting on the same H3 molecules during transcriptional induction. Further work examining rapidly inducible gene expression systems shows that following stimulation of the serum response factor (SRF) by treatment with TPA (tetradecanoylphorbol acetate), histone acetylation occurs in an ordered fashion. In this system, H3K9 is acetylated first, in conjunction with H3S10 phosphorylation (H3S10ph), and followed by H4K16ac, H3K27ac and H3K14ac, respectively, during the course of activation of immediate-early target genes

(Esnault et al., 2017). This ordered histone acetylation is then followed by the recruitment of the transcriptional machinery. Similarly, induction of IFN-β expression in HeLa cells by infection with Sendai virus generates an ordered temporal pattern of histone acetylation through the sequential recruitment of GCN5 and CBP HATs (Agalioti et al., 2000, 2002). These results suggest that gene activation follows the coordinated action of multiple HATs. In this model, HAT specificity would play a central role in facilitating robust activation of gene expression following appropriate signals but would prevent stochastic activation of gene expression through binding of a single HAT to a given gene regulatory element.

This model could be tested *in vivo* by generating a CBP enzyme with increased promiscuity through mutation of the TAZ2 domain. If this model of HAT specificity conferring robustness on gene regulation is correct, a less specific CBP would be expected to result in greater transcriptional noise and inappropriate gene activation genome-wide.

## 7.4 Histone acetylation: a mechanism for quantitative regulation of gene expression?

An additional, complementary explanation for targeting multiple HATs to the same loci would be to allow quantitative regulation of gene expression. Many genes are regulated in an essentially bistable manner, switching only between “on” and “off” states. A subset of genes, however, requires more quantitative regulation, in which the level of gene expression is proportional to the strength or duration of signalling, with regulatory elements controlling gene expression in the manner of rheostats. Examples of such loci are developmentally regulated genes that respond to morphogenic gradients in the development of animals, and immediate-early genes that respond to signalling pathways such as mitogen-activated protein kinase (MAPK) pathways (reviewed in Hazzalin and Mahadevan, 2002).

Morphogens are signalling molecules, usually proteins, examples of which include Bicoid in the specification of the anterior-posterior axis in the *Drosophila*

embryo, and Activin in *Xenopus* mesoderm formation (reviewed in Gurdon and Bourillot, 2001). Morphogens are released from a localised source and form a concentration gradient over a population of cells through diffusion. Cells within this population that are exposed to different morphogen concentrations show qualitatively different responses, giving rise to different cell fates. One way in which these different responses are achieved at the transcriptional level is through controlling the gene expression response to morphogen concentration with the use of thresholds, so that such genes operate as bistable switches. In this case, genes are switched to an “on” state when the concentration of morphogen exceeds a certain level, or remain in the “off” state when this threshold is not reached. However, a subset of genes can also respond to morphogen in a quantitative way, so that gene expression is proportional to the concentration of signalling molecule. Examples of this quantitative control of gene expression include control of expression of *Xgscd* by Activin and of *Xbra* by fibroblast growth factor (FGF) signalling in *Xenopus* development (Green et al., 1992; Gurdon et al., 1994), and of target genes by *Krüppel* in *Drosophila* (Sauer and Jäckle, 1991).

Similarly, activation of MAPK signalling pathways with increasing concentrations of lysophosphatidic acid (LPA) in Rat fibroblasts leads to proportional, quantitative increases in expression of target genes such as *c-Fos* and *c-Jun* (Cook et al., 1999). Moreover, the use of different stimuli, such as TPA or FGF, to activate MAPK pathways in mouse embryonic fibroblasts (MEFs) leads to different levels of activation of the same target genes (Hazzalin and Mahadevan, 2002). This suggests that regulation of these loci is not through a binary switch in expression between “on” and “off” states but can respond quantitatively to different signalling events.

The dCas9-based tethering experiments presented in this thesis suggest that histone acetylation could provide a mechanistic basis for this quantitative regulation of gene expression. In these experiments, targeting of different truncated forms of CBP as fusion proteins to the *Ascl1* promoter led to different levels of transcription of the target gene. One explanation for the observed differences in transcriptional activation is that the truncated enzymes show different levels of specificity, both in this experiment and *in vitro*. As a result, the

more promiscuous enzymes, such as CBP core, core-ZZ (CZ) and core-ZZ-TAZ2$^{mut}$ (CZT$^{mut}$), generated higher levels of transcription than the more specific CZT$^{wt}$ enzyme. This suggests that transcriptional activation is directly linked to overall levels of acetylation at the locus, rather than to acetylation of a particular residue, such as H3K27.

Similarly, in an artificial system in which p53 recruits CBP to a luciferase reporter under the control of the *Mdm-2* promoter, levels of upregulation were directly proportional to the level of CBP that p53 could recruit (Gu et al., 1997). This again suggests that binding of CBP and acetylation of target sites at a gene promoter does not lead to a binary switch from an "off" to an "on" state, but that intermediate levels of transcriptional activation exist for these genes. In this way, targeting of multiple HATs to a gene regulatory region to acetylate different residues could control the overall level of nucleosome acetylation, and therefore provide a mechanism to control the level of gene expression in a quantitative manner.

Similar observations of a link between acetylation levels and gene expression have been made in the endogenous MAPK system, in which levels of histone acetylation at target genes over time correspond closely to levels of gene expression (Edmunds et al., 2008; Hazzalin and Mahadevan, 2002). In this system, expression of immediate-early genes is rapidly induced in response to stimulation by factors such as epidermal growth factor (EGF), and the promoters of these genes are simultaneously acetylated. Importantly, within hours of stimulation, expression of target genes returns to normal levels, and histone acetylation levels are similarly reduced. This suggests that histone acetylation is a highly dynamic system, and that histone acetylation does not necessarily function to generate a stable shift in transcriptional status, but rather provides a mechanism to convey, in a quantitative manner, on-going signalling and gene expression.

This model could again be tested using mutations in the TAZ2 domain to generate a less specific CBP enzyme. This would be predicted to lead to more rapid and less precise upregulation of gene expression in response to signalling events. This could be particularly important during processes such as

development, where gene expression must be very finely controlled, and may explain why mutation of the TAZ2 domain in only a single allele of CBP is sufficient to generate developmental disease (Menke et al., 2016, 2018). Therefore, testing the effect of TAZ2 mutations over the course of differentiation could be an important step in understanding how CBP and other HATs function *in vivo*.

## 7.5 Towards a model for histone acetylation function?

These observations provide insight into how mechanistically histone acetylation might function to regulate gene expression. Several models have been proposed for how histone acetylation affects gene expression (Henikoff, 2005; Zentner and Henikoff, 2013). First, the “histone code” model suggests that individual histone modifications exert their function through specific binding by domains of effector proteins (Jenuwein and Allis, 2001; Turner, 2000). Histone acetylation in particular can be recognised by bromodomain-containing proteins associated with transcriptional activation, including chromatin remodellers and components of the transcriptional machinery (Filippakopoulos and Knapp, 2012). A second model suggests that certain histone acetylation marks, including H3K9ac and H3K27ac, function to activate transcription by preventing repressive methylation modifications at these residues.

However, it is difficult to reconcile these models with the results presented here. Were transcription guided by the binding of a specific effector to a specific acetylation mark such as H3K27ac, or the blocking of methylation at H3K27, it would be expected that all of the catalytically active CBP proteins tethered as dCas9 fusion proteins should give rise to similar levels of transcription, as each of these proteins generates the same levels of H3K27ac. In these experiments, however, the level of transcription corresponded to the overall level of acetylation rather than to the levels of one particular acetylation mark. Moreover, the histone code hypothesis relies on the existence of proteins that possess highly specific acetyl-lysine binding domains. In reality, however, bromodomains are rather promiscuous, with individual bromodomains binding to multiple different acetylated histone residues (Filippakopoulos and Knapp,

2012). This suggests that if proteins binding acetylated histones influence transcription, they do so by binding to multiple different acetylated lysines with limited specificity for individual sites.

A third model for histone acetylation function is that acetylation neutralises the positive charge of lysine side chains, weakening interactions between histones and the negatively charged DNA phosphate backbone, thereby allowing the transcriptional machinery access to the underlying DNA (Wade et al., 1997). This model would be consistent with the quantitative changes in gene expression observed here in dCas9-based tethering of truncated CBP proteins, and in previously reported responses to endogenous MAPK signalling, as cumulative increases in histone acetylation at multiple residues could be directly translated into increased accessibility and transcriptional activation. The charge neutralisation model is further supported by earlier work in which combinations of lysine-to-arginine mutations were made in H3 and H4 histone tails in *S. cerevisiae*, finding that different lysine residues within the same histone tail, with the possible exception of H4K16, function redundantly and cumulatively in the activation of gene expression (Dion et al., 2005; Martin et al., 2004). This suggests that the major role of histone acetylation is structural, facilitating access to the underlying DNA for the transcriptional machinery. This is consistent with the observation that directly recruiting the transcriptional machinery by tethering VP160 to a gene promoter results in far greater levels of transcriptional activation than are achieved even with promiscuous acetylation. Such a model could be further supported by extending the dCas9-based tethering experiments to additional loci to test whether the quantitative changes in histone acetylation and gene expression observed at *Ascl1* can be generalised to other gene regulatory elements with different DNA sequences and chromatin environments.

## 7.6 HAT specificity and disease: an avenue for therapy?

Central to these considerations of robustness in gene expression and quantitative transcriptional regulation is the observation that different HATs possess different specificities. This includes both substrate specificities mediated by mechanisms such as DNA binding by TAZ2 in CBP/p300, but also

specificity of recruitment mediated by interaction with transcription factors. In this way, gene activation is likely to involve concerted action of multiple HATs at the same locus, recruited by multiple transcription factors and acetylating different histone residues. Consistent with the importance of these mechanisms, multiple findings suggest that these forms of specificity are subverted in diseases such as cancer to drive aberrant gene expression, and that therefore HATs might be valid targets for future therapy.

Subversion of specificity can occur in several ways. For example, MOZ-CBP fusion proteins generated by translocation events in acute myeloid leukaemia (AML) possess two separate HAT activities, potentially conferring dual specificity and targeting, and permitting overexpression of target genes (Yang, 2015). AML can also be driven by alternative CBP/p300 fusion proteins, in which the catalytic domain and C-terminus of CBP or p300 is fused to the N-terminus of MLL1, resulting in a HAT enzyme that can be aberrantly recruited to promoters genome-wide through the ZF-CXXC domain of MLL1 (Yang, 2004). Similarly, in hormone receptor-driven cancers such as ER-positive breast cancer and androgen receptor (AR)-positive prostate cancer, transcription factors such as ER and AR can bind to target sites, recruit co-activators such as CBP/p300, and drive a transcriptional program independently of normal regulatory mechanisms (reviewed in Green and Carroll, 2007).

Studies pointing to the importance of HAT misregulation in cancer suggest that these enzymes would be promising targets for therapy. However, targeting of HATs has until recently been largely avoided. This is in part because specific and cell-permeable compounds targeting HAT enzymatic activity proved relatively elusive (Di Cerbo and Schneider, 2013). An additional problem, though, is that these enzymes are broadly expressed in different tissue types and act as general coactivators at a wide range of target genes, with HATs such as GCN5 thought to function at essentially all transcribed genes (Baptista et al., 2017; Bonnet et al., 2014). Inhibition of HATs could therefore be expected to give rise to unacceptable toxicity. However, recent work using CRISPR-Cas9 based screens has shown that loss of the HATs GCN5 and HBO1 affects cell proliferation in AML, and inhibition of GCN5 with small molecules does not lead to general toxicity in other cell types (MacPherson et al., 2019; Tzelepis et al.,

2016). Moreover, targeting of the bromodomain of CBP/p300 in AR-positive prostate cancer leads to loss of AR-dependent H3K27ac and to inhibition of cell proliferation without significant effect on AR-negative cells (Jin et al., 2017; Raisner et al., 2018), suggesting that non-transformed cells can compensate for CBP/p300 loss by the activity of other HAT enzymes.

These results show that loss of robustness in gene expression regulation, by driving gene expression through a single transcription factor and its specific coactivators in cancers, renders HAT inhibition a promising avenue for therapy, and demonstrates the importance of robust transcriptional regulation by multiple substrate-specific HATs in healthy cells.

## 7.7 Conclusions

The work presented in this thesis revealed that the substrate specificity of the histone acetyltransferase CBP is controlled by the sequence-independent DNA binding activity of the TAZ2 domain. This novel DNA binding activity points to a general role for sequence-independent DNA binding in chromatin interactions, and suggests that all HATs may have mechanisms to restrain their substrate specificity. This leads to a model in which concerted action by multiple HATs at gene regulatory regions plays an important role in ensuring robust regulation of gene expression (Fig. 7.1).

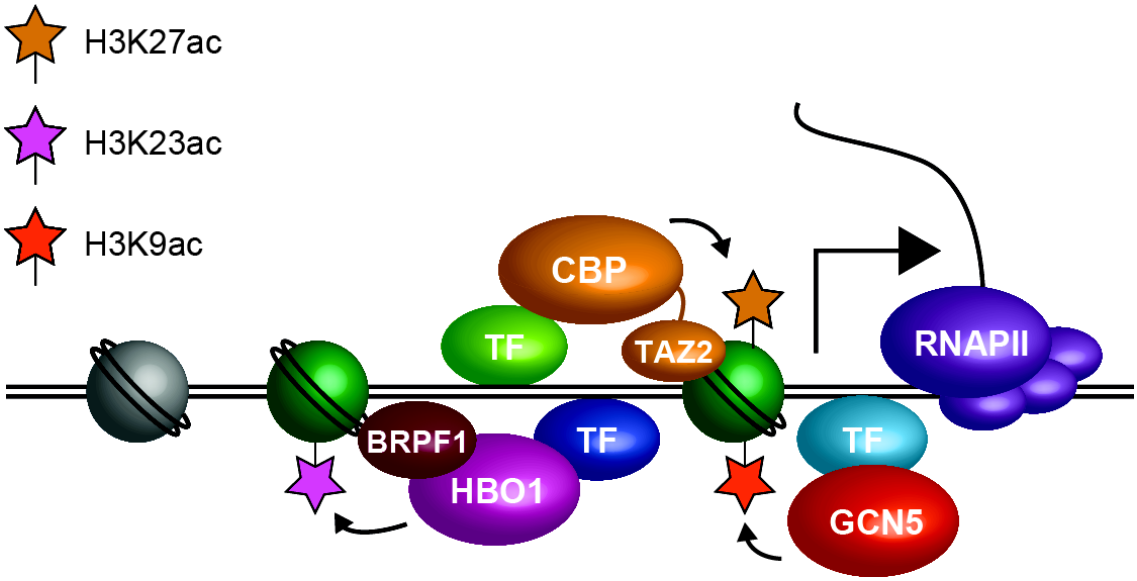



**Fig. 7.1: A model for HAT function at regulatory elements.**
In this model, representatives of all three HAT families are bound at a gene promoter. CBP specifically acetylates H3K27, with specificity mediated by TAZ2 DNA binding. HBO1 acetylates H3K23, with specificity mediated by BRPF1 DNA binding. GCN5 acetylates H3K9. Together, this coordinated acetylation facilitates access to the underlying DNA for the transcriptional machinery and leads to robust transcription of the target gene by RNA polymerase II.

# 8. Appendix

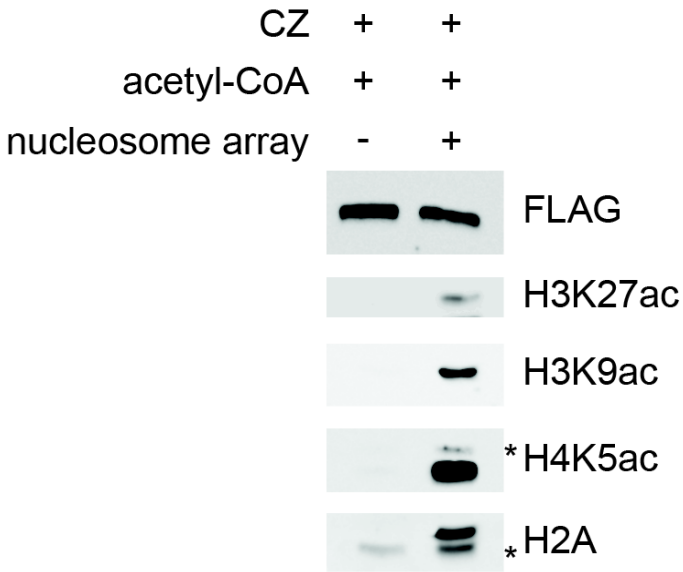



**Fig. S1: Testing the specificity of antibodies used in HAT assays.**
HAT assays were carried out for 20 mins using CBP CZ enzyme in the presence of acetyl-CoA and in the presence or absence of nucleosome array. Reactions were analysed by western blot using antibodies against H3K27ac, H3K9ac, H4K5ac, total H2A, and FLAG. The results suggest that a non-specific band (*) detected using H4K5ac antibody is only detectable in the presence of nucleosomes and therefore likely represents cross-reactivity with another acetylated histone residue. A non-specific band detected using H2A antibody is present, to a lower extent, in reactions without nucleosome, and likely represents a small amount of acetylated histone protein that co-purifies with the CZ enzyme.

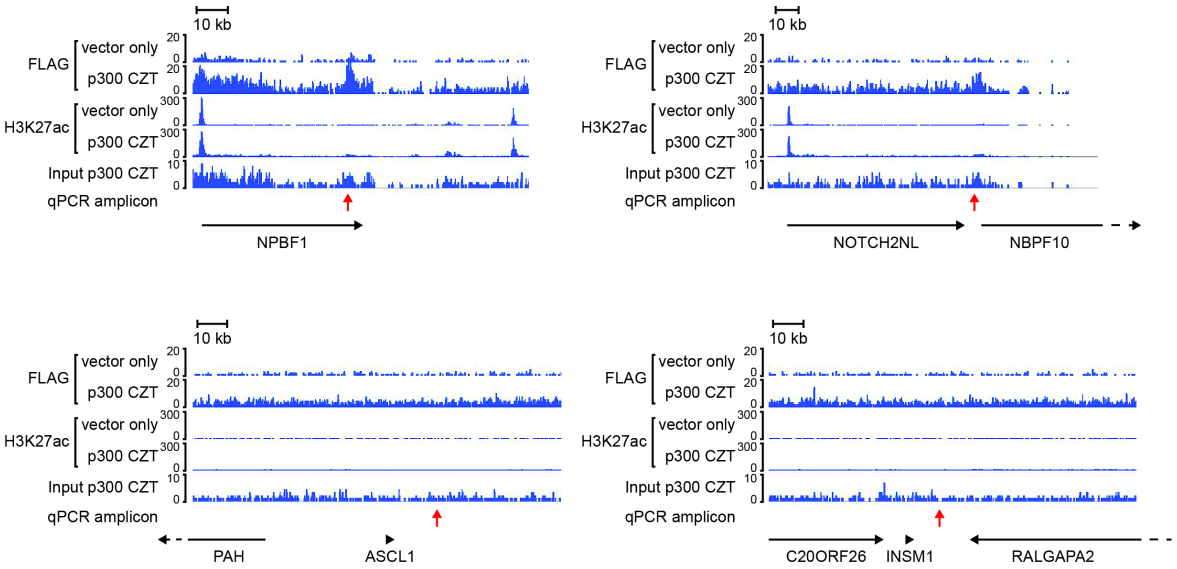



**Fig. S2: p300 CZT does not form clear peaks at sites of H3K27ac**
ChIP-seq snapshots using data from (Zhang et al., 2018), showing FLAG and H3K27ac ChIP in vector only and p300 CZT cell lines, and input for p300 CZT. Genes are indicated by horizontal black arrows below the sequencing traces and sites of amplicons used for qPCR analysis by vertical red arrows.
Top: two snapshots showing sites called as peaks of p300 CZT and with increasing H3K27ac.
Bottom: two snapshots without p300 CZT peaks or increasing H3K27ac.

# References


Abad, M.A., Ruppert, J.G., Buzuk, L., Wear, M., Zou, J., Webb, K.M., Kelly, D.A., Voigt, P., Rappsilber, J., Earnshaw, W.C., et al. (2019). Borealin-nucleosome interaction secures chromosome association of the chromosomal passenger complex. J. Cell Biol. *218*, 3912–3925.

Afek, A., and Lukatsky, D.B. (2012). Nonspecific protein-DNA binding is widespread in the yeast genome. Biophys. J. *102*, 1881–1888.

Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., and Thanos, D. (2000). Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter. Cell *103*, 667–678.

Agalioti, T., Chen, G., and Thanos, D. (2002). Deciphering the transcriptional histone acetylation code for a human gene. Cell *111*, 381–392.

Aguilar-Gurrieri, C. (2013). Structural studies of nucleosome assembly. University of Grenoble.

Aguilera, A., and Gómez-González, B. (2008). Genome instability: a mechanistic view of its causes and consequences. Nat. Rev. Genet. *9*, 204–217.

Akam, M. (1987). The molecular basis for metameric pattern in the Drosophila embryo. Development *101*, 1–22.

Alekseyenko, A., Gorchakov, A., Kharchenko, P., and Kuroda, M. (2014). Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. Proc Natl Acad Sci USA *111*, 2488–2493.

Alfieri, C., Gambetta, M.C., Matos, R., Glatt, S., Sehr, P., Fraterman, S., Wilm, M., Müller, J., and Müller, C.W. (2013). Structural basis for targeting the chromatin repressor Sfmbt to Polycomb response elements. Genes Dev. *27*, 2367–2379.

Alighieri, D., and Kirkpatrick, R. (2012). The Divine Comedy (Penguin Classics).

Allan, J., Hartman, P.G., Crane-Robinson, C., and Aviles, F.X. (1980). The structure of histone H1 and its location in chromatin. Nature *288*, 675–679.

Allard, S., Utley, R.T., Savard, J., Clarke, A., Grant, P., Brandl, C.J., Pillus, L., Workman, J.L., and Côté, J. (1999). NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. EMBO J. *18*, 5108–5119.

Allegra, P., Sterner, R., Clayton, D.F., and Allfrey, V.G. (1987). Affinity chromatographic purification of nucleosomes containing transcriptionally active DNA sequences. J. Mol. Biol. *196*, 379–388.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat. Rev. Mol. Cell Biol. *16*, 155–166.

Allen, M.D., Grummitt, C.G., Hilcenko, C., Min, S.Y., Tonkin, L.M., Johnson, C.M., Freund, S.M., Bycroft, M., and Warren, A.J. (2006). Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. EMBO J. *25*, 4503–4512.

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc. Natl. Acad. Sci. U. S. A. *51*, 786–794.

An, W., and Roeder, R.G. (2003). Direct association of p300 with unmodified

H3 and H4 N termini modulates p300-dependent acetylation and transcription of nucleosomal templates. J. Biol. Chem. *278*, 1504–1510.
An, W., Palhan, V.B., Karymov, M.A., Leuba, S.H., and Roeder, R.G. (2002). Selective requirements for histone H3 and H4 N termini in p300-dependent transcriptional activation from chromatin. Mol. Cell *9*, 811–821.
Antequera, F., and Bird, A. (1993). Number of CpG islands and genes in human and mouse. Proc. Natl. Acad. Sci. U. S. A. *90*, 11995–11999.
Arany, Z., Sellers, W.R., Livingston, D.M., and Eckner, R. (1994). E1A-associated p300 and CREB-associated CBP belong to a conserved family of coactivators. Cell *77*, 799–800.
Ardehali, M.B., Mei, A., Zobeck, K.L., Caron, M., Lis, J.T., and Kusch, T. (2011). Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. EMBO J. *30*, 2817–2828.
Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. *44*, W344–W350.
Attar, N., and Kurdistani, S.K. (2017). Exploitation of EP300 and CREBBP Lysine Acetyltransferases by Cancer. Cold Spring Harb. Perspect. Med. *7*.
Bannister, A.J., and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. Nature *384*, 641–643.
Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.
Baptista, T., Grünberg, S., Minoungou, N., Koster, M.J.E., Timmers, H.T.M., Hahn, S., Devys, D., and Tora, L. (2017). SAGA Is a General Cofactor for RNA Polymerase II Transcription. Mol. Cell *68*, 130-143.e5.
Barak, O., Lazzaro, M.A., Lane, W.S., Speicher, D.W., Picketts, D.J., and Shiekhattar, R. (2003). Isolation of human NURF: a regulator of Engrailed gene expression. EMBO J. *22*, 6089–6100.
Becker, P.B., and Workman, J.L. (2013). Nucleosome Remodeling and Epigenetics. Cold Spring Harb. Perspect. Biol. *5*, a017905–a017905.
Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A.R., Papai, G., Reymer, A., Syed, S.H., Lone, I.N., Tonchev, O., Crucifix, C., et al. (2017). Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. Mol. Cell *66*, 384-397.e8.
Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482–485.
Bentley, M.L., Corn, J.E., Dong, K.C., Phung, Q., Cheung, T.K., and Cochran, A.G. (2011). Recognition of UbcH5c and the nucleosome by the Bmi1/Ring1b ubiquitin ligase complex. EMBO J. *30*, 3285–3297.
Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.
Bestor, T.H., and Ingram, V.M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proc. Natl. Acad. Sci. U. S. A. *80*, 5559–

5563.
Bestor, T.H., and Verdine, G.L. (1994). DNA methyltransferases. Curr. Opin. Cell Biol. *6*, 380–389.
Bhaumik, P., Davis, J., Tropea, J.E., Cherry, S., Johnson, P.F., and Miller, M. (2014). Structural insights into interactions of C/EBP transcriptional activators with the Taz2 domain of p300. Acta Crystallogr. D. Biol. Crystallogr. *70*, 1914–1921.
Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209–213.
Black, J.B., Adler, A.F., Wang, H.-G., D'Ippolito, A.M., Hutchinson, H.A., Reddy, T.E., Pitt, G.S., Leong, K.W., and Gersbach, C.A. (2016). Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. Cell Stem Cell *19*, 406–414.
Blackledge, N.P., and Klose, R. (2011). CpG island chromatin: A platform for gene regulation. Epigenetics *6*, 147–152.
Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. Mol. Cell *38*, 179–190.
Blackledge, N.P., Long, H.K., Zhou, J.C., Kriaucionis, S., Patient, R., and Klose, R.J. (2012). Bio-CAP: a versatile and highly sensitive technique to purify and characterise regions of non-methylated DNA. Nucleic Acids Res. *40*, e32.
Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L.P., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. Cell 1–15.
Blackledge, N.P., Fursova, N.A., Kelley, J.R., Huseyin, M.K., Feldmann, A., and Klose, R.J. (2019). PRC1 Catalytic Activity Is Central to Polycomb System Function. Mol. Cell.
Blau, J., Xiao, H., McCracken, S., O'Hare, P., Greenblatt, J., and Bentley, D. (1996). Three functional classes of transcriptional activation domain. Mol. Cell. Biol. *16*, 2044–2055.
Bonnet, J., Wang, C.-Y., Baptista, T., Vincent, S.D., Hsiao, W.-C., Stierle, M., Kao, C.-F., Tora, L., and Devys, D. (2014). The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. Genes Dev. *28*, 1999–2012.
Bordoli, L., Netsch, M., Lüthi, U., Lutz, W., and Eckner, R. (2001). Plant orthologs of p300/CBP: conservation of a core domain in metazoan p300/CBP acetyltransferase-related proteins. Nucleic Acids Res. *29*, 589–597.
Borrow, J., Stanton, V.P., Andresen, J.M., Becher, R., Behm, F.G., Chaganti, R.S., Civin, C.I., Disteche, C., Dubé, I., Frischauf, A.M., et al. (1996). The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein. Nat. Genet. *14*, 33–41.
Bose, D.A., Donahue, G., Reinberg, D., Shiekhattar, R., Bonasio, R., and Berger, S.L. (2017). RNA Binding to CBP Stimulates Histone Acetylation and Transcription. Cell *168*, 135-149.e22.
Boveri, T. (1904). Ergebnisse über die Konstitution der chromatischen Substanz des Zellkerns (Jena: G. Fischer).
Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I.,

Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *72*, 248–254.

Breen, T.R., and Duncan, I.M. (1986). Maternal expression of genes that regulate the bithorax complex of Drosophila melanogaster. Dev. Biol. *118*, 442–456.

Brown, J.L. (2003). The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. Development *130*, 285–294.

Brown, C.E., Howe, L., Sousa, K., Alley, S.C., Carrozza, M.J., Tan, S., and Workman, J.L. (2001). Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit. Science *292*, 2333–2337.

Brown, D.A., Di Cerbo, V., Feldmann, A., Ahn, J., Ito, S., Blackledge, N.P., Nakayama, M., McClellan, M., Dimitrova, E., Turberfield, A.H., et al. (2017). The SET1 Complex Selects Actively Transcribed Target Genes via Multivalent Interaction with CpG Island Chromatin. Cell Rep. *20*, 2313–2327.

Brownell, J.E., and Allis, C.D. (1995). An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. Proc. Natl. Acad. Sci. U. S. A. *92*, 6364–6368.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843–851.

Buchwald, G., van der Stoop, P., Weichenrieder, O., Perrakis, A., van Lohuizen, M., and Sixma, T.K. (2006). Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. EMBO J. *25*, 2465–2474.

Calo, E., and Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? Mol. Cell *49*, 825–837.

Cao, R., and Zhang, Y. (2004). SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Mol. Cell *15*, 57–67.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039–1043.

Cao, R., Tsukada, Y.-I., and Zhang, Y. (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. Mol. Cell *20*, 845–854.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.-J., Anderson, S., Yates, J., Washburn, M.P., et al. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581–592.

Cerase, A., Smeets, D., Tang, Y.A., Gdula, M., Kraus, F., Spivakov, M., Moindrot, B., Leleu, M., Tattermusch, A., Demmerle, J., et al. (2014). Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. Proc. Natl. Acad. Sci. *111*, 2235–2240.

Di Cerbo, V., and Schneider, R. (2013). Cancers with wrong HATs: The impact of acetylation. Brief. Funct. Genomics *12*, 231–243.

Di Cerbo, V., Mohn, F., Ryan, D.P., Montellier, E., Kacem, S., Tropberger, P., Kallis, E., Holzner, M., Hoerner, L., Feldmann, A., et al. (2014). Acetylation of histone H3 at lysine 64 regulates nucleosome dynamics and facilitates transcription. Elife *2014*, 1–23.

Chen, T.A., and Allfrey, V.G. (1987). Rapid and reversible changes in nucleosome structure accompany the activation, repression, and superinduction of murine fibroblast protooncogenes c-fos and c-myc. Proc. Natl. Acad. Sci. U. S. A. *84*, 5252–5256.

Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. *23*, 1163–1171.

Choi, J., Bachmann, A.L., Tauscher, K., Benda, C., Fierz, B., and Müller, J. (2017). DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. Nat. Struct. Mol. Biol. *24*, 1039–1047.

Chrivia, J.C., Kwok, R.P., Lamb, N., Hagiwara, M., Montminy, M.R., and Goodman, R.H. (1993). Phosphorylated CREB binds specifically to the nuclear protein CBP. Nature *365*, 855–859.

Cierpicki, T., Risner, L.E., Grembecka, J., Lukasik, S.M., Popovic, R., Omonkowska, M., Shultis, D.D., Zeleznik-Le, N.J., and Bushweller, J.H. (2010). Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. Nat. Struct. Mol. Biol. *17*, 62–68.

Clayton, A.L., Rose, S., Barratt, M.J., and Mahadevan, L.C. (2000). Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosomes upon gene activation. EMBO J. *19*, 3714–3726.

Clayton, A.L., Hazzalin, C.A., and Mahadevan, L.C. (2006). Enhanced Histone Acetylation and Transcription: A Dynamic Perspective. Mol. Cell *23*, 289–296.

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J., Skalnik, D., and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. Genes Dev. *26*, 1714–1728.

Clouaire, T., Webb, S., and Bird, A. (2014). Cfp1 is required for gene expression-dependent H3K4 trimethylation and H3K9 acetylation in embryonic stem cells. Genome Biol. *15*, 451.

Comuzzi, B., Nemes, C., Schmidt, S., Jasarevic, Z., Lodde, M., Pycha, A., Bartsch, G., Offner, F., Culig, Z., and Hobisch, A. (2004). The androgen receptor co-activator CBP is up-regulated following androgen withdrawal and is highly expressed in advanced prostate cancer. J. Pathol. *204*, 159–166.

Conway, E., Jerman, E., Healy, E., Ito, S., Holoch, D., Oliviero, G., Deevy, O., Glancy, E., Fitzpatrick, D.J., Mucha, M., et al. (2018). A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities. Mol. Cell *70*, 408-421.e8.

Cook, S.J., Aziz, N., and McMahon, M. (1999). The repertoire of fos and jun proteins expressed during the G1 phase of the cell cycle is determined by the duration of mitogen-activated protein kinase activation. Mol. Cell. Biol. *19*, 330–341.

Cooper, D.N., Taggart, M.H., and Bird, A.P. (1983). Unmethylated domains in vertebrate DNA. Nucleic Acids Res. *11*, 647–658.

Cooper, S., Grijzenhout, A., Underwood, E., Ancelin, K., Zhang, T., Nesterova,

T.B., Anil-Kirmizitas, B., Bassett, A., Kooistra, S.M., Agger, K., et al. (2016). Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. Nat. Commun. *7*, 13661.

Copur, Ö., and Müller, J. (2013). The histone H3-K27 demethylase Utx regulates HOX gene expression in Drosophila in a temporally restricted manner. Development *140*, 3478–3485.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M. a, Frampton, G.M., Sharp, P. a, et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Cross, S.H., Meehan, R.R., Nan, X., and Bird, A. (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nat. Genet. *16*, 256–259.

Cruz-Molina, S., Respuela, P., Tebartz, C., Kolovos, P., Nikolic, M., Fueyo, R., van Ijcken, W.F.J., Grosveld, F., Frommolt, P., Bazzi, H., et al. (2017). PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. Cell Stem Cell *20*, 689-705.e9.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell *111*, 185–196.

Dancy, B.M., and Cole, P.A. (2015). Protein lysine acetylation by p300/CBP. Chem. Rev. *115*, 2419–2452.

Daujat, S., Weiss, T., Mohn, F., Lange, U.C., Ziegler-Birling, C., Zeissler, U., Lappe, M., Schübeler, D., Torres-Padilla, M.-E., and Schneider, R. (2009). H3K64 trimethylation marks heterochromatin and is dynamically remodeled during developmental reprogramming. Nat. Struct. Mol. Biol. *16*, 777–781.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J. Mol. Biol. *319*, 1097–1113.

Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. Nat. Struct. Mol. Biol. *20*, 1250–1257.

Debes, J.D., Sebo, T.J., Lohse, C.M., Murphy, L.M., Haugen, D.A.L., and Tindall, D.J. (2003). p300 in prostate cancer proliferation and progression. Cancer Res. *63*, 7638–7640.

Déjardin, J., and Kingston, R.E. (2009). Purification of Proteins Associated with Specific Genomic Loci. Cell *136*, 175–186.

Delvecchio, M., Gaucher, J., Aguilar-Gurrieri, C., Ortega, E., and Panne, D. (2013). Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation. Nat. Struct. Mol. Biol. *20*, 1040–1046.

Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., Anastassiadis, K., Stunnenberg, H.G., and Stewart, A.F. (2014). Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. Development *141*, 526–537.

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase

bromodomain. Nature *399*, 491–496.
van Dijk, T.B., Gillemans, N., Stein, C., Fanis, P., Demmers, J., van de Corput, M., Essers, J., Grosveld, F., Bauer, U.-M., and Philipsen, S. (2010). Friend of Prmt1, a novel chromatin target of protein arginine methyltransferases. Mol. Cell. Biol. *30*, 260–272.
Dillon, S.C., Zhang, X., Trievel, R.C., and Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol. *6*, 227.
Dimitrova, E., Kondo, T., Feldmann, A., Nakayama, M., Koseki, Y., Konietzny, R., Kessler, B.M., Koseki, H., and Klose, R.J. (2018). FBXL19 recruits CDK-Mediator to CpG islands of developmental genes priming them for activation during lineage commitment. Elife *7*.
Dion, M.F., Altschuler, S.J., Wu, L.F., and Rando, O.J. (2005). Genomic characterization reveals a simple histone H4 acetylation code. Proc. Natl. Acad. Sci. U. S. A. *102*, 5501–5506.
Dorighi, K.M., Swigut, T., Henriques, T., Bhanu, N. V., Scruggs, B.S., Nady, N., Still, C.D., Garcia, B.A., Adelman, K., and Wysocka, J. (2017). Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. Mol. Cell *66*, 568-576.e4.
Doyon, Y., Cayrou, C., Ullah, M., Landry, A.J., Côté, V., Selleck, W., Lane, W.S., Tan, S., Yang, X.J., and Côté, J. (2006). ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. Mol. Cell *21*, 51–64.
Durrin, L.K., Mann, R.K., Kayne, P.S., and Grunstein, M. (1991). Yeast histone H4 N-terminal sequence is required for promoter activation in vivo. Cell *65*, 1023–1031.
Dyson, H.J., and Wright, P.E. (2016). Role of intrinsic protein disorder in the function and interactions of the transcriptional coactivators CREB-binding Protein (CBP) and p300. J. Biol. Chem. *291*, 6714–6722.
Eberl, H.C., Spruijt, C.G., Kelstrup, C.D., Vermeulen, M., and Mann, M. (2013). A Map of General and Specialized Chromatin Readers in Mouse Tissues Generated by Label-free Interaction Proteomics. Mol. Cell *49*, 368–378.
Eckner, R., Ewen, M.E., Newsome, D., Gerdes, M., DeCaprio, J.A., Lawrence, J.B., and Livingston, D.M. (1994). Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. Genes Dev. *8*, 869–884.
Edmunds, J.W., Mahadevan, L.C., and Clayton, A.L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. EMBO J. *27*, 406–420.
ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
Endoh, M., Endo, T. a, Endoh, T., Isono, K., Sharif, J., Ohara, O., Toyoda, T., Ito, T., Eskeland, R., Bickmore, W. a, et al. (2012). Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. PLoS Genet. *8*, e1002774.
Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.

Erokhin, M., Elizar'ev, P., Parshikov, A., Schedl, P., Georgiev, P., and Chetverina, D. (2015). Transcriptional read-through is not sufficient to induce an epigenetic switch in the silencing activity of Polycomb response elements. Proc. Natl. Acad. Sci. U. S. A. *112*, 14930–14935.

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A., et al. (2010). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. Mol. Cell *38*, 452–464.

Esnault, C., Gualdrini, F., Horswell, S., Kelly, G., Stewart, A., East, P., Matthews, N., and Treisman, R. (2017). ERK-Induced Activation of TCF Family of SRF Cofactors Initiates a Chromatin Modification Cascade Associated with Transcription. Mol. Cell *65*, 1081-1095.e5.

Farcas, A.M., Blackledge, N.P., Sudbury, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cesare, A., Sheahan, T.W., et al. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. Elife *1*, e00205.

Feller, C., Forné, I., Imhof, A., and Becker, P.B. (2015). Global and specific responses of the histone acetylome to systematic perturbation. Mol. Cell *57*, 559–571.

Feng, H., Jenkins, L.M.M., Durell, S.R., Hayashi, R., Mazur, S.J., Cherry, S., Tropea, J.E., Miller, M., Wlodawer, A., Appella, E., et al. (2009). Structural Basis for p300 Taz2-p53 TAD1 Binding and Modulation by Phosphorylation. Structure *17*, 202–210.

Feng, Y., Vlassis, A., Roques, C., Lalonde, M., González Aguilera, C., Lambert, J., Lee, S., Zhao, X., Alabert, C., Johansen, J. V, et al. (2016). BRPF 3-HBO 1 regulates replication origin activation and histone H3K14 acetylation . EMBO J. *35*, 176–192.

Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J.Z., Innocenti, C., Ferrier, P., Spicuglia, S., Gut, M., Gut, I., et al. (2012). CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. Genome Res. *22*, 2399–2408.

Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398–402.

Filippakopoulos, P., and Knapp, S. (2012). The bromodomain interaction module. FEBS Lett. *586*, 2692–2704.

Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., et al. (2010). Selective inhibition of BET bromodomains. Nature *468*, 1067–1073.

Flemming, W. (1879). Ueber das Verhalten des Kerns bei der Zelltheilung, und über die Bedeutung mehrkerniger Zellen. Arch. Für Pathol. Anat. Und Physiol. Und Für Klin. Med.

Fortschegger, K., and Shiekhattar, R. (2011). Plant homeodomain fingers form a helping hand for transcription. Epigenetics *6*, 4–8.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science *306*, 1574–1577.

Frey, F., Sheahan, T., Finkl, K., Stoehr, G., Mann, M., Benda, C., and Müller, J. (2016). Molecular basis of PRC1 targeting to polycomb response elements by PhoRC. Genes Dev. *30*, 1116–1127.

Fritsch, C., Brown, J.L., Kassis, J. a, and Müller, J. (1999). The DNA-binding

polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. Development *126*, 3905–3913.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol. Cell *45*, 344–356.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J. Mol. Biol. *196*, 261–282.

Gates, L.A., Shi, J., Rohira, A.D., Feng, Q., Zhu, B., Bedford, M.T., Sagum, C.A., Jung, S.Y., Qin, J., Tsai, M.J., et al. (2017). Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. J. Biol. Chem. *292*, 14456–14472.

Gayther, S.A., Batley, S.J., Linger, L., Bannister, A., Thorpe, K., Chin, S.-F., Daigo, Y., Russell, P., Wilson, A., Sowter, H.M., et al. (2000). Mutations truncating the EP300 acetylase in human cancers The EP300 protein is a histone acetyltransferase 1,2 that regulates transcription via chromatin remodelling 3 and is important in the processes of cell proliferation 4 and differentiation 5. EP300 . *24*, 1–4.

Gill, G., and Ptashne, M. (1988). Negative effect of the transcriptional activator GAL4. Nature *334*, 721–724.

Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K.R., Aasland, R., Anastassiadis, K., Ang, S.-L., and Stewart, a F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. Development *133*, 1423–1432.

Goodman, R.H., and Smolik, S. (2000). CBP/p300 in cell growth, transformation, and development. Genes Dev. *14*, 1553–1577.

Gozdecka, M., Meduri, E., Mazan, M., Tzelepis, K., Dudek, M., Knights, A.J., Pardo, M., Yu, L., Choudhary, J.S., Metzakopian, E., et al. (2018). UTX-mediated enhancer and chromatin remodeling suppresses myeloid leukemogenesis through noncatalytic inverse regulation of ETS and GATA programs. Nat. Genet. *50*, 883–894.

Grant, P.A., Duggan, L., Côté, J., Roberts, S.M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., Winston, F., et al. (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes Dev. *11*, 1640–1650.

Green, K.A., and Carroll, J.S. (2007). Oestrogen-receptor-mediated transcription and the influence of co-factors and chromatin state. Nat. Rev. Cancer *7*, 713–722.

Green, J.B., New, H. V, and Smith, J.C. (1992). Responses of embryonic Xenopus cells to activin and FGF are separated by multiple dose thresholds and correspond to distinct axes of the mesoderm. Cell *71*, 731–739.

Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. Nat. Rev. Genet. *13*, 343–357.

Grijzenhout, A., Godwin, J., Koseki, H., Gdula, M., Szumska, D., McGouran, J.F., Bhattacharya, S., Kessler, B.M., Brockdorff, N., and Cooper, S. (2016). Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ES cells. Development 2716–2723.

Grimm, C., Matos, R., Ly-Hartig, N., Steuerwald, U., Lindner, D., Rybin, V., Müller, J., and Müller, C.W. (2009). Molecular recognition of histone lysine methylation by the Polycomb group repressor dSfmbt. EMBO J.

*28*, 1965–1977.
Grossman, S.R. (2001). p300/CBP/p53 interaction and regulation of the p53 response. Eur. J. Biochem. *268*, 2773–2778.
Gu, W., Shi, X.L., and Roeder, R.G. (1997). Synergistic activation of transcription by CBP and p53. Nature *387*, 819–823.
Guenther, M.G., Levine, S.S., Boyer, L. a, Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77–88.
Gurdon, J.B., and Bourillot, P.Y. (2001). Morphogen gradient interpretation. Nature *413*, 797–803.
Gurdon, J.B., Harger, P., Mitchell, A., and Lemaire, P. (1994). Activin signalling and response to a morphogen gradient. Nature *371*, 487–492.
Hall, D.B., and Struhl, K. (2002). The VP16 activation domain interacts with multiple transcriptional components as determined by protein-protein cross-linking in vivo. J. Biol. Chem. *277*, 46043–46050.
Hallson, G., Hollebakken, R.E., Li, T., Syrzycka, M., Kim, I., Cotsworth, S., Fitzpatrick, K. a, Sinclair, D. a R., and Honda, B.M. (2012). dSet1 is the main H3K4 di- and tri-methyltransferase throughout Drosophila development. Genetics *190*, 91–100.
Han, M., and Grunstein, M. (1988). Nucleosome loss activates yeast downstream promoters in vivo. Cell *55*, 1137–1145.
Hassan, A.H., Prochasson, P., Neely, K.E., Galasinski, S.C., Chandy, M., Carrozza, M.J., and Workman, J.L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell *111*, 369–379.
Hazzalin, C. a, and Mahadevan, L.C. (2005). Dynamic acetylation of all lysine 4-methylated histone H3 in the mouse nucleus: analysis at c-fos and c-jun. PLoS Biol. *3*, e393.
Hazzalin, C.A., and Mahadevan, L.C. (2002). MAPK-Regulated transcription: A continuously variable gene switch? Nat. Rev. Mol. Cell Biol. *3*, 30–40.
He, J., Shen, L., Wan, M., Taranova, O., Wu, H., and Zhang, Y. (2013). Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nat. Cell Biol. *15*.
He, W., Zhang, L., Villarreal, O.D., Fu, R., Bedford, E., Dou, J., Patel, A.Y., Bedford, M.T., Shi, X., Chen, T., et al. (2019). De novo identification of essential protein domains from CRISPR-Cas9 tiling-sgRNA knockout screens. Nat. Commun. *10*.
Hebbes, T.R., Thorne, A.W., and Crane-Robinson, C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. EMBO J. *7*, 1395–1402.
Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.
Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.
Heitz, E. (1928). Das Heterochromatin der Moose. Jahrbücher Für Wissenschaftliche Bot.
Heitz, E. (1929). Heterochromatin, Chromocentren, Chromomeren. Ber. Dtsch. Bot. Ges. *47*, 274–284.

Helmrich, A., Ballarino, M., Nudler, E., and Tora, L. (2013). Transcription-replication encounters, consequences and genomic instability. Nat. Struct. Mol. Biol. *20*, 412–418.

Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol. Cell. Biol. *18*, 6538–6547.

Henikoff, S. (2005). Histone modifications: Combinational complexity or cumulative simplicity? Proc. Natl. Acad. Sci. U. S. A. *102*, 5308–5309.

Henikoff, S., and Shilatifard, A. (2011). Histone modification: cause or cog? Trends Genet. *27*, 389–396.

Hennekam, R.C.M. (2006). Rubinstein-Taybi syndrome. Eur. J. Hum. Genet. *14*, 981–985.

Herz, H.-M., Garruss, A., and Shilatifard, A. (2013). SET for life: biochemical activities and biological functions of SET domain-containing proteins. Trends Biochem. Sci. *38*, 621–639.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Holbert, M.A., Sikorski, T., Carten, J., Snowflack, D., Hodawadekar, S., and Marmorstein, R. (2007). The human monocytic leukemia zinc finger histone acetyltransferase domain contains DNA-binding activity implicated in chromatin targeting. J. Biol. Chem. *282*, 36603–36613.

Hong, S., Cho, Y.-W., Yu, L.-R., Yu, H., Veenstra, T.D., and Ge, K. (2007). Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. Proc. Natl. Acad. Sci. U. S. A. *104*, 18439–18444.

Horton, S.J., Giotopoulos, G., Yun, H., Vohra, S., Sheppard, O., Bashford-Rogers, R., Rashid, M., Clipson, A., Chan, W.I., Sasca, D., et al. (2017). Early loss of Crebbp confers malignant stem cell properties on lymphoid progenitors. Nat. Cell Biol. *19*, 1093–1104.

Hottiger, M.O., Felzien, L.K., and Nabel, G.J. (1998). Modulation of cytokine-induced HIV gene expression by competitive binding of transcription factors to the coactivator p300. EMBO J. *17*, 3124–3134.

Hu, D., Garruss, A.S., Gao, X., Morgan, M. a, Cook, M., Smith, E.R., and Shilatifard, A. (2013a). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1093–1097.

Hu, D., Gao, X., Morgan, M. a, Herz, H.-M., Smith, E.R., and Shilatifard, A. (2013b). The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Mol. Cell. Biol. *33*, 4745–4754.

Hu, D., Gao, X., Morgan, M. a, Herz, H.-M., Smith, E.R., and Shilatifard, A. (2013c). The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Mol. Cell. Biol. *33*, 4745–4754.

Huang, R.C., and Bonner, J. (1962). Histone, a suppressor of chromosomal RNA synthesis. Proc. Natl. Acad. Sci. U. S. A. *48*, 1216–1222.

Hunkapiller, J., Shen, Y., Diaz, A., Cagney, G., McCleary, D., Ramalho-Santos, M., Krogan, N., Ren, B., Song, J.S., and Reiter, J.F. (2012). Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. PLoS Genet. *8*, e1002576.

Ianculescu, I., Wu, D.-Y., Siegmund, K.D., and Stallcup, M.R. (2012). Selective roles for cAMP response element-binding protein binding protein and p300 protein as coregulators for androgen-regulated gene expression in advanced prostate cancer cells. J. Biol. Chem. *287*, 4000–4013.

Ide, S., and Dejardin, J. (2015). End-targeting proteomics of isolated chromatin segments of a mammalian ribosomal RNA gene promoter. Nat. Commun. *6*, 6674.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. FEBS Lett. *583*, 1713–1720.

Illingworth, R., Kerr, A., DeSousa, D., Jørgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., et al. (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol. *6*, 0037–0051.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG Islands Identify numerous conserved promoters in the mammalian genome. PLoS Genet. *6*, e1001134.

Illingworth, R.S., Moffat, M., Mann, A.R., Read, D., Hunter, C.J., Pradeepa, M.M., Adams, I.R., and Bickmore, W.A. (2015). The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. Genes Dev. *29*, 1897–1902.

Imai, S., Armstrong, C.M., Kaeberlein, M., and Guarente, L. (2000). Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. Nature *403*, 795–800.

Inoue, A., and Fujimoto, D. (1969). Enzymatic deacetylation of histone. Biochem. Biophys. Res. Commun. *36*, 146–150.

Isono, K., Endo, T.A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B.E., and Koseki, H. (2013). SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. Dev. Cell *26*, 565–577.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303.

Iyer, N.G., Ozdag, H., and Caldas, C. (2004). p300/CBP and cancer. Oncogene *23*, 4225–4231.

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422–1425.

Jeltsch, A. (2006). On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. Epigenetics *1*, 63–66.

Jenkins, L.M.M., Yamaguchi, H., Hayashi, R., Cherry, S., Tropea, J.E., Miller, M., Wlodawer, A., Appella, E., and Mazur, S.J. (2009). Two distinct motifs within the p53 transactivation domain bind to the Taz2 domain of p300 and are differentially affected by phosphorylation. Biochemistry *48*, 1244–1255.

Jenkins, L.M.M., Feng, H., Durell, S.R., Tagad, H.D., Mazur, S.J., Tropea, J.E., Bai, Y., and Appella, E. (2015). Characterization of the p300 Taz2-p53 TAD2 Complex and Comparison with the p300 Taz2-p53 TAD1 Complex. Biochemistry *54*, 2001–2010.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080.

Jin, L., Garcia, J., Chan, E., de la Cruz, C., Segal, E., Merchant, M., Kharbanda, S., Raisner, R., Haverty, P.M., Modrusan, Z., et al. (2017). Therapeutic Targeting of the CBP/p300 Bromodomain Blocks the Growth of Castration-Resistant Prostate Cancer. Cancer Res. *77*, 5564–5575.

Jin, Q., Yu, L.R., Wang, L., Zhang, Z., Kasper, L.H., Lee, J.E., Wang, C., Brindle, P.K., Dent, S.Y.R., and Ge, K. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. EMBO J. *30*, 249–262.

Johnson, C.A., White, D.A., Lavender, J.S., O'Neill, L.P., and Turner, B.M. (2002). Human class I histone deacetylase complexes show enhanced catalytic activity in the presence of ATP and co-immunoprecipitate with the ATP-dependent chaperone protein Hsp70. J. Biol. Chem. *277*, 9590–9597.

Kabadi, A.M., Ousterout, D.G., Hilton, I.B., and Gersbach, C.A. (2014). Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. Nucleic Acids Res. *42*, e147.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kahn, T.G., Stenberg, P., Pirrotta, V., and Schwartz, Y.B. (2014). Combinatorial Interactions Are Required for the Efficient Recruitment of Pho Repressive Complex (PhoRC) to Polycomb Response Elements. PLoS Genet. *10*, e1004495.

Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W.T.C., Müller, C.W., Vermeulen, M., and Müller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. Nat. Struct. Mol. Biol. 10–13.

Kalkhoven, E., Teunissen, H., Houweling, A., Verrijzer, C.P., and Zantema, A. (2002). The PHD Type Zinc Finger Is an Integral Part of the CBP Acetyltransferase Domain. Mol. Cell. Biol. *22*, 1961–1970.

Kamei, Y., Xu, L., Heinzel, T., Torchia, J., Kurokawa, R., Gloss, B., Lin, S.C., Heyman, R.A., Rose, D.W., Glass, C.K., et al. (1996). A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. Cell *85*, 403–414.

Kanoksilapatham, W., Gonzalez, J., and Robb, F. (2007). Directed-Mutagenesis and Deletion Generated through an Improved Overlapping-Extension PCR Based Procedure. Silpakorn Univ. Sci. Technol. J. *1*, 7–12.

Karmodiya, K., Krebs, A.R., Oulad-Abdelghani, M., Kimura, H., and Tora, L. (2012). H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC Genomics *13*.

Kim, D.-H., Tang, Z., Shimada, M., Fierz, B., Houck-Loomis, B., Bar-Dagen, M., Lee, S., Lee, S.-K., Muir, T.W., Roeder, R.G., et al. (2013). Histone H3K27 trimethylation inhibits H3 binding and function of SET1-like H3K4 methyltransferase complexes. Mol. Cell. Biol. *33*, 4936–4946.

Kim, T.K., Kim, T.H., and Maniatis, T. (1998). Efficient recruitment of TFIIB and CBP-RNA polymerase II holoenzyme by an interferon- enhanceosome in vitro. Proc. Natl. Acad. Sci. *95*, 12191–12196.

King, I.F.G., Francis, N.J., and Kingston, R.E. (2002). Native and Recombinant Polycomb Group Complexes Establish a Selective Block to Template Accessibility To Repress Transcription In Vitro Native and Recombinant

Polycomb Group Complexes Establish a Selective Block to Template Accessibility To Repress Transc.
Kitabayashi, I., Aikawa, Y., Nguyen, L.A., Yokoyama, A., and Ohki, M. (2002). Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ-CBP fusion protein. EMBO J. *20*, 7184–7196.
Kleff, S., Andrulis, E.D., Anderson, C.W., and Sternglanz, R. (1995). Identification of a Gene Encoding a Yeast Histone H4 Acetyltransferase. J. Biol. Chem. *270*, 24674–24677.
Klein, B.J., Muthurajan, U.M., Lalonde, M.E., Gibson, M.D., Andrews, F.H., Hepler, M., Machida, S., Yan, K., Kurumizaka, H., Poirier, M.G., et al. (2016). Bivalent interaction of the PZP domain of BRPF1 with the nucleosome impacts chromatin dynamics and acetylation. Nucleic Acids Res. *44*, 472–484.
Klein, B.J., Jang, S.M., Lachance, C., Mi, W., Lyu, J., Sakuraba, S., Krajewski, K., Wang, W.W., Sidoli, S., Liu, J., et al. (2019). Histone H3K23-specific acetylation by MORF is coupled to H3K14 acylation. Nat. Commun. *10*.
Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends Biochem. Sci. *31*, 89–97.
Klose, R.J., Cooper, S., Farcas, A.M., Blackledge, N.P., and Brockdorff, N. (2013). Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins. PLoS Genet. *9*, e1003717.
Klymenko, T., Papp, B., Fischle, W., Köcher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Müller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes Dev. *20*, 1110–1122.
Knezetic, J.A., and Luse, D.S. (1986). The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. Cell *45*, 95–104.
Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868–871.
Kornberg, R.D., and Thomas, J.O. (1974). Chromatin structure; oligomers of the histones. Science *184*, 865–868.
Kossel, A. (1884). Über einen peptonartigen Bestandtheil des Zellkerns. Hoppe-Seyler's Z. Physiol. Chem *8*, 551.
Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.
Kraus, W.L., Manning, E.T., and Kadonaga, J.T. (1999). Biochemical Analysis of Distinct Activation Functions in p300 That Enhance Transcription Initiation with Chromatin Templates. Mol. Cell. Biol. *19*, 8123–8135.
Krebs, A.R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., and Tora, L. (2011). SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. Mol. Cell *44*, 410–423.
Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929–930.
Krois, A.S., Ferreon, J.C., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2016). Recognition of the disordered p53 transactivation domain by the transcriptional adapter zinc finger domains of CREB-binding protein. Proc. Natl. Acad. Sci. U. S. A. *113*, E1853–E1862.
Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008).

Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. PLoS Genet. *4*, 14.

Kuo, M.H., Brownell, J.E., Sobel, R.E., Ranalli, T.A., Cook, R.G., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. Nature *383*, 269–272.

Kuo, M.H., vom Baur, E., Struhl, K., and Allis, C.D. (2000). Gcn4 activator targets Gcn5 histone acetyltransferase to specific promoters independently of transcription. Mol. Cell *6*, 1309–1320.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev. *16*, 2893–2905.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680–685.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A.A., and Verrijzer, C.P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. Genes Dev. *22*, 2799–2810.

Lalonde, M.E., Avvakumov, N., Glass, K.C., Joncas, F.H., Saksouk, N., Holliday, M., Paquet, E., Yan, K., Tong, Q., Klein, B.J., et al. (2013). Exchange of associated factors directs a switch in HBO1 acetyltransferase histone tail specificity. Genes Dev. *27*, 2009–2024.

Lan, F., Bayliss, P.E., Rinn, J.L., Whetstine, J.R., Wang, J.K., Chen, S., Iwase, S., Alpatov, R., Issaeva, I., Canaani, E., et al. (2007). A histone H3 lysine 27 demethylase regulates animal posterior development. Nature *449*, 689–694.

Lange, U.C., Siebert, S., Wossidlo, M., Weiss, T., Ziegler-Birling, C., Walter, J., Torres-Padilla, M.-E., Daujat, S., and Schneider, R. (2013). Dissecting the role of H3K64me3 in mouse pericentromeric heterochromatin. Nat. Commun. *4*, 2233.

Laurent, B., Ruitu, L., Murn, J., Hempel, K., Ferrao, R., Xiang, Y., Liu, S., Garcia, B.A., Wu, H., Wu, F., et al. (2015). A specific LSD1/KDM1A isoform regulates neuronal differentiation through H3K9 demethylation. Mol. Cell *57*, 957–970.

Lee, J.-H., and Skalnik, D.G. (2005). CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. J. Biol. Chem. *280*, 41725–41731.

Lee, J.-H., and Skalnik, D.G. (2008). Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. Mol. Cell. Biol. *28*, 609–618.

Lee, K.K., and Workman, J.L. (2007). Histone acetyltransferase complexes: One size doesn't fit all. Nat. Rev. Mol. Cell Biol. *8*, 284–295.

Lee, D.Y., Hayes, J.J., Pruss, D., and Wolffe, A.P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell *72*, 73–84.

Lee, J.E., Wang, C., Xu, S., Cho, Y.W., Wang, L., Feng, X., Baldridge, A.,

Sartorelli, V., Zhuang, L., Peng, W., et al. (2013). H3K4 mono- And di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. Elife *2013*, 1–25.

Lee, J.H., Voo, K.S., and Skalnik, D.G. (2001). Identification and characterization of the DNA binding domain of CpG-binding protein. J. Biol. Chem. *276*, 44669–44676.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. *20*, 311–316.

Levine, S.S., Weiss, A., Erdjument-Bromage, H., Shao, Z., Tempst, P., and Kingston, R.E. (2002). The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. Mol. Cell. Biol. *22*, 6070–6078.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926.

Li, H., Liefke, R., Jiang, J., Kurland, J.V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D.J., Bulyk, M.L., et al. (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. Nature *549*, 287–291.

Liefke, R., and Shi, Y. (2015). The PRC2-associated factor C17orf96 is a novel CpG island regulator in mouse ES cells. Cell Discov. *1*, 15008.

Ling, X., Harkness, T.A., Schultz, M.C., Fisher-Adams, G., and Grunstein, M. (1996). Yeast histone H3 and H4 amino termini are important for nucleosome assembly in vivo and in vitro: redundant and position-independent functions in assembly but not in gene regulation. Genes Dev. *10*, 686–699.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Liu, L., Qin, S., Zhang, J., Ji, P., Shi, Y., and Wu, J. (2012). Solution structure of an atypical PHD finger in BRPF2 and its interaction with DNA. J. Struct. Biol. *180*, 165–173.

Liu, X., Wang, L., Zhao, K., Thompson, P.R., Hwang, Y., Marmorstein, R., and Cole, P.A. (2008). The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. Nature *451*, 846–850.

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233-247.e17.

Local, A., Huang, H., Albuquerque, C.P., Singh, N., Lee, A.Y., Wang, W., Wang, C., Hsia, J.E., Shiau, A.K., Ge, K., et al. (2018). Identification of H3K4me1-associated proteins at mammalian enhancers. Nat. Genet. *50*, 73–82.

Long, H.K., Blackledge, N.P., and Klose, R.J. (2013). ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochem. Soc. Trans. *41*, 727–740.

Lorch, Y., LaPointe, J.W., and Kornberg, R.D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell *49*, 203–210.

Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity

binding to histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. *276*, 19–42.

Lu, H., Pise-Masison, C.A., Fletcher, T.M., Schiltz, R.L., Nagaich, A.K., Radonovich, M., Hager, G., Cole, P.A., and Brady, J.N. (2002). Acetylation of nucleosomal histones by p300 facilitates transcription from tax-responsive human T-cell leukemia virus type 1 chromatin template. Mol. Cell. Biol. *22*, 4450–4462.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251–260.

Lyst, M.J., Ekiert, R., Ebert, D.H., Merusi, C., Nowak, J., Selfridge, J., Guy, J., Kastan, N.R., Robinson, N.D., de Lima Alves, F., et al. (2013). Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. Nat. Neurosci. *16*, 898–902.

Ma, H., Tu, L.-C., Naseri, A., Huisman, M., Zhang, S., Grunwald, D., and Pederson, T. (2016). CRISPR-Cas9 nuclear dynamics and target recognition in living cells. J. Cell Biol. *214*, 529–537.

Ma, Q., Alder, H., Nelson, K.K., Chatterjee, D., Gu, Y., Nakamura, T., Canaani, E., Croce, C.M., Siracusa, L.D., and Buchberg, A.M. (1993). Analysis of the murine All-1 gene reveals conserved domains with human ALL-1 and identifies a motif shared with DNA methyltransferases. Proc. Natl. Acad. Sci. U. S. A. *90*, 6350–6354.

MacPherson, L., Anokye, J., Yeung, M.M., Lam, E.Y.N., Chan, Y.-C., Weng, C.-F., Yeh, P., Knezevic, K., Butler, M.S., Hoegl, A., et al. (2019). HBO1 is required for the maintenance of leukaemia stem cells. Nature *577*.

Mancini, D.N., Singh, S.M., Archer, T.K., and Rodenhiser, D.I. (1999). Site-specific DNA methylation in the neurofibromatosis (NF1) promoter interferes with binding of CREB and SP1 transcription factors. Oncogene *18*, 4108–4119.

Mann, R.K., and Grunstein, M. (1992). Histone H3 N-terminal mutations allow hyperactivation of the yeast GAL1 gene in vivo. EMBO J. *11*, 3297–3306.

Manning, E.T., Ikehara, T., Ito, T., Kadonaga, J.T., and Kraus, W.L. (2001). p300 Forms a Stable, Template-Committed Complex with Chromatin: Role for the Bromodomain. Mol. Cell. Biol. *21*, 3876–3887.

Marcus, G.A., Silverman, N., Berger, S.L., Horiuchi, J., and Guarente, L. (1994). Functional similarity and physical association between GCN5 and ADA2: putative transcriptional adaptors. EMBO J. *13*, 4807–4815.

Marmorstein, R., and Zhou, M.M. (2014). Writers and readers of histone acetylation: Structure, mechanism, and inhibition. Cold Spring Harb. Perspect. Biol. *6*.

Martin, A.M., Pouchnik, D.J., Walker, J.L., and Wyrick, J.J. (2004). Redundant roles for histone H3 N-terminal lysine residues in subtelomeric gene repression in Saccharomyces cerevisiae. Genetics *167*, 1123–1132.

Martinez-Balbás, M.A., Bannister, A.J., Martin, K., Haus-Seuffert, P., Meisterernst, M., and Kouzarides, T. (1998). The acetyltransferase activity of CBP stimulates transcription. EMBO J. *17*, 2886–2893.

Martinez, E., Kundu, T.K., Fu, J., and Roeder, R.G. (1998). A human SPT3-TAFII31-GCN5-L acetylase complex distinct from transcription factor IID. J. Biol. Chem. *273*, 23781–23785.

Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional Regulatory Elements in the Human Genome. Annu. Rev. Genomics Hum. Genet. *7*,

29–59.
Masumoto, H., Hawke, D., Kobayashi, R., and Verreault, A. (2005). A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. Nature *436*, 294–298.
Matt, T., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2004). The CBP/p300 TAZ1 domain in its native state is not a binding partner of MDM2. Biochem. J. *381*, 685–691.
Mazo, A.M., Huang, D.H., Mozer, B.A., and Dawid, I.B. (1990). The trithorax gene, a trans-acting regulator of the bithorax complex in Drosophila, encodes a protein with zinc-binding domains. Proc. Natl. Acad. Sci. U. S. A. *87*, 2112–2116.
McMahon, S.B., Van Buskirk, H.A., Dugan, K.A., Copeland, T.D., and Cole, M.D. (1998). The Novel ATM-Related Protein TRRAP Is an Essential Cofactor for the c-Myc and E2F Oncoproteins. Cell *94*, 363–374.
Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet. *6*, e1001244.
Mendes, M.L., Fischer, L., Chen, Z.A., Barbon, M., O'Reilly, F.J., Giese, S.H., Bohlke Schneider, M., Belsom, A., Dau, T., Combe, C.W., et al. (2019). An integrated workflow for crosslinking mass spectrometry. Mol. Syst. Biol. *15*.
Menke, L.A., van Belzen, M.J., Alders, M., Cristofoli, F., Ehmke, N., Fergelot, P., Foster, A., Gerkes, E.H., Hoffer, M.J.V., Horn, D., et al. (2016). CREBBP mutations in individuals without Rubinstein–Taybi syndrome phenotype. Am. J. Med. Genet. Part A *170*, 2681–2693.
Menke, L.A., Gardeitchik, T., Hammond, P., Heimdal, K.R., Houge, G., Hufnagel, S.B., Ji, J., Johansson, S., Kant, S.G., Kinning, E., et al. (2018). Further delineation of an entity caused by CREBBP and EP300 mutations but not resembling Rubinstein–Taybi syndrome. Am. J. Med. Genet. Part A *176*, 862–876.
Meyer, M.E., Gronemeyer, H., Turcotte, B., Bocquel, M.T., Tasset, D., and Chambon, P. (1989). Steroid hormone receptors compete for factors that mediate their enhancer function. Cell *57*, 433–442.
Mi, W., Guan, H., Lyu, J., Zhao, D., Xi, Y., Jiang, S., Andrews, F.H., Wang, X., Gagea, M., Wen, H., et al. (2017). YEATS2 links histone acetylation to tumorigenesis of non-small cell lung cancer. Nat. Commun. *8*.
Miescher-Rüsch, F. (1871). Ueber die chemische Zusammensetzung der Eiterzellen [On the chemical composition of pus cells]. Med. Untersuchungen.
Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.
Miller, M., Dauter, Z., Cherry, S., Tropea, J.E., and Wlodawer, A. (2009). Structure of the Taz2 domain of p300: Insights into ligand binding. Acta Crystallogr. Sect. D Biol. Crystallogr. *65*, 1301–1308.
Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., and Shilatifard, a (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc. Natl. Acad. Sci. U. S. A. *98*, 12902–12907.
Miller, T.C.R., Simon, B., Rybin, V., Grötsch, H., Curtet, S., Khochbin, S., Carlomagno, T., and Müller, C.W. (2016). A bromodomain–DNA

interaction facilitates acetylation-dependent bivalent nucleosome recognition by the BET protein BRDT. Nat. Commun. *7*, 13855.

Milne, T.A., Dou, Y., Martin, M.E., Brock, H.W., Roeder, R.G., and Hess, J.L. (2005). MLL associates specifically with a subset of transcriptionally active target genes. Proc. Natl. Acad. Sci. U. S. A. *102*, 14765–14770.

Mirsky, A.E., and Ris, H. (1947). The chemical composition of isolated chromosomes. J. Gen. Physiol. *31*, 7–18.

Mohammed, H., D'Santos, C., Serandour, A.A., Ali, H.R., Brown, G.D., Atkins, A., Rueda, O.M., Holmes, K.A., Theodorou, V., Robinson, J.L.L., et al. (2013). Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. Cell Rep. *3*, 342–349.

Mohammed, H., Taylor, C., Brown, G.D., Papachristou, E.K., Carroll, J.S., and D'Santos, C.S. (2016). Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. Nat. Protoc. *11*, 316–326.

Mohd-Sarip, A., Venturini, F., Chalkley, G.E., and Verrijzer, C.P. (2002). Pleiohomeotic can link polycomb to DNA and mediate transcriptional repression. Mol. Cell. Biol. *22*, 7473–7483.

Möhrle, B.P., Kumpf, M., and Gauglitz, G. (2005). Determination of affinity constants of locked nucleic acid (LNA) and DNA duplex formation using label free sensor technology. Analyst *130*, 1634–1638.

Morris, S.A., Baek, S., Sung, M.-H., John, S., Wiench, M., Johnson, T.A., Schiltz, R.L., and Hager, G.L. (2014). Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. Nat. Struct. Mol. Biol. *21*, 73–81.

Morrison, E.A., Sanchez, J.C., Ronan, J.L., Farrell, D.P., Varzavand, K., Johnson, J.K., Gu, B.X., Crabtree, G.R., and Musselman, C.A. (2017). DNA binding drives the association of BRG1/hBRM bromodomains with nucleosomes. Nat. Commun. *8*, 1–14.

Mujtaba, S., He, Y., Zeng, L., Yan, S., Plotnikova, O., Sachchidanand, Sanchez, R., Zeleznik-Le, N.J., Ronai, Z., and Zhou, M.-M. (2004). Structural Mechanism of the Bromodomain of the Coactivator CBP in p53 Transcriptional Activation. Mol. Cell *13*, 251–263.

Müller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J. a (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell *111*, 197–208.

Muraoka, M., Konishi, M., Kikuchi-Yanoshita, R., Tanaka, K., Shitara, N., Chong, J.M., Iwama, T., and Miyaki, M. (1996). p300 gene alterations in colorectal and gastric carcinomas. Oncogene *12*, 1565–1569.

Murray, K. (1964). The Occurrence Of Epsilon-N-Methyl Lysine In Histones. Biochemistry *3*, 10–15.

Musselman, C.A., Mansfield, R.E., Garske, A.L., Davrazou, F., Kwan, A.H., Oliver, S.S., O'Leary, H., Denu, J.M., Mackay, J.P., and Kutateladze, T.G. (2009). Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. Biochem. J. *423*, 179–187.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110–113.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature

*393*, 386–389.
Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Köcher, T., Cohen, A., Stunnenberg, H.G., Wilm, M., and Müller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. EMBO J. *26*, 4078–4088.
Nightingale, K.P., Wellinger, R.E., Sogo, J.M., and Becker, P.B. (1998). Histone acetylation facilitates RNA polymerase II transcription of the Drosophila hsp26 gene in chromatin. EMBO J. *17*, 2865–2876.
van Nuland, R., Smits, A.H., Pallaki, P., Jansen, P.W.T.C., Vermeulen, M., and Timmers, H.T.M. (2013). Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. Mol. Cell. Biol.
Ogryzko, V. V., Schiltz, R.L., Russanova, V., Howard, B.H., and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell *87*, 953–959.
Ogryzko, V. V, Kotani, T., Zhang, X., Schiltz, R.L., Howard, T., Yang, X.J., Howard, B.H., Qin, J., and Nakatani, Y. (1998). Histone-like TAFs within the PCAF histone acetylase complex. Cell *94*, 35–44.
Oike, Y., Takakura, N., Hata, A., Kaname, T., Akizuki, M., Yamaguchi, Y., Yasue, H., Araki, K., Yamamura, K.I., and Suda, T. (1999). Mice homozygous for a truncated form of CREB-binding protein exhibit defects in hematopoiesis and vasculo-angiogenesis. Blood *93*, 2771–2779.
Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat. Genet. *19*, 219–220.
Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.
Oktaba, K., Gutiérrez, L., Gagneur, J., Girardot, C., Sengupta, A.K., Furlong, E.E.M., and Müller, J. (2008). Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in Drosophila. Dev. Cell *15*, 877–889.
Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714–717.
Ortega, E., Rengachari, S., Ibrahim, Z., Hoghoughi, N., Gaucher, J., Holehouse, A.S., Khochbin, S., and Panne, D. (2018). Transcription factor dimerization activates the p300 acetyltransferase. Nature *562*, 538–544.
Owen, D.J., Ornaghi, P., Yang, J.C., Lowe, N., Evans, P.R., Ballario, P., Neuhaus, D., Filetici, P., and Travers, A.A. (2000). The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. EMBO J. *19*, 6141–6149.
Papachristou, E.K., Kishore, K., Holding, A.N., Harvey, K., Roumeliotis, T.I., Chilamakuri, C.S.R., Omarjee, S., Chia, K.M., Swarbrick, A., Lim, E., et al. (2018). A quantitative mass spectrometry-based approach to monitor the dynamics of endogenous chromatin-associated protein complexes. Nat. Commun. *9*, 2311.
Park, S., Stanfield, R.L., Martinez-Yamout, M.A., Dyson, H.J., Wilson, I.A., and Wright, P.E. (2017). Role of the CBP catalytic core in intramolecular SUMOylation and control of histone H3 acetylation. Proc. Natl. Acad. Sci.

U. S. A. *114*, E5335–E5342.
Park, S.H., Ayoub, A., Lee, Y., Xu, J., Kim, H., Zheng, W., Zhang, B., Sha, L., An, S., Zhang, Y., et al. (2019). Cryo-EM structure of the human MLL1 core complex bound to the nucleosome. Nat. Commun. *10*, 5540.
Parthun, M.R., Widom, J., and Gottschling, D.E. (1996). The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. Cell *87*, 85–94.
Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. EMBO J. *23*, 4061–4071.
Pasini, D., Malatesta, M., Jung, H.R., Walfridsson, J., Willer, A., Olsson, L., Skotte, J., Wutz, A., Porse, B., Jensen, O.N., et al. (2010). Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. Nucleic Acids Res. *38*, 4958–4969.
Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P., et al. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature *473*, 394–397.
Pengelly, A.R., Copur, Ö., Jäckle, H., Herzig, A., and Müller, J. (2013). A histone mutant reproduces the phenotype caused by loss of histone-modifying factor polycomb. Science (80-. ). *339*, 698–699.
Pengelly, A.R., Kalb, R., Finkl, K., and Müller, J. (2015). Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. Genes Dev. *29*, 1487–1492.
Petrij, F., Giles, R.H., Dauwerse, H.G., Saris, J.J., Hennekam, R.C., Masuno, M., Tommerup, N., van Ommen, G.J., Goodman, R.H., and Peters, D.J. (1995). Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. Nature *376*, 348–351.
Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281–1295.
Ponting, C.P., Blake, D.J., Davies, K.E., Kendrick-Jones, J., and Winder, S.J. (1996). ZZ and TAZ: New putative zinc fingers in dystrophin and other proteins. Trends Biochem. Sci. *21*, 11–13.
Presman, D.M., Ball, D.A., Paakinaho, V., Grimm, J.B., Lavis, L.D., Karpova, T.S., and Hager, G.L. (2017). Quantifying transcription factor binding dynamics at the single-molecule level in live cells. Methods *123*, 76–88.
Prior, C.P., Cantor, C.R., Johnson, E.M., Littau, V.C., and Allfrey, V.G. (1983). Reversible changes in nucleosome structure and histone H3 accessibility in transcriptionally active and inactive states of rDNA chromatin. Cell *34*, 1033–1042.
Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.
Raisner, R., Kharbanda, S., Jin, L., Jeng, E., Chan, E., Merchant, M., Haverty, P.M., Bainer, R., Cheung, T., Arnott, D., et al. (2018). Enhancer Activity Requires CBP/P300 Bromodomain-Dependent Histone H3K27 Acetylation. Cell Rep. *24*, 1722–1729.
Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., and Sweet, R.M. (1993). Crystal structure of globular domain of histone H5 and its implications for

nucleosome binding. Nature *362*, 219–223.
Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A Unifying Model for the Selective Regulation of Inducible Transcription by CpG Islands and Nucleosome Remodeling. Cell *138*, 114–128.
Rappsilber, J. (2011). The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. J. Struct. Biol. *173*, 530–540.
Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593–599.
Rea, S., Xouri, G., and Akhtar, A. (2007). Males absent on the first (MOF): from flies to humans. Oncogene *26*, 5385–5394.
Reid, J.L., Iyer, V.R., Brown, P.O., and Struhl, K. (2000). Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of Esa1 histone acetylase. Mol. Cell *6*, 1297–1307.
Ren, X., and Kerppola, T.K. (2011). REST interacts with Cbx proteins and regulates polycomb repressive complex 1 occupancy at RE1 elements. Mol. Cell. Biol. *31*, 2100–2110.
Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG Islands genome wide. Mol. Cell *55*, 347–360.
Rikitake, Y., and Moran, E. (1992). DNA-binding properties of the E1A-associated 300-kilodalton protein. Mol. Cell. Biol. *12*, 2826–2836.
Robinson, P.J.J., and Rhodes, D. (2006). Structure of the "30 nm" chromatin fibre: a key role for the linker histone. Curr. Opin. Struct. Biol. *16*, 336–343.
Roeder, R.G. (2019). 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. Nat. Struct. Mol. Biol. *26*, 783–791.
Roeder, R.G., and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature *224*, 234–237.
Roguev, a, Schaft, D., Shevchenko, a, Pijnappel, W.W., Wilm, M., Aasland, R., and Stewart, a F. (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. EMBO J. *20*, 7137–7148.
Rose, N.R., King, H.W., Blackledge, N.P., Fursova, N.A., Ember, K.J., Fischer, R., Kessler, B.M., and Klose, R.J. (2016). RYBP stimulates PRC1 to shape chromatin-based communication between polycomb repressive complexes. Elife *5*.
Saksouk, N., Barth, T.K., Ziegler-Birling, C., Olova, N., Nowak, A., Rey, E., Mateos-Langerak, J., Urbach, S., Reik, W., Torres-Padilla, M.-E., et al. (2014). Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. Mol. Cell *56*, 580–594.
Sauer, F., and Jäckle, H. (1991). Concentration-dependent transcriptional activation or repression by Krüppel from a single binding site. Nature *353*, 563–566.
Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc. Natl. Acad. Sci. U. S. A. *103*, 1412–1417.
Scheuermann, J.C., Gutiérrez, L., and Müller, J. (2012). Histone H2A

monoubiquitination and Polycomb repression: The missing pieces of the puzzle. Fly (Austin). *6*, 162–168.
Schmidl, C., Klug, M., Boeld, T.J., Andreesen, R., Hoffmann, P., Edinger, M., and Rehli, M. (2009). Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. Genome Res. *19*, 1165–1174.
Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. Cell *42*, 330–341.
Schmitt, S., Prestel, M., and Paro, R. (2005). Intergenic transcription through a polycomb group response element counteracts silencing. Genes Dev. *19*, 697–708.
Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.-M., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M., et al. (2015). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. Nat. Genet. *47*, 1179–1186.
Schuettengruber, B., Oded Elkayam, N., Sexton, T., Entrevan, M., Stern, S., Thomas, A., Yaffe, E., Parrinello, H., Tanay, A., and Cavalli, G. (2014). Cooperativity, specificity, and evolutionary stability of Polycomb targeting in Drosophila. Cell Rep. *9*, 219–233.
Schwartz, Y.B., Kahn, T.G., Nix, D. a, Li, X.-Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat. Genet. *38*, 700–705.
Sealy, L., and Chalkley, R. (1978). DNA associated with hyperacetylated histone is preferentially digested by DNase I. Nucleic Acids Res. *5*, 1863–1876.
Sebé-Pedrós, A., De Mendoza, A., Lang, B.F., Degnan, B.M., and Ruiz-Trillo, I. (2011). Unexpected repertoire of metazoan transcription factors in the unicellular holozoan capsaspora owczarzaki. Mol. Biol. Evol. *28*, 1241–1254.
Sedkov, Y., Benes, J.J., Berger, J.R., Riker, K.M., Tillib, S., Jones, R.S., and Mazo, A. (1999). Molecular genetic analysis of the Drosophila trithorax-related gene which encodes a novel SET domain protein. Mech. Dev. *82*, 171–179.
Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R. a, Young, R. a, and Sharp, P. a (2008). Divergent transcription from active promoters. Science *322*, 1849–1851.
Seto, E., and Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. Cold Spring Harb. Perspect. Biol. *6*, a018713.
Shahbazian, M.D., and Grunstein, M. (2007). Functions of Site-Specific Histone Acetylation and Deacetylation. Annu. Rev. Biochem. *76*, 75–100.
Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C.T., Bender, W., and Kingston, R.E. (1999). Stabilization of chromatin structure by PRC1, a Polycomb complex. Cell *98*, 37–46.
Sharif, J., Endo, T. a, Ito, S., Ohara, O., and Koseki, H. (2013). Embracing change to remain the same: conservation of polycomb functions despite divergence of binding motifs among species. Curr. Opin. Cell Biol. *25*, 305–313.
Shen, W., Xu, C., Huang, W., Zhang, J., Carlson, J.E., Tu, X., Wu, J., and Shi, Y. (2007). Solution structure of human Brg1 bromodomain and its specific binding to acetylated histone tails. Biochemistry *46*, 2100–2110.
Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases:

mechanisms of regulation in development and disease pathogenesis. Annu. Rev. Biochem. *81*, 65–95.

Shirai, A., Kawaguchi, T., Shimojo, H., Muramatsu, D., Ishida-Yonetani, M., Nishimura, Y., Kimura, H., Nakayama, J.I., and Shinkai, Y. (2017). Impact of nucleic acid and methylated H3K9 binding activities of Suv39h1 on its heterochromatin assembly. Elife *6*.

Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844–847.

Shpargel, K.B., Sengoku, T., Yokoyama, S., and Magnuson, T. (2012). UTX and UTY demonstrate histone demethylase-independent function in mouse embryonic development. PLoS Genet. *8*, e1002964.

Simó-Riudalbas, L., Pérez-Salvia, M., Setien, F., Villanueva, A., Moutinho, C., Martínez-Cardús, A., Moran, S., Berdasco, M., Gomez, A., Vidal, E., et al. (2015). KAT6B Is a Tumor Suppressor Histone H3 Lysine 23 Acetyltransferase Undergoing Genomic Loss in Small Cell Lung Cancer. Cancer Res. *75*, 3936–3945.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nat. Rev. Mol. Cell Biol. *10*, 697–708.

Singh, M., Popowicz, G.M., Krajewski, M., and Holak, T.A. (2007). Structural ramification for acetyl-lysine recognition by the bromodomain of human BRG1 protein, a central ATPase of the SWI/SNF remodeling complex. Chembiochem *8*, 1308–1316.

Song, C.Z., Keller, K., Chen, Y., Murata, K., and Stamatoyannopoulos, G. (2002). Transcription coactivator CBP has direct DNA binding activity and stimulates transcription factor DNA binding through small domains. Biochem. Biophys. Res. Commun. *296*, 118–124.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490–495.

Stanton, B.Z., Hodges, C., Calarco, J.P., Braun, S.M.G., Ku, W.L., Kadoch, C., Zhao, K., and Crabtree, G.R. (2017). Smarca4 ATPase mutations disrupt direct eviction of PRC1 from chromatin. Nat. Genet. *49*, 282–288.

Sterner, D.E., and Berger, S.L. (2000). Acetylation of histones and transcription-related factors. Microbiol. Mol. Biol. Rev. *64*, 435–459.

Struhl, G. (1981). A gene product required for correct initiation of segmental determination in Drosophila. Nature *293*, 36–41.

Struhl, G., and Akam, M. (1985). Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. EMBO J. *4*, 3259–3264.

Suhara, W., Yoneyama, M., Kitabayashi, I., and Fujita, T. (2002). Direct involvement of CREB-binding protein/p300 in sequence-specific DNA binding of virus-activated interferon regulatory factor-3 holocomplex. J. Biol. Chem. *277*, 22304–22313.

Suka, N., Suka, Y., Carmen, A.A., Wu, J., and Grunstein, M. (2001). Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. Mol. Cell *8*, 473–479.

Sun, J., Paduch, M., Kim, S.A., Kramer, R.M., Barrios, A.F., Lu, V., Luke, J., Usatyuk, S., Kossiakoff, A.A., and Tan, S. (2018). Structural basis for activation of SAGA histone acetyltransferase Gcn5 by partner subunit Ada2. Proc. Natl. Acad. Sci. U. S. A. *115*, 10010–10015.

Swinstead, E.E., Miranda, T.B., Paakinaho, V., Baek, S., Goldstein, I., Hawkins, M., Karpova, T.S., Ball, D., Mazza, D., Lavis, L.D., et al. (2016). Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. Cell *165*, 593–605.

Szerlong, H.J., and Hansen, J.C. (2011). Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structureThis paper is one of a selection of papers published in a Special Issue entitled 31st Annual International Asilomar Chromatin and Chromosomes Conference, and. Biochem. Cell Biol. *89*, 24–34.

Taherbhoy, A.M., Huang, O.W., and Cochran, A.G. (2015). BMI1–RING1B is an autoinhibited RING E3 ubiquitin ligase. Nat. Commun. *6*, 7621.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930–935.

Takai, H., Masuda, K., Sato, T., Sakaguchi, Y., Suzuki, T., Suzuki, T., Koyama-Nasu, R., Nasu-Nishimura, Y., Katou, Y., Ogawa, H., et al. (2014). 5-Hydroxymethylcytosine plays a critical role in glioblastomagenesis by recruiting the CHTOP-methylosome complex. Cell Rep. *9*, 48–60.

Tanaka, Y., Naruse, I., Maekawa, T., Masuya, H., Shiroishi, T., and Ishii, S. (1997). Abnormal skeletal patterning in embryos lacking a single Cbp allele: A partial similarity with Rubinstein-Taybi syndrome. Proc. Natl. Acad. Sci. U. S. A. *94*, 10215–10220.

Tanner, K.G., Trievel, R.C., Kuo, M.H., Howard, R.M., Berger, S.L., Allis, C.D., Marmorstein, R., and Denu, J.M. (1999). Catalytic mechanism and function of invariant glutamic acid 173 from the histone acetyltransferase GCN5 transcriptional coactivator. J. Biol. Chem. *274*, 18157–18160.

Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science *272*, 408–411.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., et al. (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell *148*, 664–678.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. *14*, 1025–1040.

Tazi, J., and Bird, A. (1990). Alternative chromatin structure at CpG islands. Cell *60*, 909–920.

Thakur, J.K., Yadav, A., and Yadav, G. (2014). Molecular recognition by the KIX domain and its role in gene regulation. Nucleic Acids Res. *42*, 2112–2125.

Thompson, J.P., Granoff, A., and Willis, D.B. (1986). Trans-activation of a methylated adenovirus promoter by a frog virus 3 protein. Proc. Natl. Acad. Sci. U. S. A. *83*, 7688–7692.

Thompson, J.P., Granoff, A., and Willis, D.B. (1988). Methylation of the promoter for an immediate-early frog virus 3 gene does not inhibit transcription. J. Virol. *62*, 4680–4685.

Thompson, P.R., Wang, D., Wang, L., Fulco, M., Pediconi, N., Zhang, D., An, W., Ge, Q., Roeder, R.G., Wong, J., et al. (2004). Regulation of the p300 HAT domain via a novel activation loop. Nat. Struct. Mol. Biol. *11*, 308–315.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W.W., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082–1086.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development *136*, 3131–3141.

Trievel, R.C., Rojas, J.R., Sterner, D.E., Venkataramani, R.N., Wang, L., Zhou, J., Allis, C.D., Berger, S.L., and Marmorstein, R. (1999). Crystal structure and mechanism of histone acetylation of the yeast GCN5 transcriptional coactivator. Proc. Natl. Acad. Sci. U. S. A. *96*, 8931–8936.

Tropberger, P., Pott, S., Keller, C., Kamieniarz-Gdula, K., Caron, M., Richter, F., Li, G., Mittler, G., Liu, E.T., Bühler, M., et al. (2013). Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. Cell *152*, 859–872.

Tropea, J.E., Cherry, S., and Waugh, D.S. (2009). Expression and purification of soluble His6-tagged TEV protease. Methods Mol. Biol.

Turner, B.M. (2000). Histone acetylation and an epigenetic code. BioEssays *22*, 836–845.

Tzelepis, K., Koike-Yusa, H., De Braekeleer, E., Li, Y., Metzakopian, E., Dovey, O.M., Mupo, A., Grinkevich, V., Li, M., Mazan, M., et al. (2016). A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. Cell Rep. *17*, 1193–1205.

Ullah, M., Pelletier, N., Xiao, L., Zhao, S.P., Wang, K., Degerny, C., Tahmasebi, S., Cayrou, C., Doyon, Y., Goh, S.-L., et al. (2008). Molecular architecture of quartet MOZ/MORF histone acetyltransferase complexes. Mol. Cell. Biol. *28*, 6828–6843.

Vaquero, A., Scher, M., Lee, D., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2004). Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. Mol. Cell *16*, 93–105.

Vermeulen, M., Carrozza, M.J., Lasonder, E., Workman, J.L., Logie, C., and Stunnenberg, H.G. (2004). In vitro targeting reveals intrinsic histone tail specificity of the Sin3/histone deacetylase and N-CoR/SMRT corepressor complexes. Mol. Cell. Biol. *24*, 2364–2372.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W.M.P., van Schaik, F.M. a, Varier, R. a, Baltissen, M.P. a, Stunnenberg, H.G., Mann, M., and Timmers, H.T.M. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58–69.

Vettese-Dadey, M., Grant, P.A., Hebbes, T.R., Crane- Robinson, C., Allis, C.D., and Workman, J.L. (1996). Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. EMBO J. *15*, 2508–2518.

Vidal, M., and Gaber, R.F. (1991). RPD3 encodes a second factor required to achieve maximum positive and negative transcriptional states in Saccharomyces cerevisiae. Mol. Cell. Biol. *11*, 6317–6327.

Vidali, G., Boffa, L.C., Bradbury, E.M., and Allfrey, V.G. (1978). Butyrate

suppression of histone deacetylation leads to accumulation of multiacetylated forms of histones H3 and H4 and increased DNase I sensitivity of the associated DNA sequences. Proc. Natl. Acad. Sci. U. S. A. *75*, 2239–2243.

Voigt, P., LeRoy, G., Drury, W.J., Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181–193.

Voo, K.S., Carlone, D.L., Jacobsen, B.M., Flodin, A., and Skalnik, D.G. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol. Cell. Biol. *20*, 2108–2121.

Wachter, E., Quante, T., Merusi, C., Arczewska, A., Stewart, F., Webb, S., and Bird, A. (2014). Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. Elife *3*, e03397.

Wade, P.A., Pruss, D., and Wolffe, A.P. (1997). Histone acetylation: chromatin in action. Trends Biochem. Sci. *22*, 128–132.

Walker, E., Chang, W.Y., Hunkapiller, J., Cagney, G., Garcha, K., Torchia, J., Krogan, N.J., Reiter, J.F., and Stanford, W.L. (2010). Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. Cell Stem Cell *6*, 153–166.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004a). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873–878.

Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. (2004b). Hierarchical recruitment of polycomb group silencing complexes. Mol. Cell *14*, 637–646.

Wang, S.-P., Tang, Z., Chen, C.-W., Shimada, M., Koche, R.P., Wang, L.-H., Nakadai, T., Chramiec, A., Krivtsov, A. V, Armstrong, S.A., et al. (2017a). A UTX-MLL4-p300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer Landscapes for Eliciting Transcription. Mol. Cell *67*, 308-321.e6.

Wang, X., Paucek, R.D., Gooding, A.R., Brown, Z.Z., Ge, E.J., Muir, T.W., and Cech, T.R. (2017b). Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. Nat. Struct. Mol. Biol. *24*, 1028–1038.

Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009). Genome-wide Mapping of HATs and HDACs Reveals Distinct Functions in Active and Inactive Genes. Cell *138*, 1019–1031.

Waters, L., Yue, B., Veverka, V., Renshaw, P., Bramham, J., Matsuda, S., Frenkiel, T., Kelly, G., Muskett, F., Carr, M., et al. (2006). Structural diversity in p160/CREB-binding protein coactivator complexes. J. Biol. Chem. *281*, 14787–14795.

Weinert, B.T., Narita, T., Satpathy, S., Srinivasan, B., Hansen, B.K., Schölz, C., Hamilton, W.B., Zucconi, B.E., Wang, W.W., Liu, W.R., et al. (2018). Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/p300 Acetylome. Cell *174*, 231-244.e12.

Whyte, P., Williamson, N.M., and Harlow, E. (1989). Cellular targets for transformation by the adenovirus E1A proteins. Cell *56*, 67–75.

Williams, K., Christensen, J., and Helin, K. (2012). DNA methylation: TET proteins-guardians of CpG islands? EMBO Rep. *13*, 28–35.

Wojciak, J.M., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2009). Structural basis for recruitment of CBP/p300 coactivators by STAT1 and STAT2 transactivation domains. EMBO J. *28*, 948–958.

Worden, E.J., Hoffmann, N.A., Hicks, C.W., and Wolberger, C. (2019). Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. Cell *176*, 1490-1501.e12.

Wu, H., and Zhang, Y. (2011). Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells. Cell Cycle *10*, 2428–2436.

Wu, M., Wang, P.F., Lee, J.S., Martin-Brown, S., Florens, L., Washburn, M., and Shilatifard, A. (2008). Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Mol. Cell. Biol. *28*, 7337–7344.

Wu, X., Johansen, J.V., and Helin, K. (2013). Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. Mol. Cell *1*, 1–13.

Wysocka, J., Swigut, T., Xiao, H., Milne, T. a, Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., et al. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86–90.

Xu, C., Bian, C., Lam, R., Dong, A., and Min, J. (2011). The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain. Nat. Commun. *2*, 227.

Xu, W., Edmondson, D.G., Evrard, Y.A., Wakamiya, M., Behringer, R.R., and Roth, S.Y. (2000). Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development. Nat. Genet. *26*, 229–232.

Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J.G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., et al. (2012). Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. Cell *151*, 1200–1213.

Yamauchi, T., Yamauchi, J., Kuwata, T., Tamura, T., Yamashita, T., Bae, N., Westphal, H., Ozato, K., and Nakatani, Y. (2000). Distinct but overlapping roles of histone acetylase PCAF and of the closely related PCAF-B/GCN5 in mouse embryogenesis. Proc. Natl. Acad. Sci. U. S. A. *97*, 11303–11306.

Yan, K., Rousseau, J., Littlejohn, R.O., Kiss, C., Lehman, A., Rosenfeld, J.A., Stumpel, C.T.R., Stegmann, A.P.A., Robak, L., Scaglia, F., et al. (2017). Mutations in the Chromatin Regulator Gene BRPF1 Cause Syndromic Intellectual Disability and Deficient Histone Acetylation. Am. J. Hum. Genet. *100*, 91–104.

Yan, Y., Barlev, N.A., Haley, R.H., Berger, S.L., and Marmorstein, R. (2000). Crystal structure of yeast Esa1 suggests a unified mechanism for catalysis and substrate binding by histone acetyltransferases. Mol. Cell *6*, 1195–1205.

Yan, Y., Harper, S., Speicher, D.W., and Marmorstein, R. (2002). The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate. Nat. Struct. Biol. *9*, 862–869.

Yang, X.J. (2004). The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. Nucleic Acids Res. *32*, 959–976.

Yang, X.J. (2015). MOZ and MORF acetyltransferases: Molecular interaction, animal development and human disease. Biochim. Biophys. Acta - Mol.

Cell Res. *1853*, 1818–1826.
Yang, W.M., Inouye, C., Zeng, Y., Bearss, D., and Seto, E. (1996). Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. Proc. Natl. Acad. Sci. U. S. A. *93*, 12845–12850.
Yao, T.P., Oh, S.P., Fuchs, M., Zhou, N.D., Ch'ng, L.E., Newsome, D., Bronson, R.T., Li, E., Livingston, D.M., and Eckner, R. (1998). Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. Cell *93*, 361–372.
Ye, J., Ai, X., Eugeni, E.E., Zhang, L., Carpenter, L.R., Jelinek, M.A., Freitas, M.A., and Parthun, M.R. (2005). Histone H4 lysine 91 acetylation a core domain modification associated with chromatin assembly. Mol. Cell *18*, 123–130.
Yee, S.P., and Branton, P.E. (1985). Detection of cellular proteins associated with human adenovirus type 5 early region 1A polypeptides. Virology *147*, 142–153.
Yu, B.D., Hess, J.L., Horning, S.E., Brown, G.A., and Korsmeyer, S.J. (1995). Altered Hox expression and segmental identity in Mll-mutant mice. Nature *378*, 505–508.
Yuan, L.W., and Giordano, A. (2002). Acetyltransferase machinery conserved in p300/CBP-family proteins. Oncogene *21*, 2253–2260.
Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T., et al. (2012). Dense chromatin activates polycomb repressive complex 2 to regulate H3 lysine 27 methylation. Science (80-. ). *337*, 971–975.
Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355–364.
Zentner, G.E., and Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. Nat. Struct. Mol. Biol. *20*, 259–266.
Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A.P., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. *13*, 1924–1935.
Zhang, Y., Mittal, A., Reid, J., Reich, S., Gamblin, S.J., and Wilson, J.R. (2015). Evolving catalytic properties of the MLL family SET domain. Structure *23*, 1921–1933.
Zhang, Y., Xue, Y., Shi, J., Ahn, J.W., Mi, W., Ali, M., Wang, X., Klein, B.J., Wen, H., Li, W., et al. (2018). The ZZ domain of p300 mediates specificity of the adjacent HAT domain for histone H3. Nat. Struct. Mol. Biol. *25*, 841–849.
Zhou, J.C., Blackledge, N.P., Farcas, A.M., and Klose, R.J. (2012). Recognition of CpG island chromatin by KDM2A requires direct and specific interaction with linker DNA. Mol. Cell. Biol. *32*, 479–489.
Zuber, J., Shi, J., Wang, E., Rappaport, A.R., Herrmann, H., Sison, E.A., Magoon, D., Qi, J., Blatt, K., Wunderlich, M., et al. (2011). RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature *478*, 524–528.
Van Zundert, G.C.P., Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastritis, P.L., Karaca, E., Melquiond, A.S.J., Van Dijk, M., De Vries, S.J., and Bonvin, A.M.J.J. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J. Mol. Biol. *428*, 720–

725.
Zwart, W., Theodorou, V., Kok, M., Canisius, S., Linn, S., and Carroll, J.S. (2011). Oestrogen receptor-co-factor chromatin specificity in the transcriptional regulation of breast cancer. EMBO J. *30*, 4764–4776.