The H2A.Z and PWWP2A associated NuRD interactor HMG20A controls transcriptional programs in head and heart development

A thesis presented for the degree of Doctor rerum naturalium (Dr.rer.nat)

by

Andreas Herchenröther

submitted to the Institute for Genetics, Justus-Liebig-University Giessen

1 Preamble

Dean:

Prof. Dr Thomas Wilke,
Department of Animal Ecology & Systematics,
Justus Liebig University Giessen

First Referee:

Prof. Dr. Sandra B. Hake,
Institute for Genetics,
Justus Liebig University Giessen

Second Referee:

Prof. Dr. Alexander Brehm,
Institute for Molecular Biology & Tumor Research
Phillips University Marburg

1.1 Curriculum Vitae

Publications associated with this dissertation

Herchenröther, A., Gossen, S., Friedrich, T., Reim, A., Daus, N., Diegmüller, F., Leers, J., Sani, H. M., Gerstner, S., Schwarz, L., et al. (2023a). The H2A.Z and NuRD associated protein HMG20A controls early head and heart developmental transcription programs. *Nat Commun* 14, 472.

Herchenröther, A., Wunderlich, T. M., Lan, J. and Hake, S. B. (2023b). Spotlight on histone H2A variants: From B to X to Z. *Semin Cell Dev Biol* 135, 3–12.

Giaimo, B. D., Ferrante, F., Herchenröther, A., Hake, S. B. and Borggrefe, T. (2019). The histone variant H2A.Z in gene regulation. *Epigenetics & Chromatin* 12, 37–37.

1.2 Eidesstattliche Versicherung

Ich erkläre: Ich habe die vorgelegte Dissertation selbstständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Ich stimme einer evtl. Überprüfung meiner Dissertation durch eine Antiplagiat-Software zu. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der 'Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis' niedergelegt sind, eingehalten.

——————————————— ——————————————————

Ort, Datum (Andreas Herchenröther)

1.3 Abstract/ Summary

The inheritable information of all eukaryotic organisms is organized in DNA-protein complexes called chromatin. Dedicated chromatin-binding proteins are required for DNA-based processes during development. The recently established direct histone variant H2A.Z interactor PWWP2A is involved in craniofacial development. During my PhD work, I identified the H2A.Z/PWWP2A-associated High mobility group protein 20A (HMG20A) as part of several chromatin-modifying complexes, including the nucleosome remodeling and deacetylase (NuRD) complex, and showed its localization to distinct genomic regulatory regions. Furthermore, HMG20A depletion causes severe head and heart developmental defects in *Xenopus laevis*. Data gathered here suggest that craniofacial malformations are caused by defects in neural crest cell (NCC) migration and cartilage formation. These developmental defects are replicated in HMG20A-depleted mouse embryonic stem cells (mESCs), which show inefficient differentiation into NCCs and cardiomyocytes (CM). Loss of HMG20A, which marks promoters and enhancers, results in chromatin accessibility changes and a striking deregulation of transcription programs involved in epithelial-mesenchymal transition (EMT) and differentiation processes. Collectively, my study implicates HMG20A as part of the H2A.Z/PWWP2A/NuRD-axis and reveal it as a key modulator of sophisticated developmental transcription programs that guide the differentiation of NCCs and CMs.

1.4 Zusammenfassung

In eukaryontischen Organismen ist die vererbbare Information in DNA-Protein-Komplexen, dem sogenannten Chromatin, organisiert. Spezielle Chromatin-bindende Proteine sind für DNA-basierte Prozesse während der Entwicklung erforderlich. Das kürzlich entdeckte PWWP2A Protein, das die Histonvariante H2A.Z direkt binden kann, ist an der Kopfentwicklung beteiligt. Darauf aufbauend meine Arbeit das H2A.Z/PWWP2A-assoziierte Protein HMG20A als Teil mehrerer Chromatin-modifizierender Komplexe, einschließlich dem Nucleosome Remodelling and Deacetylase (NuRD) Komplex, sowie seine Lokalisierung in verschiedenen regulatorischen genomischen Regionen identifizieren. Darüber hinaus führt die Deletion von HMG20A in *Xenopus laevis* zu schweren Entwicklungsstörungen von Kopf und Herz. Die hier gesammelten Daten deuten darauf hin, dass kraniofaziale Fehlbildungen durch Defekte bei der Migration von Neuralleistenzellen (NCCs) und folglich der Knorpelbildung verursacht werden. Diese Entwicklungsdefekte wurden in Hmg20A-depletierten embryonalen Stammzellen der Maus (mESCs) repliziert, die eine ineffiziente Differenzierung in NCCs und Kardiomyozyten (CM) zeigen. Der Verlust von HMG20A, das offene Promotoren und Enhancer bindet, führt zu Veränderungen der Chromatinzugänglichkeit und einer starken Deregulierung von Transkriptionsprogrammen, die an der Epithelial-mesenchymalen Transition (EMT) und an Differenzierungsprozessen beteiligt sind. Insgesamt zeigt meine Studie, dass HMG20A ein Teil der H2A.Z/PWWP2A/NuRD-Achse und ein wichtiger Modulator von fein abgestimmten Transkriptionsprogrammen ist, die die Differenzierung von NCCs und CMs steuern.

1 PREAMBLE ..... **A**

1.1 CURRICULUM VITAE ..... **B**
1.2 EIDESSTATTLICHE VERSICHERUNG ..... **C**
1.3 ABSTRACT / SUMMARY ..... **E**
1.4 ZUSAMMENFASSUNG ..... **F**
1.5 LIST OF ABBREVIATIONS ..... **L**

2 INTRODUCTION **1**

2.1 CHROMATIN STRUCTURE AND ITS IMPLICATIONS ON TRANSCRIPTIONAL REGULATION **1**
2.1.1 CHROMATIN DENSITY IS RELATED TO TRANSCRIPTIONAL ACTIVITY 2
2.1.2 CHROMATIN-ASSOCIATED RNAS INFLUENCE ITS ORGANIZATION 3
2.1.3 5'METHYL-CYTOSINE REPRESSES TRANSCRIPTION 4
2.1.4 HISTONE ACETYLATION AND METHYLATION MODIFY TRANSCRIPTION CONTEXT-DEPENDENT 4
2.1.5 CHROMATIN REMODELERS ALTER CHROMATIN STRUCTURE AND DETERMINE DNA ACCESSIBILITY 7
2.1.5.1 Paralogs of Nucleosome and deacetylase complex (NuRD) members determine its function in gene regulation 9
2.1.6 THE HISTONE VARIANT H2A.Z 10
2.1.7 PHF14, RAI1, TCF20 & HMG20A IN GENE REGULATION 16
2.2 AIM OF THIS STUDY **19**

3 MATERIAL **20**

3.1 LABORATORY EQUIPMENT AND SOFTWARE **20**
3.2 CONSUMABLES **23**
3.3 CHEMICALS **24**
3.4 COMMERCIAL KITS **26**
3.5 ANTIBODIES **27**

3.6 DNA OLIGONUCLEOTIDES FOR (QUANTITATIVE) POLYMERASE CHAIN REACTION PRIMER AND CRISPR/CAS9 GUIDE RNA SEQUENCES 31

3.7 GENERAL BUFFERS 34

4 METHODS 37

4.1 CELL PROPAGATION, CELL TRANSFECTION, AND CELL BIOLOGICAL ANALYSIS 37

4.1.1 CULTURING, PASSAGING, FREEZING AND THAWING OF CELLS 37

4.1.1.1 Hela cells 37

4.1.1.2 Mouse embryonic stem cells (mESC) 38

4.1.1.2.1 Differentiation protocols for mouse embryonic stem cells 40

4.1.1.2.1.1 Neural crest cell differentiation and migration 40

4.1.1.2.1.2 Cardiomyocyte differentiation 40

4.1.1.3 Sf9 cells 41

4.1.2 TRANSFECTION OF PLASMIDS TO ECTOPICALLY EXPRESS FUSION PROTEINS 41

4.1.2.1 Hela cells 41

4.1.2.2 Mouse embryonic stem cells (mESC) 41

4.1.2.3 Sf9 cells 42

4.1.3 GENERATION OF HELA CELL LINES THAT ECTOPICALLY EXPRESS FUSION PROTEINS 42

4.1.4 TRANSFECTION OF SIRNAS TO DEPLETE SPECIFIC MRNAS 42

4.1.5 ENDOGENOUS PROTEIN DEPLETION IN MOUSE EMBRYONIC STEM CELLS 42

4.1.6 IMMUNOFLUORESCENCE (IF) MICROSCOPY 43

4.1.7 FLOW CYTOMETRY ANALYSIS 44

4.1.7.1 Flow cytometry analysis to ensure purity of cells expressing GFP fusion protein 44

4.2 DNA-BASED METHODS 45

4.2.1 RESTRICTION ENZYME-BASED CLONING 45

4.2.2 RECOMBINATION-BASED CLONING (DNA ASSEMBLY) 45

4.2.3 TRANSFORMATION OF DH5A ESCHERICHIA COLI AND EXTRACTION OF PLASMID DNA FROM BACTERIAL SUSPENSION CULTURE 46

4.2.4 EXTRACTION OF GENOMIC DNA FROM MAMMALIAN CELLS (CRUDE LYSATES) .. 47

4.2.5 POLYMERASE CHAIN REACTION (PCR) 47

4.2.6 QUANTITATIVE POLYMERASE CHAIN REACTION (QPCR) 48
4.3 BIOCHEMICAL / PROTEIN-BASED METHODS 48
4.3.1 PREPARATION OF WHOLE CELL LYSATES 48
4.3.2 COOMASSIE STAINING OF PROTEINS 49
4.3.3 IMMUNOBLOT 49
4.3.4 (CO-) IMMUNOPRECIPITATION OF GFP-HMG20A AND NURD COMPONENTS IN HEK239T CELLS IN THE LABORATORY OF PROF. DR. JOEL MACKAY 50
4.3.5 MICROCOCCAL NUCLEASE IMMUNOPRECIPITATION FOLLOWED BY LABEL-FREE QUANTITATIVE MASS SPECTROMETRY OR IMMUNOBLOT 52
4.4 TRANSCRIPTOMIC ANALYZES 56
4.4.1 EXTRACTION OF TOTAL RNA FROM LIVING CELLS AND DEPLETION OF GENOMIC DNA CONTAMINANTS 56
4.4.2 REVERSE TRANSCRIPTION 56
4.4.3 MRNA SEQUENCING AND ANALYSIS 56
4.4.4 TRAJECTORY OF GENE EXPRESSION OVER TIME IN CARDIOMYOCYTE DIFFERENTIATION 57
4.5 CHROMATIN ANALYSIS 58
4.5.1 CHIP OF GFP-HMG20A IN HELA CELLS 58
4.5.1.1 Sample preparation 58
4.5.1.2 Bioinformatic analysis 61
4.5.2 CLEAVAGE UNDER TARGETS AND RELEASE USING NUCLEASE (CUT&RUN) OF HMG20A IN MOUSE EMBRYONIC STEM CELLS 61
4.5.2.1 Sample preparation 61
4.5.2.2 Bioinformatic analysis 62
4.5.3 ATAC- SEQUENCING 63
4.5.3.1 Sample preparation 63
4.5.3.2 Bioinformatic analysis 63
4.6 XENOPUS LAEVIS STUDIES LED BY STEFANIE GOSSEN AT ANNETTE BORCHER'S LABORATORY AT THE UNIVERSITY OF MARBURG 64
4.6.1 WHOLE-MOUNT RNA IN SITU HYBRIDIZATION 64
4.6.2 HMG20A DEPLETION IN DEVELOPING XENOPUS LAEVIS TADPOLES 64
4.6.3 COLLAGEN II STAINING IN XENOPUS LAEVIS TADPOLES 65

5 RESULTS 66

5.1 HMG20A INTERACTS WITH PRTH, NURD AND BHC COMPLEXES 66

5.1.1 GENERATION OF HELA CELL LINES EXPRESSING GFP-PRTH 67

5.1.2 HMG20A INTERACTS WITH COMPLEXES THAT MODIFY CHROMATIN 68

5.1.2.1 HMG20A is a new NuRD-interacting protein. 72

5.2 HMG20A BINDS TO DNA 76

5.3 HMG20A BINDS TO CHROMATIN IN AN H2A.Z-DEPENDENT AND –INDEPENDENT MANNER 78

5.4 LOSS OF HMG20A HAS LITTLE EFFECT ON TRANSCRIPTIONAL REGULATION IN DIFFERENTIATED HUMAN CARCINOMA CELL LINES 86

5.5 HMG20A REGULATES THE FORMATION OF THE NEURAL CREST AND HEART 92

5.5.1 DEPLETION OF HMG20A IN MOUSE EMBRYONIC STEM CELLS AFFECTS NEURAL CREST DIFFERENTIATION IN VITRO 94

5.5.2 HMG20A IS ESSENTIAL FOR IN VITRO CARDIOMYOCYTE DIFFERENTIATION 98

5.5.2.1.1 Transcriptional repression of Hmg20a correlates with chromatin accessibility in Hmg20a DP 102

5.5.2.1.2 Loss of HMG20A alters transcriptomic programs of cardiomyocyte differentiation 104

5.5.2.1.3 Pioneer transcription factors and master regulators of differentiation are deregulated in Day2 Hmg20a DP cells 107

5.6 HMG20A LOCALIZED TO PROMOTERS AND ENHANCERS IN DAY2 MESCS REGULATING CHROMATIN ORGANIZATION AND EMBRYONIC DEVELOPMENT. 110

5.6.1 HMG20A DAMPENS CHROMATIN ACCESSIBILITY OF TRANSCRIPTIONALLY ACTIVE GENES 111

5.6.2 GENES INVOLVED IN DEVELOPMENTAL PROCESSES AND CELL MIGRATION ARE DIRECTLY REPRESSED BY HMG20A 113

5.6.3 HMG20A COLOCALIZES WITH NURD AND LSD1/BHC COMPLEX TO H2A.Z OCCUPIED PROMOTORS AND H2A.Z INDEPENDENT ENHANCERS 114

5.6.4 HMG20A REGULATES EITHER GENES INVOLVED IN CHROMATIN ORGANIZATION OR EMBRYONIC DEVELOPMENT, DEPENDING ON ITS CO-LOCALIZATION TO H2A.Z 116

6 DISCUSSION 120

6.1 HMG20A BINDS TRANSCRIPTIONALLY REPRESSIVE COMPLEXES.....120

6.1.1 HMG20A IS PART OF AN INTRICATE NETWORK OF INTERACTIONS OF COMPONENTS OF THE NURD COMPLEX. 122

6.1.2 DNA SEQUENCE-SPECIFIC BINDING OF HMG20A 123

6.2 HMG20A BINDS TO OPEN, REGULATORY REGIONS OF THE GENOME, LIMITING THEIR EXPRESSION 127

6.3 DO SMALL AMOUNTS OF HMG20A ENFORCE ITS FUNCTION?.....128

6.4 LOSS OF HMG20A SKEWS NEURAL CREST DIFFERENTIATION TOWARDS MELANOCYTES AND PERTURBS CARDIOMYOCYTE DIFFERENTIATION.....129

6.5 HMG20A AND ITS ASSOCIATED COMPLEXES REGULATE TRANSCRIPTION PROGRAMS DURING DIFFERENTIATION – A MODEL.....133

6.6 OUTLOOK.....135

7 ACKNOWLEDGEMENTS 136

8 REFERENCES.....CXXXVII

9 APPENDICES XXIV

9.1 DATA ON THE GENERATION OF HELA CELLS EXPRESSING GFP-PRTH XXIV

9.2 HMG20A REGULATES NEURAL CREST AND HEART DIFFERENTIATION IN XENOPUS LAEVIS XXV

9.3 VECTOR MAPS AND CLONING STRATEGIES XXIX

9.3.1 CLONING STRATEGIES APPLYING RESTRICTION ENZYME FOLLOWED BY DNA LIGATION..... XXIX

9.3.1.1 Generation of gRNA and Cas9 expression vectors xxxii

9.3.2 RECOMBINATION BASED CLONING STRATEGIES XXXIII

1.5 List of Abbreviations

4WJ ..... Four-way junction
5mc ..... 5'-methylcytosin
acetyl-CoA ..... Acetyl-coenzyme A
Acta2 ..... Actin Alpha 2
AID ..... Auxin-inducible-degron
ANP32E ..... Acidic Nuclear Phosphoprotein 32 Family Member E
ATAC-seq ..... Assay for Transposase-Accessible Chromatin using sequencing
BAF ..... BRG1-associated factor
BAH ..... Bromo-adjacent homology
BEND3 ..... BEN Domain Containing 3
BHC/CoREST ..... BRAF-HDAC/REST-Corepressor complex
BHLH ..... Basic-helix-loop-helix
bp ..... *Base pairs*
BPTF ..... Bromodomain PHD Finger Transcription Factor
BRD2 ..... Bromodomain-containing protein 2
BRG1 ..... Brahma-related gene-1
BSA ..... Bovine serum albumin fraction V
CBP ..... CREB-binding protein
CD ..... Chromodomain
Cdh1 ..... *Cadherin 1*
Cdh2 ..... *Cadherin 2*
Cebpa ..... *CCAAT Enhancer Binding Protein Alpha*
CERC2 ..... Cat Eye Syndrome Critical Region Protein 2
CHD ..... Chromodomain protein
CHD-M ..... Middle part of CHD4
ChIP ..... Chromatin immunoprecipitation

CM ..... Cardiomyocyte
CUT&RUN ..... Cleavage under targets and release using nuclease
DMEM ..... Dulbecco's modified Eagle's Medium
DNA-FISH ..... DNA-Fluorecence *in situ* hybridization
DNMT ..... DNA methyl transferase
EB ..... Embryoid body
EHMT2 ..... Euchromatic histone-lysine N-methyltransferase 2
EMSA ..... Electrophoretic mobility shift assays
EMT ..... Epithelial-mesenchymal transition
EZH2 ..... Enhancer Of Zeste Homolog 2
FACT ..... Facilitator of transcription
FBS ..... Fetal bovine serum
FOS/JUN ..... Jun Proto-Oncogene / Fos Proto-Oncogene
FSC ..... Forward scatter
Gata2 ..... GATA Binding Protein 2
GATAD ..... GATA Zinc Finger Domain Containing
Gcn5 ..... Histone acetyltransferase GCN5
GFP ..... Green fluorescence protein
GO ..... Gene ontology
GSE1 ..... Genetic Suppressor Element 1
GTF ..... Gene transfer format
HAT ..... Histone acetyltransferase
HCD ..... Higher-energy collisional dissociation
HDAC ..... Histone deacetylase
HMG box ..... High mobility group box
HMG20A ..... High mobility group protein 20A
HNRNPU ..... Heterogeneous nuclear ribonucleoprotein U
Hox ..... *Homeobox Protein*
HRP ..... Horse-radish-peroxidase
IB ..... Immunoblot

INO80 ..... Inositol-requiring mutant 80
ISWI ..... Imitation switch
KAT ..... Lysine acetyltransferase
KAT2B ..... Histone acetyltransferase KAT2B
KAT8 ..... Histone acetyltransferase KAT8
KDM1A ..... Lysine Demethylase 1A
KDM1B ..... Lysine demetylase 1 B
KDM5 ..... Lysine demethylase 5
KLF ..... Krueppel-Like Factor
Ku70/80 ..... Lupus Ku autoantigen protein p70/80
L3MBTL3 ..... Lethal(3)Malignant Brain Tumor-Like Protein 3
LB ..... Lysogeny Broth medium
LIF ..... Leukemia inhibitory factor
LSD1 ..... Lysine demethylase 1
M1HR ..... MTA1 specific histone deacetylation complex
MAGE ..... Melanoma Antigen
MBD ..... Methyl-binding domain protein
MeCP2 ..... Methyl-CpG-binding protein 2
Mef2c ..... Myocyte Enhancer Factor 2C
mESC ..... Mouse embryonic stem cell
MLL ..... Mixed Lineage Leukemia
MNase ..... Micrococcal-nuclease
MTA ..... Metastasis associated
NAA ..... 1-Naphthaleneacetic acid
NCC ..... Neural crest cell
ncRNA ..... Noncoding RNA
ND ..... Not defined
NDR ..... Nucleosome-depleted region
Neurod1 ..... *Neuronal Differentiation 1*
NGS ..... New generation sequencing

N

NuRD ..... Nucleosome remodeling and deacetylase complex

NURF ..... Nucleosome remodeling factor

Oct4..... Octamer-Binding Protein 4

ONECUT3/FOXE1... One Cut Domain Family Member 3/ Forkhead Box E1

P/S..... Penicillin/Streptomycin

p300..... Histone acetyltransferase p300

PARS2 ..... Prolyl-TRNA Synthetase 2

*Pax*3 ..... *Paired Box* 3

PBS ..... Phosphate buffered saline

PCA ..... Principle component analysis, Principal component analysis

PCR ..... Polymerase chain reaction

PDGFRα..... Platelet-derived Growth Factor Receptor α

PHD..... Plant homeodomain

PHF14.....PHD Finger Protein 14

PHF21A ..... PHD Finger Protein 21A

PRTH..... PHF14-RAI1-TCF20-HMG20A

PTM..... Post-translational modification

PVDF .....Polyvinylidene difluoride

PWWP .....Proline-Tryptophan-Tryptophan-Proline

PWWP2A..... PWWP domain-containing protein 2A

qPCR .....Quantitative PCR

RAI1 ..... Retinoic acid induced 1

RBBP .....Retinoblastoma-Binding Protein

RCOR1..... REST Corepressor 1

RD.....Replication-dependent

REST ..... RE1 silencing transcription factor/neural restrictive silencing factor

RNAPII ..... RNA Polymerase II

RNase ..... Ribonuclease

RPL11 ..... Ribosomal Protein L11

RT-qPCR ..... Reverse transcription-qPCR

SAM ..... S-adenosylhomocysteine

SANT Swi3-Ada2-N-Cor-TFIIIB

SDS-PAGE ..... SDS polyacrylamide gel electrophoresis

Sf2 ..... Super family 2

Six ..... *Sine Oculis Homeobox*

Slug ..... Snail Family Transcriptional Repressor 2

SP ..... Specificity Protein

SRCAP ..... Snf2-related CREB-binding protein activator protein

SSC ..... Sideward scatter

Ste11 ..... Transcription factor ste11

SUMO ..... Small ubiquitin-like modification

SUV39H1 ..... Suppressor of variegation 3-9 Homolog 1

SWI/SNF ..... Switch/sucrose-non-fermenting

SWR1 ..... SWI2/SNF2-Related 1 Chromatin Remodeling Complex

TAF1 ..... Transcription initiation factor TFIID subunit

Tbx5 ..... T-Box Transcription Factor 5

TCF20 ..... Transcription factor 20

TCF4 ..... Transcription factor 4

TEAD ..... TEF-1 and abaA domain transcription factor

TF ..... Transcription factor

TFA ..... Trifluoroacetic acid

Tgfb1i1 ..... *Transforming Growth Factor Beta Induced Factor Homeobox 1 like 1*

Tgfbr ..... *Transforming Growth Factor Beta Receptor*

TIP60 ..... Tat interactive protein 60 kDa

TIR1 ..... Transport Inhibitor Response 1

TSS ..... Transcriptional start site

Twist ..... Twist Family BHLH Transcription Factor

P

Wnt ..... Wingless-related integration site

WT ..... Wild-type

Xist ..... X-inactive specific transcript

YEATS YNK7-ENL-AF-9-TFIIF-small-subunit

ZNF512B ..... Zinc finger protein 512B

R

2 Introduction

2.1 Chromatin structure and its implications on transcriptional regulation

In eukaryotic cells, DNA is localized within compartments called nuclei, where it is condensed into a packaging form called chromatin. Chromatin is made of repeating units of nucleosomes (Figure 1A), consisting of 147 base pairs (bp) of DNA wrapped around proteins called histones, which are organized as octamers, with so-called linker DNA connecting them in between (Kornberg, 1974; Olins and Olins, 1974; Olins and Olins, 2003). Histone octamers comprise pairs of the core H2A, H2B, H3, and H4 histone proteins (Luger et al., 1997; Richmond and Davey, 2003). These core histones have flexible N- and C-terminal tails that are frequently post-translationally modified (Figure 1). The linker DNA is often bound by H1 or other DNA-binding proteins, which in turn allow chromatin to form higher-order structures. The distribution and density of histone octamers along DNA is variable and, therefore, chromatin structure is highly dynamic. Chromatin allows eukaryotes to physically regulate gene expression, mainly by masking genomic features or making them accessible to other regulatory proteins, such as transcription factors (TFs), in a controlled manner. Chromatin structure and architecture are highly dynamic and are regulated by various mechanisms (Figure 1B). These include, (1) histone post-translational modifications (PTMs) that not only specifically attract or repel certain chromatin binders, such as transcriptional regulators, but also influence the accessibility of DNA for DNA-dependent polymerases, (2) DNA methylation (in higher eukaryotes), (3) noncoding RNAs

(ncRNAs), (4) chromatin remodelers, and (5) deposition of histone variants. All of them are interconnected and reversible in principle.

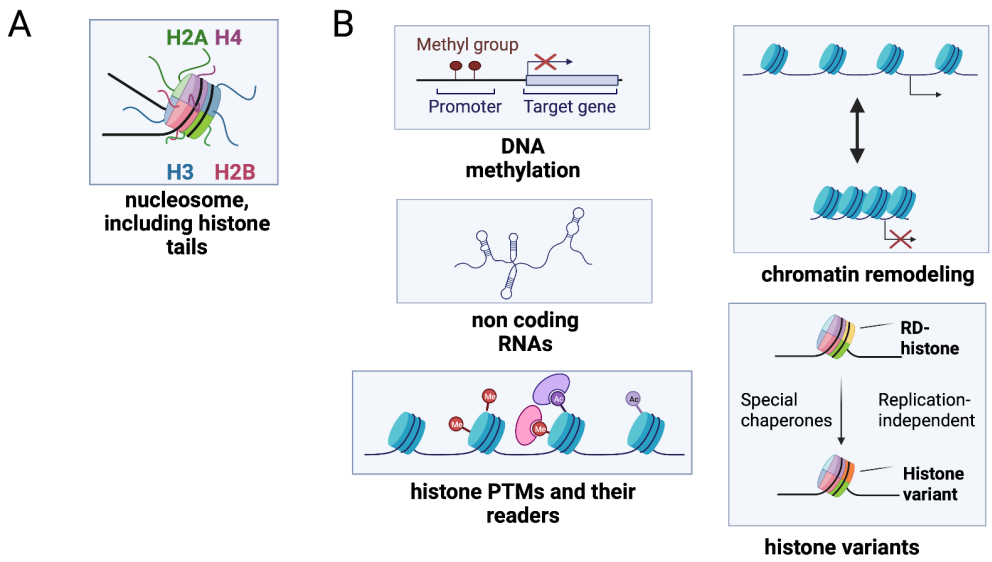


Figure 1: The eukaryotic nucleosome and general mechanisms of gene regulation
(A) Schematic depiction of a nucleosome and histone tails. (B) Schematic representation of general gene regulatory mechanisms in mammals. PTM: post-translational modification, RD: replication dependent. This figure was created using Biorender.

2.1.1 *Chromatin density is related to transcriptional activity*

Already more than a hundred years ago, it was evident that there are distinct density patterns in chromosomes, with denser parts of chromatin, coined 'heterochromatin', visually distinct from the loser, so-called 'euchromatin' regions (Flemming, 1882; Heitz, 1928). DNA-Fluorescence *in situ* hybridization (DNA-FISH) experiments starting from the 1980s finally revealed that in most cells transcriptionally active genes reside primarily in euchromatin in the interior of the nucleus, while transcriptionally repressed genes are more likely to reside in heterochromatin near the nuclear periphery during interphase of the cell cycle (Branco and Pombo, 2006; Chambeyron and Bickmore, 2004; Cremer et al., 1982; Cremer et al., 2012; Ferrai et al., 2010; Fritz et al.,

2019; Manuelidis, 1985; Nguyen et al., 2020; Payne et al., 2021; Schardin et al., 1985; Su et al., 2020; Takei et al., 2021). These categorizations are not as static as they appeared in initial experiments. The fluidity of chromatin allows certain regions, called facultative heterochromatin (Trojer and Reinberg, 2007), to switch from an euchromatic to a heterochromatic state and vice versa, depending on the needs of cells (Żylicz and Heard, 2020). However, about 45% of chromatin in humans (Consortium et al., 2001; Venter et al., 2001), 30% in *Drosophila* (Kaminker et al., 2002), and 50–80% of certain grasses (Meyers et al., 2001; Sanmiguel and Bennetzen, 1998; Vicient et al., 1999) are highly condensed and represent transcriptionally repressed constitutive heterochromatin of transposable elements (Marsano and Dimitri, 2022) e.g., centromeres (Talbert and Henikoff, 2020), telomeres and subtelomeres. These regions are, for the most part, not dynamic throughout the cell cycle or during metazoan development and remain silenced at all times.

2.1.2 *Chromatin-associated RNAs influence its organization*

As for proteins and DNA, RNA, the third macromolecular polymer in living cells, can also influence the structure and dynamics of chromatin. Famous examples of regulatory RNAs acting as scaffold to stabilize or establish certain chromatin conformations are the Heterogeneous nuclear ribonucleoprotein U (HNRNPU)-bound RNA that regulates higher-order chromatin organization (Fan et al., 2018; Zhang et al., 2019). X-inactive specific transcript (Xist), on the other hand, is of central importance in mammalian X-chromosome inactivation in female cells (Brockdorff et al., 1991; Hong et al., 2000; Nora et al., 2012; Żylicz and Heard, 2020).

2.1.3 5'methyl-cytosine represses transcription

Like histones in eukaryotes, DNA is modified in all kingdoms of life. In higher metazoans, DNA methylation is implemented mostly for gene regulatory processes (Nasrullah et al., 2022). During its development, differentiation and gametogenesis, 5'-methylcytosin (5mc) is set de novo by the DNA methyl transferases (DNMTs) DNMT3A, DNMT3B and DNMT3L (Ooi et al., 2007; Yanagisawa et al., 2002) and is inherited by DNA methyltransferase 1 (DNMT1) methylation over replication events with about 96% maintenance efficiency per replication (Bestor and Ingram, 1983; Gruenbaum et al., 1982; Laird et al., 2004). In general, 5mc is associated with gene silencing (Razin and Riggs, 1980), since it alters DNA binding affinities of transcription factors to their motifs and/or recruits repressive readers such as Methyl-CpG-binding protein 2 (MeCP2) (Meehan et al., 1989; Nan et al., 1997), Methyl-binding domain protein 1, 2 or 4 (MBD1, MBD2 or MBD4) (Hendrich and Bird, 1998; Ohki et al., 2001). Therefore, 5mc enhances chromatin binding of repressive complexes, such as NuRD, to methylated CpG islands of promoters and other *cis*-regulatory elements (Bird et al., 1985; Brackertz et al., 2006; Zhang et al., 1999), causing gene repression.

2.1.4 Histone acetylation and methylation modify transcription context-dependent

As already mentioned above, PTMs of histone proteins are a major effector of DNA-templated processes in eukaryotes. There have been numerous histone PTMs identified, such as acetylation, methylation, phosphorylation, ubiquitination, ADP-ribosylation et cetera (Ramazi et al., 2020). With respect to the relevance of this study, only histone methylation and acetylation will be discussed in more detail.

The common model of histone PTM function postulates, that patterns of PTMs 'code' (the so-called 'histone-code') for features of chromatin regions. Like code in human communication, it is (1) written, (2) read and (3) erased, in a coordinated manner (Jenuwein and Allis, 2001; Strahl and Allis, 2000).

Studies by Allfrey, Falkner and Mirsky in 1964 showed direct relationships between histone acetylation and transcription, which were later specified with acetyl-specific histone antibodies by Hebbes et al. (Allfrey et al., 1964; Hebbes et al., 1988). Histone and lysine acetyltransferases (HATs; KATs) such as Histone acetyltransferase GCN5 (Gcn5), Transcription initiation factor TFIID subunit (TAF1), Histone acetyltransferase KAT8 (KAT8), Histone acetyltransferase KAT2B (KAT2B), Histone acetyltransferase p300/CREB-binding protein (p300/CBP), transfer ('write') acetyl-groups from acetyl-coenzyme A (acetyl-CoA) (Bannister and Kouzarides, 1996; Brownell and Allis, 1995; Brownell et al., 1996; Kleff et al., 1995; Marmorstein and Zhou, 2014; Mizzen et al., 1996; Ogryzko et al., 1996) onto lysine residues of histones (Megee et al., 1990). Acetyl-groups neutralize positive charges of histone tails, leading to detachment of DNA from the histone octamer and in turn more gene activity (*cis*-effect) (Clarke et al., 1993; Grunstein, 1997; Hizume et al., 2011; O'Neill and Turner, 1995), while at the same time they can be bound ('read') by factors with conserved bromo, YNK7-ENL-AF-9-TFIIF-small-subunit (YEATS) (Dhalluin et al., 1999; Haynes et al., 1992; Li et al., 2014), or certain plant homeodomain (PHD) finger domains (Lange et al., 2008; Zeng et al., 2010), which in turn facilitate recruitment of transcriptional activators (*trans*-effect). These effects can be reversed ('erased') by histone deacetylation via histone deacetylases (HDACs) such as HDAC1 or HDAC2 (Taunton et al., 1996), or the family of

Sirtunin proteins (SIRT1-7 in mammals; Sir1 and Sir2 in yeast (Imai et al., 2000)).

SET domains (Jenuwein et al., 1998) of lysine methyl transferases (KTMs) transfer the reactive methyl group from S-adenosylhomocysteine (SAM) to certain lysine residues of histones (Aagaard et al., 1999; Liao and Seebeck, 2019; Tachibana et al., 2001; Tschiersch et al., 1994). Their impact on transcriptional output is more complicated since the position of this modification within the histone protein is of critical importance. Mono-, di- or triple-methylation of histone 3 lysine residue 4 (H3K4me1/2/3) is associated with active transcription events (Krogan et al., 2002; Santos-Rosa et al., 2002) and prevents DNA methylation in mammals (Ooi et al., 2007). These PTMs are catalyzed by, for instance, the Mixed Lineage Leukemia (MLL) protein family (Krogan et al., 2002; Milne et al., 2002). In contrast, heterochromatic methylation of lysine 9 (H3K9me) by, for example, Suppressor of variegation 3-9 Homolog 1 SUV39H1 and Euchromatic histone-lysine N-methyltransferase 2 (EHMT2), or lysine 27 (H3K27me) by Enhancer Of Zeste Homolog 2 (EZH2) (Müller et al., 2002) of the polycomb complex is connected to gene silencing (Jenuwein and Allis, 2001; Kuzmichev et al., 2002; Rea et al., 2000; Tachibana et al., 2001). Again, depending on the position of a lysine methylation within a histone, it is read and interpreted by different reader domains. Well-described methyl-lysine reader domains are chromodomains (CD) (Flanagan et al., 2005; Paro and Hogness, 1991; Pray-Grant et al., 2005; Sims et al., 2005), PHD fingers (Shi et al., 2006; Wysocka et al., 2006), and DNA/H3K36me3 binding PWWP domains (Qiu et al., 2002; Vezzoli et al., 2010). Methyl-lysine binding domains are extensively reviewed in (Musselman et al., 2014). Different lysine methylations are also removed by different demethylases, in the case of di-methylations or triple-methylations, in a stepwise manner. H3K4me3

and H3K4me2 are usually demethylated by proteins of the Lysine demethylase 5 (KDM5) protein family (Christensen et al., 2007; Sinha et al., 2010; Yamane et al., 2007), whilst H3K4me1 and H3K4me2 are demethylated by lysine demethylase 1 (LSD1; also named KDM1A (Shi et al., 2004)) of the BRAF-HDAC/ Corepressor of REST (RE1 silencing transcription factor/neural restrictive silencing factor) complex (BHC/CoREST) (Lee et al., 2005) and its homologue Lysine demethylase 1 B (KDM1B) (Ciccone et al., 2009) in the KDM1B/NPAC complex (Fang et al., 2013; Marabelli et al., 2019).

2.1.5 *Chromatin remodelers alter chromatin structure and determine DNA accessibility*

Besides the *cis*-effects of acetylation mentioned above and the intermediate impact of DNA polymerases and their auxiliary factors in transcription or replication (Leidescher et al., 2022; Wang et al., 2021a), nucleosome spacing and occupancy are determined and actively altered by nucleosome remodeling factors or complexes. Chromatin remodelers do not separate double stranded DNA, but rather 'push' and 'pull' DNA along the nucleosome (Dürr et al., 2005). They hydrolyze ATP through a conserved Super family 2 (Sf2) helicase family ATPase domain and use the energy generated to translocate the histone core down the minor groove of DNA (Côté et al., 1994).

Based on differences in structural domains, N- and C-terminally of the ATPase domain, chromatin remodelers are classified into four families (Tyagi et al., 2016): Switch/sucrose-non-fermenting (SWI/SNF), inositol-requiring mutant 80 (INO80), imitation switch (ISWI) and chromodomain (CHD). SWI/SNF (Neigeborn and Carlson, 1984; Pazin and Kadonaga, 1997; Stern et al., 1984) and INO80 remodelers (Ebbert et al., 1999; Shen et al., 2000) bind to actin and actin-related proteins via

helicase Swi3-Ada2-N-Cor-TFIIIB (SANT) domains, INO80s lack an acetyl-histone tail binding bromodomain (Awad and Hassan, 2008) for increased performance on nucleosomes containing acetylated histones (Hassan et al., 2006). On the contrary, nucleosome and DNA binding is mediated by C-terminal HAND-SANT-SLIDE domains in ISWI (Elfring et al., 1994) nucleosome remodelers (Clapier and Cairns, 2009), and by N-terminal, methyl-histone binding Chromodomains in CHD (Chromodomain-Helicase-DNA binding) remodelers (Delmas et al., 1993; Woodage et al., 1997).

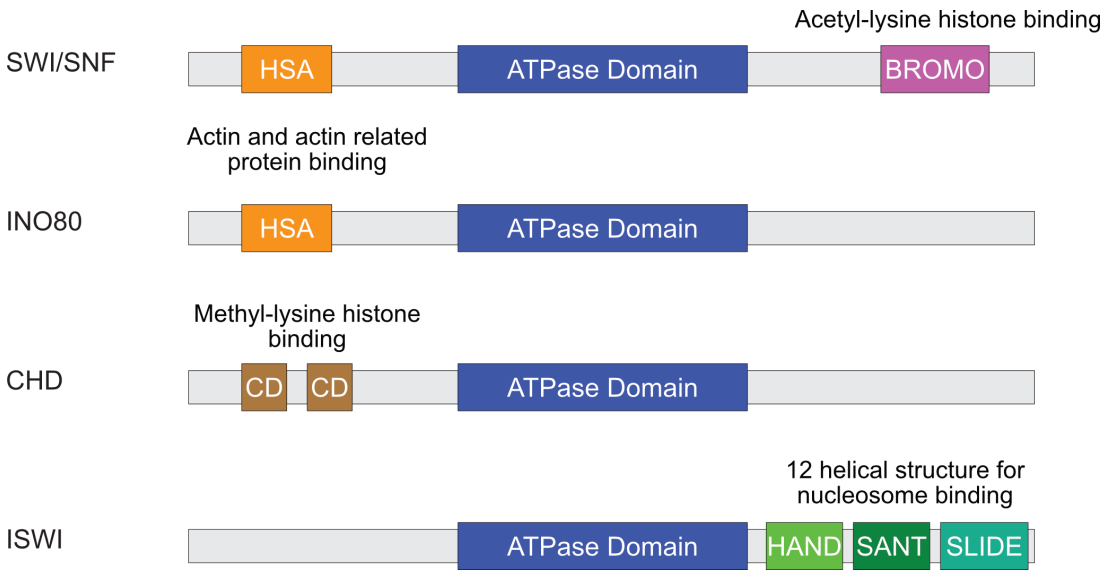


Figure 2: Schematic depiction of ATP-dependent chromatin remodeling proteins and their domains
ATP-dependent chromatin remodelers are classified by their structural differences. HSA: Helicase-SANT-associated; CD: Chromodomain. Adapted from (Tyagi et al., 2016)

SWI/SNF remodelers, namely Brahma-related gene-1 BRG1 and BRG1-associated factor (BAF) in mammals, can read histone acetylation and position nucleosomes in a coordinated manner to create or close nucleosome-depleted regions (NDRs) around transcriptional start sites (TSSs) with the help of actin and actin-like proteins (Rando et al., 2002;

Szerlong et al., 2003), modulating transcription (Krebs et al., 1999; Kwon et al., 1994; Roberts and Winston, 1997).

INO80s are, along with transcription repressors, well-established chaperones for histone variants. They are again subdivided into INO80 and SWI2/SNF2-Related 1 Chromatin Remodeling Complex (SWR1) subfamilies (Bao and Shen, 2007). In humans, conserved SWR1 complexes called Snf2-related CREB-binding protein activator protein, SRCAP, the related Tat-interactive-protein-60kDa/p400 (TIP60/p400) complex, and INO80 recognize and deposit the histone variant H2A.Z into chromatin (Kobor et al., 2004; Krogan et al., 2003; Luk et al., 2010; Mizuguchi et al., 2004), modulating its structure.

2.1.5.1 *Paralogs of Nucleosome and deacetylase complex (NuRD) members determine its function in gene regulation*

This study identified NuRD as an interactor of HMG20A. To estimate the role of HMG20A it is crucial to internalize, that NuRD is an essential, ubiquitous, and abundant transcriptional regulator with multiple subunits. Essentially, all of its components exist in paralogues. Firstly, a histone tail deacetylase comprising HDAC1/2, Metastasis associated 1/2/3 (MTA1/2/3) and Retinoblastoma-Binding Protein 4/7 (RBBP4/7). Secondly, a nucleosome remodeling subunit consisting of MBD2/3, GATA Zinc Finger Domain Containing 2A/2B (GATAD2A/B) and CHD3/4. Recent reports indicate that switching between paralogues causes different NuRD structures and allows NuRD to modulate its functional output (Reid et al., 2023).

2.1.6 *The histone variant* H2A.Z

Histone variant proteins are histones with distinct regulatory functions. Unlike their respective replication-dependent (RD) equivalents, they are deposited into chromatin through the cell cycle, usually in a non-random fashion, have a different amino acid composition, and are encoded in single-copy, or duplicated genes. Histone variants for all core histone families have been reported to exist in humans (Draizen et al., 2016; Long et al., 2019).

One of the highly studied H2A variants is H2A.Z. There are two genes coding for H2A.Z proteins H2A.Z.1 (*H2AFZ*) and H2A.Z.2.1 (*H2AFV*), called H2A.Z.2 throughout this work, with a primate specific splice variant H2A.Z.2.2 (Bönisch et al., 2012) H2A.Z is reported to be involved in most chromatin-based processes (Coon et al., 2005; Giaimo et al., 2019; Herchenröther et al., 2023; Kreienbaum et al., 2022). H2A.Z is essential for proper control of gastrulation, embryogenesis, craniofacial, and neural development (Colino-Sanguino et al., 2022; Daal and Elgin, 1992; Faast et al., 2001; Greenberg et al., 2019; Iouzalen et al., 1996), spermatogenesis (Greaves et al., 2006), memory formation (Zovkic et al., 2014), unique structure of the centromere (Greaves et al., 2007) and is overexpressed in many cancer types (Vardabasso et al., 2014). It is deposited into regulatory regions such as promoters, enhancers as well as heterochromatic domains (Bönisch et al., 2012), where it alters the nucleosome structure, provides a different post-transcriptional modification landscape, and recruits specific interacting proteins compared to RD H2A (Cole et al., 2021; Corujo and Buschbeck, 2018; Draker et al., 2012; Faast et al., 2001; Fan et al., 2004; Fujimoto et al., 2012; Giaimo et al., 2019; Greaves et al., 2007; Herchenröther et al., 2023; Hu et al., 2013; Jin et al., 2009; Kreienbaum et al., 2022; Lamaa et al., 2020;

Lewis et al., 2021; Link et al., 2018; Perell et al., 2017; Procida et al., 2021; Pünzeler et al., 2017; Ryan and Tremethick, 2018) (Figure 3, left).

The deposition and eviction of H2A.Z-H2B dimers is facilitated by several different histone chaperone complexes in a still unresolved molecular orchestration. In addition to the general histone chaperone complex Facilitator of transcription (FACT), H2A.Z is specifically incorporated by Lupus Ku autoantigen protein p70/80 (Ku70/80, also called XRCC6/ XRCC5), TIP60/p400/NuA4 and SRCAP complexes (Bönisch et al., 2012; Liang et al., 2016; Procida et al., 2021) complexes, while it is specifically evicted by Acidic Nuclear Phosphoprotein 32 Family Member E (ANP32E) or INO80 (Alatwi and Downs, 2015; Gursoy-Yuzugullu et al., 2015; Mao et al., 2014; Obri et al., 2014; Papamichos-Chronakis et al., 2011).

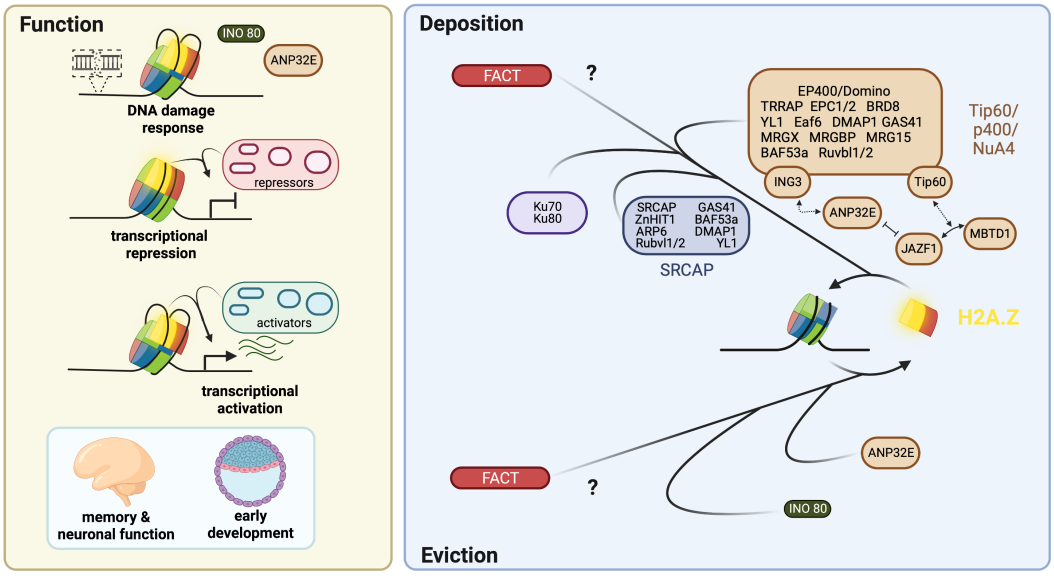


Figure 3: Schematic representation of H2A.Z's function and the factors involved in its chromatin transition.
Left: H2A.Z is evicted by INO80 and ANP32E during DNA damage response (top), it can recriuit transcriptional repressors or activators in different chromatin contexts (middle) its most studied roles in development are in early development, like gastrulation and in memory formation. Right: H2A.Z-H2B dimers are deposited by FACT, Tip60, Ku and SRCAP complexes, while their eviction is mediated by ANP32E and INO80 and FACT. Adapted from (Herchenröther et al., 2023)

The key differences between H2A.Z and RD H2A are in structure and amino acid composition of the N-terminal tail, an extension of loop1, and the C-terminal docking domain (Figure 4). When additional lysins in the N-terminus are acetylated, they are recognized by Bromodomain PHD Finger Transcription Factor BPTF (K4acK11ac) (Perell et al., 2017) and increase the nucleosome binding affinity of Bromodomain-containing protein 2 (BRD2) (Draker et al., 2012), promoting transcription. H2A.Z.1 and H2A.Z.2 both have an enlarged Loop1, compared to RD H2A, altering the H2A(.Z)-H2B interdimer interaction. A 38S/38T dimorphism between H2A.Z.1 and H2A.Z.2 makes the Loop1 of H2A.Z.2 more flexible and is responsible for the decreased stability of H2A.Z.2 containing nucleosomes *in vivo* (Horikoshi et al., 2013). Low similarity compared to RD H2A in H2A.Z's docking domain (40% sequence identity), modulates interaction with the H3-H4 dimer within the nucleosome and recognition by remodeling factors (Luger et al., 1997; Obri et al., 2014; Shukla et al., 2011), with the shortened C-terminus of H2A.Z.2 splicing variant H2A.Z.2.2, further drastically decreasing nucleosome stability (Bönisch et al., 2012).

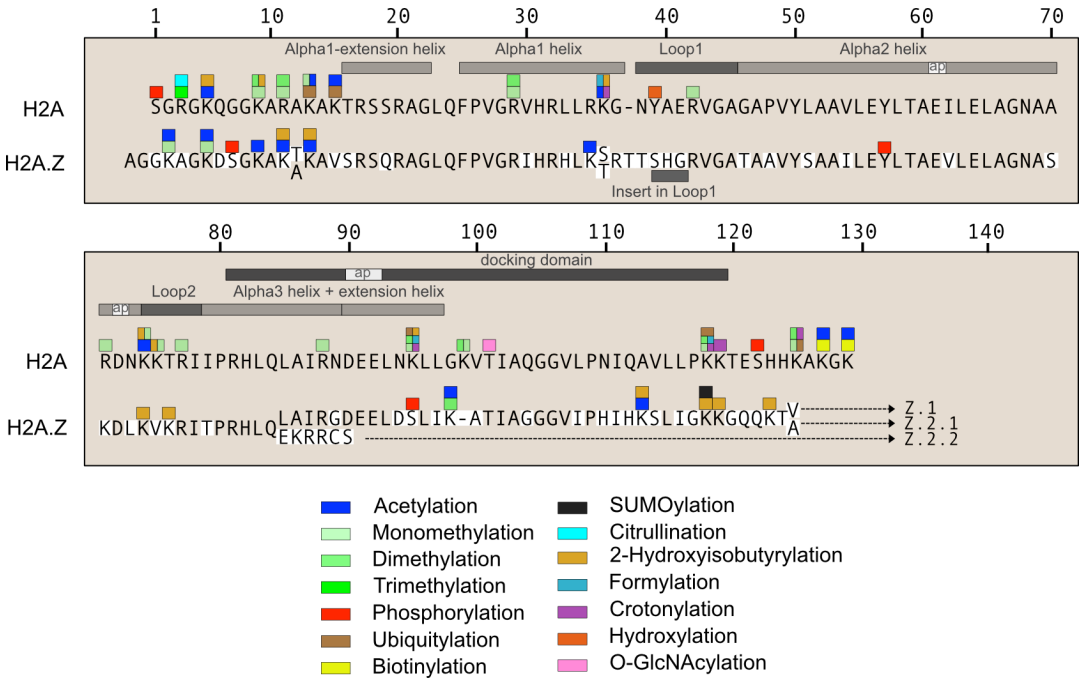


Figure 4: Alignment of amino acid sequence of H2A.Z proteins with replication-dependent H2A amino acid sequence
H2A.Z's different amino acid sequence results in differences in structural features, PTM patterns. PTMs are indicated by color. See main text for detailed description. Adapted from (Herchenröther et al., 2023)

Besides acetyl-lysine residue binding transcriptional regulators and chaperones mentioned above, multiple studies have shown that H2A.Z specifically binds to various chromatin-regulating factors and complexes (Choi et al., 2009; Draker et al., 2012; Lamaa et al., 2020; Obri et al., 2014; Pünzeler et al., 2017; Zhang et al., 2017) (Table 1). This list of H2A.Z-interacting proteins reflects its function in many different biological processes (e.g., DNA repair, splicing, or chromatin remodeling), while at the same time highlights its ambiguous role in gene regulation (reviewed in (Giaimo et al., 2019; Herchenröther et al., 2023)), as it binds to transcriptional activators, e.g., MLL, as well as repressors, e.g., HDACs. Intriguingly, Vardabasso et al.'s and Pünzeler et al.'s H2A.Z interaction study revealed that H2A.Z interacts with PHD Finger Protein 14 (PHF14), Retinoic acid induced 1 (RAI1), Transcription factor 20 (TCF20) and HMG20A, which form a complex, coined PRTH

(Eberl et al., 2013; Pünzeler et al., 2017; Vardabasso et al., 2015). Of particular interest was the association of HMG20A, as it is reported to be part of reciprocal BHC and MLL complexes, also (Ceballos-Chávez et al., 2012; Wynder et al., 2005).

Table 1: List of H2A.Z interacting proteins

Summary of interaction data from multiple studies, each protein was identified at least once

List of H2A.Z interacting proteins collected from several studies(Choi et al., 2009; Draker et al., 2012; Lamaa et al., 2020; Obri et al., 2014; Pünzeler et al., 2017; Zhang et al., 2017). ND: Not defined; NURF: Nucleosome remodeling factor; CERC2: Cat Eye Syndrome Critical Region Protein 2; BHA: bromo-adjacent homology; RNAPII: RNA Polymerase II; MAGE: Melanoma antigen. Adapted from (Giaimo et al., 2019)

Protein	Complex/Family	Function
H2A.Z		
Ep400	P400	H2A.Z deposition or ejection
TRRAP		
EPC1		
EPC2		
TIP60		
MRG15		
MRGX		
MRGBP		
YL1	P400/SRCAP	
DMAP1		
TIP49		
TIP48		
YEATS4		
BAF53		
ARP6	SRCAP	
SRCAP		
ZnHIT1		
ANP32E	P400/ANP32E	
NAP1L1	Histone-chaperone	Nucleosome assembly
NAP1L4		
SMCA5	SWI/SNF-complex	Chromatin remodeling
SMCA1	NURF and CECR2 complexes	
PHIP		Cell proliferation
SNUT2	snRNP	Splicing
KHDR1	RNA-binding	
SF3B1	Splicing factor	

Protein	Complex/Family	Function
BRD2		Gene regulation
PWWP2A	M1HR	
MTA1		
MTA2		
HDAC2		
RBBP4		
RBBP7		
MLL	H3K4me3 KMT complex	
MEN1		
HCFC2		
RBBP5		
WBP7		
WDR5		
EAF6	NuA4 complex	
ING3		
BRD8	Bromo-domain	
KDM2A	H3K36me2 demethylase	
BAHD1	BAH domain	
BCORL1		
MIER1	ELM2/SANT domains	
CDYL	Chromo-domain	
DIDO1	TF	
MYPOP		
ZFX/Y		
PHF2	H3K9me2 demethylase	
PHF6	PhD finger	
TAF7	General transcription complex	
TFII-I		
RPB1	RNAPII subunit	
TIP27/JAZF1	Zinc finger	

Protein	Complex/Family	Function
PHF14	PRTH	ND
RAI1		
HMG20A		
TCF20		
ZNF512B	Zinc finger	
MAGEA10	MAGE domain	
PHF20L1	PhD finger	
ZNF768	Zinc finger	
Myosin18A/ MYO18A	Golgi membrane trafficking	Trafficking regulation
HSP7C	Protein-chaperone	Protein folding
HSP70		
TOP2A	DNA topo-isomerase	Control of DNA topology
Actin		Cell motility
CUL4A	Ubiquitin ligase complex	Protein degradation
CUL4B		
NEDD8		
BRWD3		
TIP120		
PRKDC	Kinase	DNA repair
Pir51		
RAD23B		
XPC		
XRCC1		
XRCC5		
XRCC6		
RFS1		
MSH2	Mismatch repair complex	
MSH6		
MBTD1*	P400/MBTD1	

2.1.7 PHF14, RAI1, TCF20 & HMG20A *in gene regulation*

Wynder et al. reported that the HMG20A homolog HMG20B, a canonical member of BHC, is involved in BHC-mediated repression of *SynapsinI* in neural differentiation of murine P19 cells (Wynder et al., 2005). *SynapsinI* is a regulator of synaptic vesicles trafficking, it involved in the control of neurotransmitter release at the pre-synaptic terminal and regulates and promotes axon outgrowth and synaptogenesis (Fassio et al., 2011; Lignani et al., 2013). During neural differentiation, HMG20A levels increase, while HMG20B levels decrease (developingmouse.brain-map.org). This leads to a displacement of HMG20B from *SynapsinI*, while HMG20A now binds to it and recruits the transcriptional activator and H3 lysine 4 methyltransferase MLL1, promoting its expression and ensuring proper progression of neuronal differentiation programs (Wynder et al., 2005).

In addition to MLL1 recruitment, HMG20A promotes neural differentiation in a second way. Ceballos-Chávez and colleagues showed that HMG20A heterodimerizes with HMG20B and interferes with its small ubiquitin-like modification (SUMOylation), necessary for its regulatory activity (Ceballos-Chávez et al., 2012). At the same time, it competes with HMG20B to be part of the BHC complex, interfering with its repression of neural genes. Again, creating a positive feedback loop for neuronal differentiation programs (Ceballos-Chávez et al., 2012). The exact way how HMG20A competes with HMG20B to be a part of BHC is not yet completely resolved.

The fact that HMG20A was shown to be part of yet another chromatin complex is intriguing in several ways (Eberl et al., 2013): Mechanistically, PRTH is a putative chromatin-regulating complex that senses histone modifications, as it is repelled by H3K4me3 (Eberl et al., 2013). Therefore,

HMG20A could be involved in the detection of H3K4 methylation states (via PRTH), its establishment (via MLL) and its removal (via BHC).

The establishment and removal of H3K4me3 by MLL and LSD1 has been extensively studied in the past. But how PRTH would be able to recognize histone methylation states was not investigated until recently. Although the paralogs RAI1 and TCF20 contain several extended PHD domains, it is stated that PHF14 reads unmodified H3 tails and loses binding in case of H3 modifications (Zheng et al., 2021).

Although TCF20 is reported to be mutated in neurodevelopmental disorders (Babbs et al., 2014; Hao et al., 2022; Lévy et al., 2022; Svorenova et al., 2022; Upadia et al., 2018; Vetrini et al., 2019; Yamanaka et al., 2014; Zhou et al., 2022), there are only two studies investigating the mechanistic role of TCF20 in neural development. TCF20, along with PHF14, binds to MeCP2 and modulates MeCP2-dependent gene regulation and modifies synaptic and behavioral deficits induced by loss of MeCP2 (Zhou et al., 2022). Feng et al. reported defects in neurogenesis and behavior in TCF20 knock out mice. They claim that TCF20 is an indirect regulator of Transcription factor 4 (TCF4), an autism-related transcription factor (Feng et al., 2020; Zweier et al., 2008).

PHF14 is reported to be essential for lung development. Knocking it out in mouse models leads to death shortly after birth, because of respiratory failure, while heterozygotic knock outs are healthy and fertile, implying a dosage-dependent effect (Huang et al., 2013; Kitagawa et al., 2012). It is believed to do so by controlling mesenchymal growth by regulating the expression of Platelet-derived Growth Factor Receptor α (PDGFRα).

RAI1 mutations result in somewhat similar defects, although lung development seems not to be affected by them. They rather coincide with craniofacial malformations in Smith-Magenis syndrome (loss of function) and Potocki-Lupski syndrome (gain of function) patients. In loss of function studies in *Xenopus laevis* and mice these malformations are documented as well (Bi et al., 2005; Chong et al., 2016; Elsea and Williams, 2011; Jones et al., 2012; Joober et al., 1999; Potocki et al., 2007; Swarr et al., 2010; Tahir et al., 2014; Tunovic et al., 2014; Williams et al., 2010; Yan et al., 2007; Zwaag et al., 2009). Smith-Magenis and Potocki-Lupski syndrome affected cell types, namely NCCs, undergo EMT during the neurula stage in early development and migrate to their designated locations in the embryo to eventually differentiate to a multitude of different cell types including bones, cartilage, glia cells and melanocytes.

Like RAI1, the H2A.Z binding protein PWWP2A regulates the differentiation and migration of the neural crest. PWWP2A depletion in *Xenopus laevis* resulted in craniofacial defects very similar to those documented in loss of RAI1 (Link et al., 2018; Pünzeler et al., 2017; Tahir et al., 2014). In contrast to PRTH and HMG20A specifically, PWWP2A appears to be not directly related to H3K4 methylation, but rather to histone acetylation. By binding to MTA1, PWWP2A prohibits the assembly of the remodeling cassette of NuRD, provoking formation of an MTA1 specific histone deacetylation complex (M1HR), which acts as rheostat in enhancers of highly transcribed genes (Link et al., 2018; Low et al., 2020; Zhang et al., 2018). Intriguingly, label-free quantitation of PWWP2A's interacting proteins reveals, that PRTH is a PWWP2A interacting protein complex as well (Link et al., 2018). This brings PRTH not only into association with H2A.Z-mediated gene regulation, but vice versa, it could explain the function of H2A.Z/PWWP2A in neural crest differentiation and craniofacial development. Defective H2A.Z

deposition has been shown to be the cause of craniofacial defects in floating-harbor syndrome (Greenberg et al., 2019). This implies, that relation of H2A.Z to PWWP2A and/or PRTH could be a crucial neural crest differentiation regulator.

2.2 Aim of this study

H2A.Z-containing nucleosomes mark various regulatory regions in chromatin and are involved in numerous different DNA-based processes. Its direct interaction partner PWWP2A recruits the M1HR complex to H2A.Z occupied enhancers, where it curbs expression of highly active genes. Depletion of PWWP2A causes defects in neural crest migration that result in craniofacial malformations. HMG20A, a member of the PRTH complex, has been repeatedly reported as regulator of neuronal development. In addition to PWWP2A, its interaction partner RAI1 is involved in neural crest differentiation. Since HMG20A is an interactor of H2A.Z and PWWP2A, the aim of this study is to investigate a possible functional connection of PWWP2A and HMG20A in H2A.Z related biology.

To investigate this hypothesized connection, this study presented here applies proteome, transcriptome and (epi)genome-wide analyses in differentiated, differentiating and organismic systems.

3 Material

3.1 Laboratory Equipment and Software

Table 2: List of laboratory equipment and software

Description	Supplier
Accuri C6 Plus Flow Cytometer	Becton Dickinson BD
Axiocam 506 mono system	Carl Zeiss
Bioruptor Next Gen	Diagenode
Tissue culture hood	Thermo Fisher Scientific
Centrifuges	Beckman-Coulter Allegra X-30
	Beckman-Coulter Allegra X-30R
	Eppendorf 5424 R
	Eppendorf 5430 R
	Eppendorf 5417 R
CFX96 real-time cycler	Bio-Rad
Countess automated cell counter	Invitrogen
ECL Chemostar developer	Intas
Electrophoresis chamber (nucleic acids)	VWR Peqlab
Electrophoresis chamber	Bio-Rad
Fragment Analyzer	Agilent
Freezer (-20 °C)	Bosch
	Privileg
	Liebherr
Freezer (-80 °C)	Thermo Fisher Scientific
Fridge	Beko
	Liebherr
Gel documentation printer	Mitsubishi

GelStick Touch Documentation	Intas
Handcast gel system	Bio-Rad
H_2O purification system	Millipore
Incubator (bacteria)	Infors
Incubator (tissue culture)	Heraeus
Magnetic rack	Diagenode and GE (1.5 mL)
	Permagen (5-200 µL 8-strip)
Magnetic Stirrer	IKA
Microscopes	Carl Zeiss Axio Observer.Z1
	Carl Zeiss Telaval 31
	Leica DM IL LED
Microwave	Clatronic International
	Privileg
NanoPhotometer NP80	Implen
pH meter	Xylem Analytics
Pipette controler	Neolab
Pipette set	Labgene Scientific
Power supply unit (microscope)	Eplax
Power supply unit (nucleic acids)	Phase
Power supply unit (proteins)	Bio-Rad
QIAcube	Qiagen
Qubit 4 Fluorometer	Invitrogen
Roller mixer	LLG Labware
Rotating Wheel	Heidolph Instruments
Scale	Mettler
Semi-dry blotting system	Bio-Rad
Orbital shaker	Heidolph
Tabletop centrifuge	StarLab
Thermocycler (PCR)	Eppendorf Mastercycler
	SensoQuest
Thermomixer Comfort	Eppendorf 5436

UV-lamp (microscope)	EXFO X-cite series 120
Vacuum Pump	LLG Labware
Vortex shaker	Genie
Water bath	Köttermann
White-light plate	Kaiser slimlite plano

Description	**Version**
Affinity designer	1.10.5
BD Accuri C6 software	V1.0
Bio-Rad CFX Manager software	
ChemoStar Touch	V2.1
Readcube Papers	Cloud-based
Fragment Analyzer System	1.2.0.11
Integrative Genomics Viewer	
Intas GDS Touch 2	V1.0.1.5
Microsoft Office	2016
NCBI	web-based browser
Primer3	web-based browser
SnapGene	V5.1.5 and newer
UCSC	web-based browser
Zeiss microscope software	Zen 3.1 (blue edition)
Crispor	web-based browser

3.2 Consumables

Table 3: List of consumables

Description	Supplier
1.5 mL and 2 mL reaction tubes	Eppendorf
1.5 mL low-binding tubes	Sarstedt
15 mL and 50 mL centrifuge tubes	Greiner
15 mL conical hard plastic tubes	Sarstedt
96-well PCR plate	Sarstedt
Tissue culture plates (6-well and 24-well)	Greiner
Tissue culture plates (10 cm and 14.5 cm)	Greiner
Whatman cellulose paper	GE Healthcare Life Sciences
Cryotubes	Carl Roth
Disposable needle	B. Braun Melsungen
Disposable scalpel	B. Braun Melsungen
Disposable syringe	Henke-Sass, Wolf
Filter tips	Nerbe
Glass pipettes	HBG Henneberg-Sander
Glassware	Schott
Laboratory Bunsen burner	Campingaz
Laboratory sealing film (Parafilm)	Sigma-Aldrich
Measuring cylinder (plastic)	Brand
Microscope cover glasses (coverslips)	Paul Marienfeld
Microscope Slides	Carl Roth
Mr. Frosty-freezing container	Thermo Fisher Scientific
Nitrile gloves	StarLab
Nitrocellulose	GE Healthcare Life Sciences

Pasteur pipettes	Merck Millipore
PCR reaction tubes	Carl Roth
Pipette tips	Ratiolab
Qubit assay tubes	Invitrogen
Sealing foil	Bio-Rad
Serological pipettes	Sarstedt
Lint-free tissue	Kimberly-Clark Professional

3.3 Chemicals

Table 4: List of chemicals

Description	Supplier
4-(2-hydroxyethyl)-1-piperazineethanesulfonic	Carl Roth
Acetic acid (CH_3COOH)	Carl Roth
Acetone	Carl Roth
Agar	Carl Roth
Agarose	Carl Roth
Albumin fraction V (BSA)	Carl Roth
Ammonium persulfate (APS)	Carl Roth
Ampicillin	Carl Roth
AMPure XP beads	Beckman-
Aprotinin	AppliChem
Bromophenol blue sodium salt	Carl Roth
Calcium chloride dihydrate ($CaCl_2$ 2 × H_2O)	Carl Roth
Coomassie brilliant blue R-250	Fluka
Dimethyl sulfoxide (DMSO)	Carl Roth
DirectPCR Lysis Reagent (Cell)	Viagen
Disodium hydrogen phosphate dihydrate (Na_2HPO_4 2 × H_2O)	Carl Roth
Dithiotheitol (DTT)	Carl Roth
Dulbecco's modified Eagle medium (DMEM)	Gibco

Dynabeads (Protein G)	Invitrogen
Absolute ethanol (EtOH)	Carl Roth
Ethanol-denatured (EtOH)	Carl Roth
Ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid (EGTA)	Carl Roth
Ethylenediaminetetraacetic acid (EDTA)	Carl Roth
Fetal bovine serum (FBS)	Thermo Fisher
Fluoromount-G mounting medium	VWR
Formaldehyde (37%)	Thermo Fisher
FuGENE HD transfection reagent	Promega
GFP-Trap Dynabeads	Chromotek
Glycerol	Carl Roth
Glycine	Carl Roth
Glycogen	Thermo Fisher
Hoechst bisbenzimide H33342	Sigma-Aldrich
Hydrochloric acid (HCl)	Carl Roth
IGEPAL CA-630	Sigma-Aldrich
Immersion oil 'Immersol' 518 F	Th. Geyer
Isopropanol	Carl Roth
Kanamycin	Carl Roth
Leupeptin	AppliChem
Lithium chloride (LiCl)	Sigma-Aldrich
Magnesium chloride ($MgCl_2$)	Carl Roth
Magnesium sulfate heptahydrate ($MgSO_4$ 7 × H_2O)	Carl Roth
Methanol (MeOH)	Carl Roth
Non-fat dry milk	Carl Roth
Nonidet P-40 substitute (NP 40)	Sigma-Aldrich
Oligofectamine transfection reagent	Invitrogen
Opti-MEM	Gibco
Orange G	Merck
Paraformaldehyde (PFA)	Carl Roth
Penicillin/streptomycin	Gibco
Pepstatin	AppliChem
Phenylmethanesulfonyl fluoride (PMSF)	Carl Roth
Potassium chloride (KCl)	Carl Roth
Potassium dihydrogen phosphate (KH_2PO_4)	Carl Roth

Propidium iodide (PI)	Sigma-Aldrich
Proteinase K	Thermo Fisher
Acrylamide/Bisacrylamide mix (37.5:1 ratio)	Carl Roth
siRNAs	Dharmacon or
Sodium chloride (NaCl)	Carl Roth
Sodium deoxycholate	Sigma-Aldrich
Sodium dodecyl sulfate (SDS)	Carl Roth
Sodium hydroxide (NaOH)	Carl Roth
Sulfuric acid (H_2SO_4)	Carl Roth
Tetramethylethylenediamine (TEMED)	Carl Roth
Trident femto Western HRP substrate	GeneTex
Tris (hydrxymethyl)aminomethan (Tris)	Carl Roth
Triton X-100	Carl Roth
Tryptone/peptone	Carl Roth
Tween20	Carl Roth
Yeast Extract	Carl Roth

3.4 Commercial kits

Table 5: List of commercially available kits

Description	Supplier
HS Small Fragment Kit	Agilent
MicroPlex V2 Library Preparation Kit	Diagenode
NebNextUltra II Library Preparation Kit	New England
MinElute PCR Purification Kit	Qiagen
Plasmid Midi Kit	BLIRT
Plasmid Midi Kit	BLIRT
Qubit dsDNA HS Assay Kit	Invitrogen
RNase-Free DNase Set	Qiagen
RNeasy Mini Kit	Qiagen
Transcriptor First Strand cDNA Synthesis	Roche
ATAC-Seq Kit	Actif motif

3.5 Antibodies

The following antibodies were used in this study.

Table 6: List of antibodies, their concentration and application

Antibody	Host	Supplier	Catalog-number	Appli-cation	Dilution
α-HMG20B	Rabbit	Protein-tech	14582-1-AP	IB	1:1000
α-HMG20A	Rabbit	Protein-tech	12085-1-AP	IB	1:1000
				IF	1:200
				CUT& RUN	2 µL (0.5 µg)
α-GSE1	Rabbit	Protein-tech	24947-1-AP	IB	1:1000
α-PWWP2A	Rabbit	Novusbio	NBP2-13833	IB	1:1000
α-H3	Rabbit	abcam	ab1791	IB	1:1000
α-BRD2	Rabbit	Protein-tech	22236-1-AP	IB	1:1000
α-GFP	Mouse	Roche	11814460 001	IB	1:1000
α-GFP	Rabbit	Abcam	ab290	ChIP	1 µL (1 µg)
α-RBBP4	Rabbit	Abcam	ab488	IB	1:1000
α-MTA	Rabbit	Abcam	ab71153	IB	1:500
α-FLAG	Mouse	Sigma-Aldrich	F3165	IB	1:6000
α-FLAG (HRP-coupled)	Mouse	Sigma-Aldrich	A8592	IB	1:1000

α-HA (6E2) (HRP-coupled)	Mouse	Cell Signaling Technology	2999	IB	1:1000
α-HDAC1 (10E2) (HRP-coupled)	Mouse	Cell Signaling Tech-nology	59581	IB	1:1000
α-GFP (D5.1)	Rabbit	Cell Signaling Tech-nology	2956	IB	1:1000
α-Rabbit IgG H&L (HRP-coupled)	Goat	Abcam	ab97051	IB	1: 2,000
α-HDAC2	Mouse	Abcam	ab124974	IB	1:1000
α-IgG	Rabbit	Epicypher	13-0042	CUT& Run	0.5 µg
α-Mouse (HRP-coupled)	Goat	Thermo Fisher Scientific	31430	IB	1:20000
α-Rabbit (HRP-coupled)	Goat	Thermo Fisher Scientific	31460	IB	1:20000
α-Rabbit-Alexa 488	Goat	Thermo Fisher Scientific	A-11070	IF	1:200
α-Rabbit-Alexa 594	Goat	Thermo Fisher Scientific	A-11012	IF	1:200

Antibody	Host	Supplier	Catalog-number	Appli-cation	Dilution
α-HMG20B	Rabbit	Protein-tech	14582-1-AP	IB	1:1000
α-HMG20A	Rabbit	Protein-tech	12085-1-AP	IB	1:1000
				IB	1:200
				CUT& RUN	2 µL (0.5 µg)
α-GSE1	Rabbit	Protein-tech	24947-1-AP	IB	1:1000
α-PWWP2A	Rabbit	Norvus-bio	NBP2-13833	IB	1:1000
α-H3	Rabbit	abcam	ab1791	IB	1:1000
α-BRD2	Rabbit	Protein-tech	22236-1-AP	IB	1:1000
α-GFP	Mouse	Roche	11814460 001	IB	1:1000
α-GFP	Rabbit	abcam	ab290	ChIP	1 µL (1 µg)
α-RBBP4	Rabbit	abcam	ab488	IB	1:1000
α-MTA	Rabbit	abcam	ab71153	IB	1:500
α-FLAG	Mouse	Sigma-Aldrich	F3165	IB	1:6000
α-FLAG (HRP-coupled)	Mouse	Sigma-Aldrich	A8592	IB	1:1000

α-HA (6E2) (HRP-coupled)	Mouse	Cell Signaling Tech-nology	2999	IB	1:1000
α-HDAC1 (10E2) (HRP-coupled)	Mouse	Cell Signaling Tech-nology	59581	IB	1:1000
α-GFP (D5.1)	Rabbit	Cell Signaling Tech-nology	2956	IB	1:1000
α-Rabbit IgG H&L (HRP-coupled)	Goat	abcam	ab97051	IB	1: 2,000
α-HDAC2	Mouse	abcam	ab124974	IB	1:1000
α-IgG	Rabbit	Epicypher	13-0042	CUT& Run	0.5 µg
α-Mouse HRP	Goat	Thermo Fisher Scientific	31430	IB	1:20000
α-Rabbit HRP	Goat	Thermo Fisher Scientific	31460	IB	1:20000
α-Rabbit-Alexa 488	Goat	Thermo Fisher Scientific	A-11070	IF	1:200
α-Rabbit-Alexa 594	Goat	Thermo Fisher Scientific	A-11012	IF	1:200

3.6 DNA oligonucleotides for (quantitative) polymerase chain reaction primer and CRISPR/Cas9 guide RNA sequences

The following oligonucleotides were synthesized by IDT or Thermo Fischer Scientific.

Table 7: List of DNA oligonucleotides, their sequence, and application

Sequence (5'->3')	Name	Application
AACGTTTGAACAGAGCACAGTG	RNU5B-1 downstream F	ChIP qPCR
AAGGGTGAGAAGCAATGGGAAT	RNU5B-1 downstream R	
GGTTACCGACTCACAAGCGA	RNU5E-1 downstream F	
GAAACTGTGCCCCTCGTCA	RNU5E-1 downstream R	
TCTCCAGGTCACCTCCCG	EIF4H promoter F	
CCTACGCGGCCCATTATGT	EIF4H promoter R	
CGTTCCTTTCCACTGGTCTTTTC	ADAMTS3 genebody F	
CTCTCCTTCCTGCTGTGTGG	ADAMTS3 genebody R	
ACAGCTTTGGGTGATGCAGT	RPL11_2gb F	
TTGTTGGACCAAAACACGGC	RPL11_2gb R	
AGAACAAAAGCATGGATGACAGC	GNAI1 F	
AGCCAAGATTGTTGTGCCAACTACA	GNAI1 R	

TTTGGATCCATGGAAAACTTGATGACTAGCTCCACC	BamHInterm HMG20A F	Cloning
TTTGCGGCCGCTATACCTTTCTTTTCTTTTTGGGACGATCGAGTCTGTTCACAACTTCT	Not1HMG20A cterm R	
TTTGGATCCATGGAAAACTTGATGACTAGCTCCACC	BamHInterm HMG20A F	
TTTGCGGCCGCTATACCTTTCTTTTCTTTTTGGGATCATGAGTGGCCTGCCG	Not1HMG20A nterm R	
TTTGGATCCCGGCAGGCCACTCATGATC	BamHMG20A cterm F	
TTTGCGGCCGCTATACCTTTCTTTTCTTTTTGGGACGATCGAGTCTGTTCACAACTTCT	Not1HMG20A cterm R	
AGTGCCAAGCTTACCCTCCCCCAACCCCCAC	LHA F	*Hmg20a* DP recombination template
GCCCTTGCTCACCATCTCCCTGCAAGGAAGAG	LHA R	
TGCAGGGAGATGGTGAGCAAGGGCGAGGAG	mCherry F	
ATGCGTTTGAGTAATTTATTCATCCCACATAACTGAAATTTTATACCCT	tripple term R	
GATGAATAAATTACTCAAACGCATTTGAACACGC	RHA F	
CACACAGGAAACCTATTAGTGAATACATTTCTTCTTTGGTAAATAGCTTCTAAATATCA	RHA R	
ATTCACTAATAGGTTTCCTGTGTGAAATTGTTATCCGC	pUC18 F	
TTGGGGGAGGGTAAGCTTGGCACTGGCCG	pUC18 R	
CGAGCAAAGAGACCAGAGGTTC	mHMG20A F	

CTCGGTCTTCTGATACTGCTCC	mHMG20A R	Reverse transcription qPCR (RT-qPCR) in mESCs
GGACAAGCTGAGCAAGATTCA	Twist1 F	
CGGAGAAGGCGTAGCTGAG	Twist1 R	
TGGTCAAGAAACATTTCAACGCC	Slug F	
GGTGAGGATCTCTGGTTTTGGTA	Slug R	
CAGGTCTCCTCATGGCTTTGC	Cdh1 F	
CTTCCGAAAAGAAGGCTGTCC	Cdh1 R	
AGCGCAGTCTTACCGAAGG	Cdh2 F	
TCGCTGCTTTCATACTGAACTTT	Chd2 R	
TTT CAC CTC AGG TAA TGG GAC T	Pax3 F	
GAA CGT CCA AGG CTT ACT TTG T	Pax3 R	
GTGGTTTCCGTAGCAACTCCTAC	Mef2c F	
GGCAGTGTTGAAGCCAGACAGA	Mef2c R	
GCTTTTATCGCTGTGACTTCGTAC	Tbx5 F	
GTAACTCCAGGTCATCACTGCC	Tbx5 R	
TGCTGACAGAGGCACCACTGAA	Acta2 F	
CAGTTGTACGTCCAGAGGCATAG	Acta2 R	
GCCTCTATCACAAGATGAACGGC	Gata4 F	
TACAGGCTCACCCTCGGCATTA	Gata4 R	
CCAGTGAAGATCTTGTTGCGG	xlhmg20aRT R	RT-qPCR in *Xenopus laevis*
TGCTCCACTCACTCCCTACA	xlhmg20aRT F	
GCACGATGTGTCTTTGACATGG	Odc R	
CAGGGTGAAAGATGAGGCAAC	Odc F	
TTTGAATTCGAAAGAATCCATTTTGAAG GAAAGGCCAA	xhmg20aprobe R	RNA *in situ* probe generation
TTTCTCGAGAGAGAACCATATGAGCGA TAAAAAAAAAGTTGA	xhmg20aprobe F	
CACCGATCTCTTCCTTGCAGGGAGA	HMG20A KO guide 1 F	

AAACTCTCCCTGCAAGGAAGAGAT	HMG20A KO guide 1 R	*Hmg20a* DP guide RNA oligo-nucleotides
CACCGCAGGGGCGGCAGGGTAGAAC	HMG20A KO guide 2 F	
AAACCAGGGGCGGCAGGGTAGAAC	HMG20A KO guide 2 R	
CACCGTTTAGTGTGTTCTACGTGAC	HMG20A KO guide 3 F	
AAACTTTAGTGTGTTCTACGTGAC	HMG20A KO guide 3 R	
CACCGCAGCACCGTGGGCCTGGCAC	HMG20A KO guide 4 F	
AAACGTGCCAGGCCCACGGTGCTG	HMG20A KO guide 4 R	

3.7 General buffers

Table 8: List of general buffers

Buffer/solution	**Components**
Coomassie Staining Solution	10% Acetic acid (v/v)
	30% Methanol (v / v)
Coomassie staining solution	10% Acetic acid (v/v)
	50% Methanol (v / v)
	0.1% Coomassie brilliant blue R-250 (w/v)
Ethidium bromide	10 mg/ mL Ethidium bromide (Carl Roth)
Laemmli buffer (10×)	1.29 M Glycine
	0.25 M Tris

	1% SDS (v/v)
Laemmli sample buffer (5×)	0.5 M DTT
	250 mM Tris pH 8
	0.02% Bromophenol blue (w/v)
	30% Glycerol (v/v)
	10% SDS (v/v)
LB agar	1.5% LB agar (v/v)
LB medium	1% NaCl (w/v)
	1% Tryptone / Propitol (w / v)
	0.5% yeast extract (w/v)
Orange G loading dye buffer (6×)	0.01 M TE (pH 7.6)
	60% Glycerol (v / v)
	10% Orange G (w/v)
PBS (10×)	0.02 M KCl
	0.014 M KH_2PO_4
	0.1 M Na_2HPO_4 2 × H_2O
	1.37 M NaCl
	pH 7.4 adjusted with NaOH
Protease inhibitor mix	1: 1000 Aporotinin (1 mg/ mL)
	1: 1000 Leupeptin (1 mg/ mL)
	1:1000 Pepstatin (0.7 mg/ mL)
	1:1000 0.2 M PMSF
	1:1000 1 M DTT
Semi-dry transfer buffer (1×)	39 mM Glycine
	48 mM Tris
	20% methanol (v / v)

Separation gel (SDS-PAGE)	375 mM Tris/HCl, pH 8.8
	10% or 12% or 15% or 18% Acrylamide/Bisacrylamide mix (v/v)
	0.1% SDS (w/v)
	0.13% TEMED (v/v)
	0.13% APS (w/v)
4% stack gel (SDS-PAGE)	125 mM Tris/HCl, pH 6.8
	4% Acrylamide / Bisacrylamide mix (v/v)
	0.1% SDS (w/v)
	0.13% TEMED (v/v)
	0.13% APS (w/v)
TAE buffer (50×)	0.1 M EDTA
	2 M Tris
	pH 7.8 adjusted with acetic acid
Trypan blue solution	0.4% Trypan blue stain (Thermo Fisher Scientific)
Trypsin/EDTA	0.6 mM $CaCl_2$ 2 × H_2O
	3 mM EDTA
	2.6 mM KCl

4 Methods

4.1 Cell Propagation, Cell Transfection, and Cell Biological Analysis

4.1.1 *Culturing, Passaging, Freezing and Thawing of Cells*

4.1.1.1 *Hela cells*

Adherent Hela Kyoto cells, form here on referred to Hela, were grown in Dulbecco's modified Eagle's Medium (DMEM) supplemented with 10 % heat-inactivated fetal bovine serum (FBS) and 1 % penicillin/ streptomycin (P/S) in a humidified atmosphere at 37 °C and 5 % CO2. Cell growth medium was changed every second day. If cells reached 80-90 % confluency, they were passed in a 1:20 or 1:10 ratio into 10 cm cell culture plates (covered with 10 mL of medium). After aspiration of old medium, cells they were washed once with 10 mL of phosphate buffered saline (PBS). To dissociate cells, 2 mL of Trypsin/EDTA was placed on cells for 5 minutes at 37 °C. Complete cell detachment was ensured by gently tapping the plate. The unattached cells were resuspended in 8 mL of growth medium to stop the trypsin reaction and fully separate the cells. For cultivation, cell suspension (according to the splitting ratio) was added to a fresh cell culture plate containing the full growth medium. If cells were harvested or seeded for experiments, cell viability and cell number were determined using Countess cell counter (Invitrogen) or Fuchs-Rosenthal chambers. To store cells, trypsinized cells were pelleted by centrifugation for 5 minutes at 1,200 rpm and washed in PBS. Cells were resuspended 1 mL of freezing medium (90 % FBS + 10 % DMSO), transferred to cryotubes and stored in isopropanol-filled freezing containers at -80 °C. For long-term storage, cryotubes

were relocated and kept in liquid nitrogen tanks. Cultured cells were replaced every 2-3 months with freshly thawed cells and routinely mycoplasma contamination was PCR-tested by an internal service. To thaw cells, cryotubes were removed from liquid nitrogen containers and briefly incubated in a water bath at 37 °C. To quickly remove toxic DMSO, the thawed cells were resuspended in growth medium, spun for 5 minutes at 1,200 rpm, and the supernatant was discarded. After the cells were washed once with PBS, full growth medium was added, and cells were plated on a fresh cell culture plate.

4.1.1.2 *Mouse embryonic stem cells (mESC)*

Naive mESC growth medium was prepared in house (Table 9) and stored at -80 ° C in 50 mL aliquots. Before use 50 mL aliquots were thawed, 1 mM PD03259010 (blocking the Erk1 / 2 signaling pathway), 3 mM CHIR99021 (partially blocking glycogen synthase kinase 3 (GSK3)) inhibitors and 1000U / mL of Leukemia inhibitory factor (LIF) (Millipore) were added and mixed well. The medium was wrapped in aluminum at all times. To wash cells, wash medium was prepared in house by adding 8 mL of bovine serum albumin fraction V (BSA) (7.5%) to 500 mL of DMEM / F12 (Gibco).

Cells were cultured in galantine-covered plates (prepared with 0.1 % gelatin for at least 30-40 minutes at 37 °C and aspire supernatant) in 2 mL growth medium. The cells were split every two days at a density of 1.425×10^5 cells per 6-wellplate. Separation was carried out by adding 0.5 mL of Accutase (Millipore) incubation at room temperature for up to 6 minutes followed by a light tapping to ensure detachment. To prepare a single cell suspension, 1 mL of wash medium was added and cells were pipetted 20 times, without touching the bottom of the well. Single cell suspension was ensured by bright-field microscopy. Subsequently, cells

were transferred to 15 mL of canonical tubes prefilled with 3 mL of wash medium. Cells were pelleted by centrifugation at 300 ×g for 3.5 minutes and the supernatant was carefully aspired. The pellet was intensely but carefully resolved in 0.5–2 mL of growth medium and the cells were counted as previously described. Upon seeding, make sure to evenly distribute cells by rocking plates 3-5 times on the left and right.

The cells were frozen as described above, but with different freezing medium (see table below) and 3×10^5 per cryo-vial. To completely remove serum after thawing, cells were washed twice in wash medium.

Table 9: Composition of media for culturing mouse embryonic stem cells (naive state)

Medium	Amount	Compound
Growth medium	237.5 mL	DMEM/F12
	2.5 mL	N2 supplement
	237.5 mL	Neurobasal™ Medium
	5 mL	B-27™ Supplement
	0.5 mL	2-Mercaptoethanol 50 mM
	0.5 mL	L-Glutamine 200 mM
	6.65 mL	Sodium bicarbonate (7.5 %)
	7.5 mL	Bovine albumine fraction V 7.5 %
Wash medium	500 mL	DMEM/F12
	8 mL	Bovine albumine fraction V 7.5 %
Freeze medium	50 %	growth medium
	40 %	FBS
	10 %	DMSO (Carl Roth)

Unless labeled differently, all compounds were purchased from Thermo Fischer Scientific or its daughter companies.

4.1.1.2.1 *Differentiation protocols for mouse embryonic stem cells*

4.1.1.2.1.1 Neural crest cell differentiation and migration

On differentiation Day0, naive mouse embryonic stem cells (2i+LIF) were adapted to primed state for 2 days (Day2). Hanging drops containing 1000 cells were prepared in 25 µL of differentiation media (without LIF) for two days. Then, embryoid bodies (EBs) were pooled and cultured in suspension in differentiation media (not containing LIF) supplemented with 0.1 µM retinoic acid for three days, followed by two days without retinoic acid. On differentiation Day9, EBs containing NCCs were collected for RT-qPCR or migration analysis. For protocol scheme see Figure 25A.

The neural crest cell migration assay was applied according to a published protocol in which *Xenopus laevis* NCC explants were monitored on petri dishes (Barriga et al., 2019). Likewise, on differentiation Day9, each EB, considered as an NCC explant, was carefully transferred to galantine-coated 48-well plates and cultured for at least 24 hours. The migration was monitored by manually acquiring a microscopy picture of each attached EB. The migration ability was evaluated by cell velocity and general morphology.

4.1.1.2.1.2 Cardiomyocyte differentiation

On differentiation Day0, naive mouse embryonic stem cells were adapted to primed state in differentiation media supplemented with. 10 % FBS, 2 mM L-glutamine, 1 % nonessential amino acids, 0.1 mM β-mercaptoethanol and 1000U/ mL LIF for 2 days. Hanging drops containing 1000 cells were prepared in 25 µL of differentiation medium (without LIF) supplemented with 50 µg/mL vitamin C. On

differentiation Day6 (4 days in suspension), each droplet containing one EB was carefully transferred to a 24-well plate coated with 0.1 % gelatin. Beating cardiomyocytes were observed starting from differentiation Day7. For protocol scheme see Figure 27A.

4.1.1.3 *Sf9 cells*

Sf9 cells were cultured in Sf-900TM II SFM medium (Gibco) and maintained at 27 °C and 90 rpm. The cell freezing was carried out as previously described. The Sf9 freezing medium was prepared by adding 30% fetal bovine serum and 10 % DMSO to the Sf-900TM II SFM medium.

4.1.2 *Transfection of plasmids to ectopically express fusion proteins*

To promote the expression of a (fusion) protein of interest by eukaryotic cells, external DNA, usually plasmids extracted from bacterial cells, is brought into them by transfections. To achieve the highest efficiency, each transfection has to be adapted to the cells of interest.

4.1.2.1 *Hela cells*

To genetically modify Hela cells by transfection of plasmids, FuGENE (Promega) was applied according to the manufacturer's instructions.

4.1.2.2 *Mouse embryonic stem cells (mESC)*

MESCs were transfected with FuGENE (Promega) according to the manufacturer's instructions. One day after transfection (GFP positive) mouse embryonic stem cells were single-cell sorted in 96-well plates to obtain single-cell clones.

4.1.2.3 *Sf9 cells*

Bacmids (FastBac1 system, Thermo Fisher Scientific) were transfected into Sf9 cells to express proteins of interest. Transfection was performed applying ExpiFectamine Sf transfection reagent (Gibco) according to the manufacturer's instructions. To amplify viral production, one day after transfection, transfected cell medium was used to infect fresh cells. After 1-2 rounds of infections, the medium of the infected cells was stored as viral stocks at 4 °C.

4.1.3 *Generation of Hela cell lines that ectopically express fusion proteins*

To select cells that incorporate transfected plasmids in chromatin, appropriate cytostatics (Neomycin (G418) (60 µg/ µL), Blasticidin, Pyromycin (2 µg/ µL), Hygromycin) were added to growth medium. Cells were incubated for at least 2 weeks before determining the proper expression of the ectopically expressed fusion protein.

4.1.4 *Transfection of siRNAs to deplete specific mRNAs*

To deplete mRNAs in Hela, 2×10^5 cells were transfected with 20 pmol of ON- TARGETplus Human siRNA-SMART pool (Dharmacon) using Oligofectamine according to the manufacturer's instructions (Invitrogen). Cells were cultured for 3 days before being harvested for follow-up experiments.

4.1.5 *Endogenous protein depletion in mouse embryonic stem cells*

To generate mouse embryonic cell lines with depleted Hmg20a protein level, CRISPR / Cas9 technology was used. The scaffold-guide RNAs (sgRNAs) were designed to target the start codon (ATG) and the first

intron of the gene of interest using the online tool (http://crispor.tefor.net), synthesized by Integrated DNA Technology (IDT), and cloned into the vector pX461 (Addgene). pUC19-based donor vectors that contain a selectable marker, the mCherry or puromycin resistance gene, and the mammalian transcriptional triple terminators bGH+hGH+SV40 (synthesized by GENEWIZ) flanked by homology arms (1,000 bp), generated by PCR from genomic DNA mESC obtained using the QIAamp DNA Mini Kit (QIAGEN), were constructed by HiFi DNA Assembly (New England Biolabs) according to the manufacturer's instructions.

4.1.6 *Immunofluorescence (IF) microscopy*

To assess the location of proteins within cells Immunofluorescence microscopy was applied. For that, 2×10^5 cells were seeded in 6-well cell culture plates containing coverslips, as well as 2 mL of appropriate medium. For fixation, the cover slips were transferred to 24-well plates and treated with 500 µL of 3 % paraformaldehyde (or 1 % formaldehyde, depending on the protein of interest, tested before) in PBS for 10 minutes at room temperature after three initial PBS washes. The fixation solution was aspired and the cells were washed three times with PBS. Cells were permeabilized and blocked with 500 µL of PBS containing 0.1 % Triton X-100 for 15 minutes. For protein detection, coverslips were incubated stepwise (with three washing steps in between) with 100 µL of primary and then Alexa Fluor conjugated secondary antibody dilutions (in blocking solution) at desired concentrations (see Antibody Table on page 27) for 30 minutes at room temperature in a dark chamber. After three washes of PBS, DNA was counterstained with 200 µL of 10 µg/mL Hoechst H33342 10 g/mL (Sigma-Aldrich) in PBS for 3 minutes in a dark chamber. Finally, the coverslips were washed in 400 µL ultra-pure water and mounted on

slides with a small drop of Fluoromount-G mounting medium (VWR International). The slides were dried overnight at room temperature in a dark chamber and subjected to microscopy.

Images were acquired using an Axio Observer.Z1 inverted microscope (Carl Zeiss) equipped with the Zeiss Zen 3.1 software (blue edition) software and the Axiocam 506 mono system (Carl Zeiss). Images were taken using the EGFP ET and DAPI Ultra Bandpass filter sets (AHF Analysentechnik) and processed with software tools.

4.1.7 *Flow cytometry analysis*

Flow cytometry analysis was performed using the BD Accuri C6 Plus Flow Cytometer (Becton Dickinson BD Biosciences), together with BD reagents and the BD Accuri C6 Plus v1.0 system software. The data were received as .eps graphs by the software. Data processing was carried out within the software of the BD Accuri C6 Plus v1.0 system.

4.1.7.1 *Flow cytometry analysis to ensure purity of cells expressing GFP fusion protein*

To monitor the expression of Green fluorescence protein (GFP) or GFP fusion protein, cells were subjected to flow cytometry analysis before further analysis. Cells were harvested and 200 µL of cell suspension was saved for flow cytometry measurements. The GFP signal of 25,000 events was measured with the FL1 533/30nm optical filter (formerly 530/30nm) and the gate plotted to forward (FSC) and sideward (SSC) scatter into the viable cell population of wild-type (WT) cells. The software determines the percentage of GFP positive cells. Only if > 95% of viable cells had a higher signal in FL1 533/30nm were they used for analysis. To avoid contamination of WT cells, this procedure was

performed routinely every two weeks in case the cells were in culture for several weeks.

4.2 DNA-based methods

4.2.1 *Restriction enzyme-based cloning*

To construct plasmids, one can introduce defined ends (single stranded overhangs and blunt) to DNA fragments of interest that can be ligated. These newly designed plasmids were amplified in *Escherichia coli* strains as described below. The DNA fragments of interest were usually amplified by Polymerase chain reaction (PCR) (Q5 polymerase, New England Biolabs) using primers that harbor a specific restriction enzyme recognition/cutting site followed by three consecutive thymidines at their 3'-end. Cloning strategies of plasmids created in this work are found in Section 9.3 in the Appendices. All restriction enzymes used were produced by Thermo Fischer Scientific. If a restriction enzyme was not available from this supplier, it was acquired from New England Biolabs. Restriction digests were performed according to the manufacturer's instructions. For Primer sequence information refer to Table 7.

All HMG20A cDNA containing plasmids were generated with cDNA originating from human Hela cells.

4.2.2 *Recombination-based cloning* (DNA Assembly)

Recombination templates for the CRISPR/Cas9 manipulation of mESCs genomes were constructed by DNA HiFi Assembly (New England Biolabs) according to the manufacturer's instructions. Cloning strategies of plasmids created in this work are found in Section 9.3 in

the Appendices. For Primer and oligonucleotide sequence information refer to Table 7.

4.2.3 *Transformation of DH5α Escherichia coli and extraction of plasmid DNA from bacterial suspension culture*

To amplify plasmid DNA, stock plasmids or plasmid ligation products were heat shock transformed (90 seconds 42 °C) into rubidium chloride competent DH5α *Escherichia coli* strains. After 1 hour of incubation (37 °C, rotation) in Lysogeny Broth medium (LB), cells were pelleted, resuspended in 50 µL LB and seeded in antibiotics containing LB-Agar plates to select for successful transformation. Plates were stored up to 24 hours at 37 °C. When appropriate, the presence of the desired plasmid was tested by colony PCR. PCR was performed under standard conditions (see page 47) adding a 5 minutes 95 °C step before the initial denaturation step to lyse bacteria. The desired bacterial strains were inoculated in 5 mL of LB containing appropriate antibiotics and amplified for 16 hours at 37 °C. Plasmid extraction was performed using either QIAprep Spin Miniprep Kit (Cat. No: 27106X4) (QIAGEN) or Extractme Plasmid Mini Kit (Cat. No: EM01.1) (BLIRT) according to the manufacturer's instructions. For newly created plasmids, Sanger sequencing at Mycrosynth (Goettingen) using appropriate sequencing primers assessed the proper sequence of critical plasmid sections. When larger amounts of plasmids were desired, 100 mL of LB containing appropriate antibiotics were inoculated and plasmids were extracted using the Extractme Plasmid Midi Kit (Cat. No: EM16) (BLIRT).

4.2.4 *Extraction of genomic DNA from mammalian cells (crude lysates)*

To genotype CRISPR/Cas9 edited genomes of mouse embryonic stem cells by PCR in genomic DNA from clonal populations, a fraction of cells was lysed using DirectPCR Lysis Reagent (Cell) (Cat No.: 301-C) (Viagen) according to the manufacturer's instructions, including protein digestion with Proteinase K (powder, SERVA). For PCR, 1 µL of crude lysates was put in PCR assays.

4.2.5 *Polymerase chain reaction (PCR)*

PCR can be used to amplify specific DNA sequences either for the detection of a specific DNA sequence in the sample or to modify the ends of the desired amplicon. If sequence fidelity was of importance, the Q5 Proofreading Polymerase Kit (New England Biolabs) was used for amplification following the manufacturer's instructions. Otherwise, the Promega Gotaq Polymerase Kit was applied according to the manufacturer's instructions. Primer sequences and corresponding annealing temperatures are listed on page 31.

Table 9: Cycling program for polymerase chain reaction

Cycles	Step	Temperature	Duration
1×	Initial denaturation	95 °C	0.5 minutes
25-30×	Denaturation	95 °C	20 seconds
	Annealing of primers	see Primer list	20 seconds
	Elongation	72 °C	0.5 minutes per 1,000 bp

4.2.6 *Quantitative Polymerase Chain Reaction (qPCR)*

Quantification of original template amounts in a PCR sample can be achieved by adding the intercalating fluorescent dye SYBR green to PCR reactions. For qPCR, iTaq universal SYBR green Supermix (Biorad) was used according to the manufacturer's instructions.

Primer sequences are listed in Table 7 on page 31. The primers were designed to have annealing temperatures of 60 °C.

Table 10: Cycling program for quantitative polymerase chain reaction

Cycles	Step	Temperature	Duration	Other
1×	Pre-incubation	95 °C	5 minutes	Polymerase activation and DNA
40×	Ampli-fication	95 °C 60 °C	3 seconds 20 seconds	Denaturation and annealing/exten-sion (plate read)
1×	Melting curve	65 °C to 95 °C	12 °C/min	+1 °C increments (plate read)

4.3 Biochemical / protein-based methods

4.3.1 *Preparation of Whole Cell Lysates*

To prepare whole cell lysates, varying amounts of cells were harvested and then used to generate cell extracts by lysing cells in 100-200 µL Laemmli sample buffer (1×), followed by 30 seconds to 1 minute of mild sonication and boiling for 5 minutes at 95 °C. At this point, the lysates

were either stored at -20 °C for several months or directly subjected to SDS polyacrylamide gel electrophoresis (SDS-PAGE).

4.3.2 *Coomassie Staining of Proteins*

To visualize SDS-PAGE separated proteins, gels were subjected to Coomassie staining overnight at room temperature followed by several destaining steps the next day. Before protein staining was documented, gels were saturated in gel saver solution. Staining was documented with a white-light plate (Kaiser slimlite plano) and a Nikon D3000 DSLR. The buffer composition is listed in starting on page 34.

4.3.3 *Immunoblot*

To display and roughly estimate the amount of a specific protein in a sample, immunoblot (IB) can be applied after SDS-PAGE separation of a given cell lysate. For this, gels were saturated in semi-dry transfer buffer for at least 10 minutes to remove SDS, before subjected to semi-dry protein transfer on a nitrocellulose or polyvinylidene difluoride (PVDF) membrane (constant 200 mA for 1 hour). Subsequently, the membranes were blocked in 5-10 % milk powder PBS, supplemented with 0.1 % Tween20, to prevent specific binding of antibodies to the membrane. Primary antibody incubation occurred routinely overnight in indicated dilutions at 4 °C (rotating, 3 % milk powder in PBS, supplemented with 0.1 % Tween20). The next day, the membranes were washed at least three times for five minutes at room temperature in PBS, supplemented with 0.1 % Tween20, before secondary antibodies (coupled with horse-radish-peroxidase (HRP)) were bound to primary antibodies for one hour at room temperature (rotating, 1,5 % milk powder in PBS, supplemented with 0.1 % Tween20). The membranes

were washed at least three times for five minutes at room temperature, before the signals were visualized on a chemiluminescence Imager.

4.3.4 (Co-) *Immunoprecipitation of GFP-HMG20A and NuRD components in HEK239T cells in the laboratory of Prof. Dr. Joel Mackay*

Suspension-adapted HEK Expi293F cells (Thermo Fisher Scientific, Waltham, MA, USA) were grown to a density of 2×10^6 cells/mL in Expi293 Expression Medium (Thermo Fisher Scientific). Combinations of equimolar amounts of indicated plasmids were cotransfected into cells using linear 25-kDa polyethylenimine (PEI) (Polysciences, Warrington, PA, USA). 3.8 µg of DNA mix was first diluted in 205 µL of PBS and briefly vortexed. 7.6 µg of PEI was then added and the mixture was vortexed again, incubated for 20 minutes at room temperature, and then added to 1.9 mL of HEK cell culture. Cells were incubated for 65 hours at 37 °C with 5 % CO_2 and horizontal orbital shaking at 130 rpm. The cells were then harvested, washed twice with PBS, centrifuged (300 ×g, 5 minutes), snap frozen in liquid nitrogen and stored at -80 °C. The lysates were prepared by sonicating the thawed cell pellets in 0.5 mL of lysis buffer (50mM Tris/HCl, 150 or 500 mM NaCl, 1% Triton X-100 (v/v), 1 × cOmplete EDTA-free protease inhibitor (Roche, Basel, Switzerland), 0.2 mm DTT, pH 7.9), incubated on ice for 30 minutes to precipitate chromatin and then clarifying the lysate by centrifugation (≥ 16,000 ×g, 20 min, 4 °C). The cleared supernatant was used for GFP-nanobody pulldowns. To prepare the GFP-binding beads, streptavidin beads (Thermo Fisher Scientific, Waltham, MA, USA) were first loaded with the 6×His-SUMO-streptag-GFP nanobody protein expressed and purified from *E. coli* BL21 cells. The immobilized GFP nanobody on beads captures soluble GFP-HMG20A and other proteins that interact with it.

The cleared supernatant samples were mixed with 20 µL of streptavidin beads preloaded with 3 µg of GFP- nanobodies and incubated with rotation for 2 hours at 4 °C. After incubation, the beads were first washed with 3 times 1 mL of wash buffer and then 2 times 1 mL of wash. The bound proteins were eluted by 3 times 20 µL treatment with elution buffer.

Table 11: GFP-IP buffer composition in HEK293 cells

Buffer	**Concentration**	**Compound**
Lysis buffer	500 mM	Tris pH 7.9
	500 mM	NaCl
	1%	Triton X-100
	3 mM	ATP
	3 mM	MgCl
	1 mM	PMSV
	1×	Protease inhibitor cocktail (Roche)
	0.02 mM	DTT
Wash buffer A	50 mM	HEPES pH 7.5
	500 mM	NaCl
	0.5 %	IGEPAL CA-630
	3 mM	ATP
	3 mM	MgCl
	0.02 mM	DTT
Wash buffer B	50 mM	HEPES pH 7.5
	150 mM	NaCl
	0.5 %	IGEPAL CA-630

Elution buffer	20 mM	HEPES pH 7.5
	150 mM	NaCl
	100 mM	Biotin
	0.02 mM	DTT

4.3.5 *Micrococcal nuclease immunoprecipitation followed by label-free quantitative mass spectrometry or Immunoblot*

To evaluate the interaction of proteins with GFP-HMG20A from digested micrococcal nuclease nuclei, 2×10^7 cells expressing GFP-HMG20A and GFP were harvested. Their nuclei were isolated by incubation with 0.3 % Triton-X-100 in PBS for 10 minutes at 4 °C, followed by three washing steps. The pelted nuclei were resolved in 500 µL freshly prepared Ex100 buffer and 1.5 µL micrococcal nuclease was added and then incubated 20 minutes at 26 °C. To stop micrococcal nuclease digestion, 10 mM EGTA was added followed by careful mixing of reaction tubes and transfer to 4 °C. The insoluble cell components were pelleted by centrifugation for 10 minutes at 13,000 rpm at 4 °C and the supernatants containing mononucleosomes were transferred to fresh reaction tubes. To assess proper integrity of mononucleosomes, a 10 µL aliquot was are taken, DNA fragments were isolated using PCR purification columns (QIAGEN) and subjected to agarose gel electrophoresis.

Soluble mononucleosomes were incubated with GFP-TRAP beads (Chromotek) according to the manufacturer's instructions overnight at 4 °C rotating end over end. Instead of washing incubated beads as suggested by the manufacturer, the beads were washed twice with 1 mL of 150 mM IP wash buffer 1, followed by two washes with 1 mL of each

150 mM IP wash buffer 2. For low-throughput analysis of precipitated proteins, the remaining proteins were eluted by boiling them in 50 µL SDS-loading buffer and compared with the input material (5 or 2.5 %) by Immunoblot. For high-throughput analyzes, precipitated proteins were eluted for 30 minutes at 37 °C, shaking at 1,400 rpm in the dark in 50 µL elution buffer (eluted proteins in the supernatant were transferred to a fresh reaction tube), followed by an alkylation/elution step in 50 µL alkylation buffer for 5 minutes at 37 °C, shaking at 1,400 rpm in the dark. Both eluates were combined and the eluted proteins were further alkylated and digested by trypsin (TrypsinGold, Promega) over night at 25 °C shaking at 800 rpm in the dark. The next day trypsin digestion was stopped by adding 1% trifluoroacetic acid (TFA) (Thermo Fischer Scientific) to the assays. The peptides were stored at -20 °C until shipment to Matthias Mann Laboratory (Max Plank Institute for Biochemistry, Munich, Germany), where they were subjected to label-free quantitative mass spectrometry performed and analyzed by Alexander Reim (Max Plank Institute for Biochemistry, Munich, Germany), comparing the GFP-originating peptides with cells expressing GFP-HMG20A. Mass spectrometry experiments were performed twice (biological replicates) with three technical replicates each.

For free quantitative mass spectrometry, the peptides were analyzed by reversed phase liquid chromatography on an EASY-nLC 1000 or 1200 system (Thermo Fisher Scientific, Odense, Denmark) coupled to a Q Exactive plus or HF mass spectrometer (Thermo Fisher Scientific). HPLC columns of 50 cm length and an inner diameter of 75 µm were in-house packed with ReproSil-Pur 120 C18-AQ 1.9 µm particles (Dr. Maisch GmbH, Germany). The peptide mixtures were separated using linear gradients of 120 or 140 minutes (total run time + washout) and a two

buffer system: buffer A++ (0.1% formic acid) and buffer B++ (0.1% formic acid in 80% acetonitrile). The mass spectrometer was operated in a data-dependent top 10 or top 15 mode. The peptides were fragmented by higher-energy collisional dissociation (HCD) with a normalized collision energy of 27.

The mass spectrometry data was processed using MaxQuant software version 1.4.3.1354. Fragmentation spectra were searched against a Human Sequence Database obtained from Uniprot in May 2013 and a file containing frequently observed contaminants such as human keratins. Cysteine carbamidomethylation was set as a fixed modification; N-terminal acetylation and methionine oxidation were set as variable modifications. Trypsin was chosen as specific enzyme, with maximum missed cleavages allowed. Protein and peptide identifications were filtered at 1 % FDR. Label-free quantification was performed using the MaxLFQ algorithm integrated in MaxQuant. The match between runs option was enabled with a matching time window of 0.5 minutes and an alignment time window of 20 minimum. All other parameters were left at standard settings. MaxQuant output tables were analyzed in Perseus55 version 1.5.8.6 as follows: After deleting proteins only identified with modified peptides, hits to the reverse database, contaminants and proteins with one or less razor and unique peptides, and label-free intensities were transformed into log2. The proteins were then required to have three3 valid values in at least one triplicate, then the remaining missing values in the data matrix were imputed with values representing a normal distribution around the detection limit of the mass spectrometer. Now, a two-sample t-test was performed to identify proteins enriched in the HMG20A pull-downs compared to the input control. Only those proteins were kept for further analysis. The S0

and FDR parameters were set at 0.5 and 0.05, respectively, and the edit is 25 %.

Table 12: Buffer for GFP-IP followed by label-free quantitative mass spectrometry in micrococcal nuclease digested nuclei

Buffer	**Concentration**	**Components**
Ex100	10 mM	HEPES pH 7.5
	100 mM	NaCl
	1.5 mM	MgCl
	10 %	Glycerol
	10 mM	β-Glycerol phosphate
	1 mM	DTT (add freshly)
	2 mM	$CaCl_2$ (add freshly)
150 mM IP wash buffer 1	10 mM	Tris
	150 mM	NaCl
	0.1 %	Nonidet P40 substitute (v/v)
150 mM IP wash buffer 1	10 mM	Tris pH 7.5
	150 mM	NaCl
Elution buffer	2 M	Urea
	50 mM	Tris
	2 mM	DTT (add freshly)
	20 µg/ml	Trypsin
Alkylation buffer	2 M	Urea
	50 mM	Tris pH 7.5
	10 mM	Chloroacetamide

4.4 Transcriptomic Analyzes

4.4.1 *Extraction of Total RNA from Living Cells and Depletion of Genomic DNA Contaminants*

To assess the expression of a specific and/or all mRNAs of a given cell population, total RNA was extracted using the RNeasy Mini Kit (Cat No.: 74106) following the manufacturer's instructions (QUIAGEN) including the column DNA digestion protocol (Cat. No.: 79254). The extraction was either performed by manual pipetting or by using a QIAcube pipetting robot. Upon purification, total RNA extracts were controlled for purity by UV-vis spectroscopy and integrity by agarose gel electrophoresis.

4.4.2 *Reverse transcription*

To quantify specific mRNAs in a given cell population, extracted total RNA was reverse transcribed to cDNA applying the Transcriptor First Strand cDNA Synthesis Kit (Roche) according to the manufacturer's instructions using 2 µL of random hexamer primers.

4.4.3 *mRNA Sequencing and Analysis*

To quantify global changes in mRNA expression, approximately 1 µg of extracted total RNA was sent for the preparation of the mRNA sequencing library preparation (Poly-A enrichment) to the Marburg Genome Core Facility (Hela samples) or Novogene Co., Ltd (mESC samples). Hela samples were sequenced with 40 million single end read sequencing, while mESC samples were sequenced with 20 million paired end reads. Analysis of sequencing data was performed together with

Tobias Friedrich (Borggrefe Group, Institute for Biochemistry, Giessen, Germany):

Trimming was performed identically to the CUT & RUN data (see page 61). The alignment of the FASTQ files trimmed against the mm9 genome (or hg19 for Hela data) was carried out applying hisat2 v.2.2.1 with parameters '--min-intronlen 30 -- max-intronlen 3000'. The following analysis steps were performed within R v.4.1.258 using a modified version of the R / BioConductor package systemPipeR for various steps. Based on the BAM files and the mouse mm9 GTF (or hg19 GTF for Hela data) read counts per gene for each sample were calculated using the summarizedOverlaps function of the GenomicAlignments60 R package. The resulting read counts were normalized using DESeq2 v.1.28.161. DESeq2 was used to identify differentially expressed genes (log2FC > 2 or log2FC < -2 for mESCs and log2FC > 0.8 or log2FC < -0.8 for Hela and adjusted p-value < 0.05) for the contrasts displayed, unless otherwise indicated. Principle component analysis (PCA) was calculated using DESeq2 and plotted using ggplot2.

4.4.4 *Trajectory of gene expression over time in cardiomyocyte differentiation*

The heat map on scale z was clustered according to the Euclidian distance using the 'ward.D2' method. Line The plots for gene expression on different days were Min-Max normalized based on all expression values for each gene. Gene ontology analysis for genes from different clusters was performed using the Metascape (Zhou et al., 2019) web interface (www.metascape.org) and plotted using ggplot2.

4.5 Chromatin analysis

To assess the localization of a specific protein within chromatin and along genomic sequences, chromatin-associated proteins can be cross-linked to DNA and precipitated by specific antibodies; subsequently, DNA is purified and sequenced (ChIP seq). Alternatively, cleavage under targets and release using nuclease (CUT&RUN) can be applied. Here, proteins are targeted by specific antibodies in extracted and permeabilized nuclei, and surrounding DNA is cleaved by a Protein-A-Micrococcal-nuclease (MNase) fusion protein. Increasing the temperature will release DNA fragments that can be purified and sequenced.

4.5.1 *ChIP of GFP-HMG20A in Hela cells*

4.5.1.1 *Sample preparation*

10 µL of magnetic Dynabeads Protein G (Invitrogen) were washed one time with the Dilution buffer mix and incubated with the indicated antibodies, overnight, at 4 °C (end-over-end rotation). Upon trypsinization, 1×10^7 Hela cells were pelleted at 1,000 rpm and the pellet was resolved in 2 mL of culture medium. Cross-linking was performed by adding 55 µL of 37% formaldehyde (1% final concentration) at room temperature for 10 minutes with end-to-end rotation. The crosslink reaction was stopped by adding 125 mM glycine. Cells were washed three times with 5 mL of ice-cold PBS (5 minutes 2,000 rpm, 4 °C centrifugation). Fixed cells were resuspended in 0.2 mL of SDS-Lysis buffer supplemented with DTT (1 mM) and protease inhibitors. To shear chromatin, the lysates were transferred to 15 mL conical hard plastic tubes and subjected to Bioruptor sonication for 20 cycles high energy, 30 seconds on - 30 seconds off (Diagenode, Toyama, Japan) to generate

300 bp chromatin fragments in average size. The sheared chromatin was centrifuged for 10 min, at 18,400 ×g, at 4 °C, and the supernatant was used for immunoprecipitations (IP). For the input fraction, 100 µL (10 %) of lysates were saved.

Meanwhile, the antibody-coupled Dynabeads were magnetically separated and washed three times with a dilution buffer mix (Lysisbuffer: Dilutionbuffer, 1:9). Immunoprecipitations were carried out by resuspending beads in 900 µL Dilution buffer supplemented with DTT (1mM) and protease inhibitors and by adding 100 µL of chromatin, followed by incubation overnight at 4 °C, rotating. The next day, beads were collected and washed 1 time with low-salt buffer, 1 time with high-salt buffer, 1 time LiCl buffer and 2 times with TE buffer (p H 7.6) buffer. Subsequently, the beads and input samples were resuspended in 500 µL TE buffer. To purify DNA, 1 µL Ribonuclease A (RNase A) (10 mg /mL) was added and the samples were incubated at 37 °C for 30 min. Next, 25 µL of 10 % SDS and 2.5 µL of Proteinase K (20 mg/mL) were added followed by incubation at 37 °C for 4 h and then switched to incubation at 65 °C. DNA was purified using the DNA Min Elute Purification Kit (QIAGEN) according to the manufacturer's instructions and eluted in 20 µL MilliQ water. DNA concentrations were determined with a Qubit 4 fluorometer (Invitrogen) and the dsDNA HS Assay Kit (Invitrogen). For Illumina sequencing, libraries were generated with the MicroPlex V2 library preparation kit (Diagenode), following the manufacturer's protocol. Libraries were eluted in 20 µL 0.1× TE buffer pH 8.0 and quantified with a Fragment Analyzer (Agilent), using the HS Small Fragment Kit (Agilent). For sequencing, prepared and purified libraries were sent for sequencing to the Marburg Genome Center and sequenced with 40 million, 75 bp single-end reads. Sequencing data analysis was performed

in conjunction with Prof. Dr Marek Bartkuhn (Biomedical Informatics and Systems Medicine, Gießen, Germany).

Table 13: Composition of buffers for immunoprecipitation of chromatin

Buffer	Concentration	Compound
SDS-lysis buffer	50 mM	Tris pH 8.1
	10 mM	EDTA
	1%	Sodium dodecyl sulfate (v/v)
Dilution buffer	16.7 mM	Tris pH 8.1
	167 mM	NACl
	1.2 mM	EDTA
	1%	Triton X-100 (v/v)
Low-salt buffer	20 mM	Tris pH 8.1
	150 mM	NaCl
	2 mM	EDTA
	1%	Triton X-100 (v/v)
	0.1%	Sodium dodecyl sulfate (v/v)
High-salt buffer	20 mM	Tris pH 8.1
	500 mM	NaCl
	2 mM	EDTA
	1%	Triton X-100 (v/v)
	0.1%	Sodium dodecyl sulfate (v/v)
LiCl buffer	10 mM	Tris pH 8.1
	250 mM	Lithium chloride
	1 mM	EDTA
	1%	Nonidet P40 substitute (v/v)

0.1%	Sodium deoxycholate (w/v)

For qPCR purified input DNA was diluted 10-fold before addition to the qPCR reaction mix to ensure robust Ct values.

4.5.1.2 *Bioinformatic analysis*

Sequencing data analysis was performed in conjunction with Prof. Dr. Marek Bartkuhn (Biomedical Informatics and Systems Medicine, Gießen, Germany). Manipulation of sequencing reads was done using Rsamtools, and genomic intervals were represented as GenomicRanges objects. Analysis of association between peak intervals and known genomic annotation feature was done using the ChIPseeker package (Yu et al., 2015) with default setting using the UCSC hg19 gene definitions (BioConductor package TxDb.Hsapiens.UCSC.hg19.knownGene). As statistical tests, we performed Wilcoxon rank sum tests.

4.5.2 *Cleavage under targets and release using nuclease (CUT&RUN) of HMG20A in mouse embryonic stem cells*

4.5.2.1 *Sample preparation*

An alternative method to determine the localization of a protein of interest along a genome of a given cell population is cleavage under targets and release using nuclease (CUT&RUN). Here, a micrococcal nuclease is targeted to the protein of interest by a specific antibody in unfixed, permeabilized nuclei and cleaves the DNA up- and downstream of the protein binding sites. Fragments are released by heating the nucleus suspension to 37 °C. The DNA fragments were purified and prepared as a new generation sequencing (NGS) library. In this work

CUT&RUN was performed using the CUTANA V2.1 Kit (Epicypher), library preparation was performed by applying the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) according to the manufacturer's instructions. For sequencing, prepared and purified libraries were sent for sequencing to Novogene CO., Ltd. and sequenced with 8 million, 150 bp paired-end reads. The analysis of the sequencing data was performed in collaboration with Tobias Friedrich (Borggrefe Group, Institute for Biochemistry, Gießen, Germany).

4.5.2.2 *Bioinformatic analysis*

Paired-end raw FASTQ files were quality and adapter trimmed using trimGalore v.1.1873. The trimmed FASTQ files were aligned with the mouse mm9 reference genome (Illumina's iGenomes) using hisat2 v.2.2.174 with the parameter '--no spliced alignment' and stored as binary alignment map (BAM) files. PCR duplicate reads were removed from BAM files using Picard tools v.2.21.9 (http://picard.sourceforge.net). The resulting BAM files were used to generate individual coverage tracks (bigWig) for each sample using the deepTools bamCoverage function. MACS2 v.2.2.7.176 with wild-type or knock out IGG as input was used for the peak calling on the two wild type and two *Hmg20a* DP samples. Only peaks from wild-type samples that were not identified in one of the *Hmg20a* DP samples were used as the real HMG20A binding sites. Additionally, those sites were filtered for known mouse mm9 blacklisted regions. Based on those 2545 bona fide HMG20A sites and the individual coverage tracks for each sample, the deepTools computeMatrix and plotHeatmap commands were used to generate the binding heatmap. ChIPseeker with UCSC's mm9 gene transfer format (GTF) file was used to identify genomic characteristics that are associated with HMG20A binding sites. MEME-Suite was used for the motif discovery analysis of the HMG20A binding sites.

4.5.3 ATAC- *Sequencing*

4.5.3.1 Sample preparation

100,000 cells were harvested and the preparation of the Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) was performed applying the ATAC-Seq Kit by ACTIV MOTIF according to the manufacturer's instructions. For sequencing, prepared and purified libraries were sent for sequencing to Novogene CO., Ltd. and sequenced with 30 million, 150 bp paired end reads.

4.5.3.2 Bioinformatic analysis

Paired end raw ATAC-seq FASTQ files were trimmed, aligned and filtered for PCR duplicates identical to the CUT&RUN data. Peak calling for each BAM file was performed using MACS2 v.2.2.7.1 without input and "-g 2.8e9 -q 0.01 --nomodel" as parameters. Only peaks that were conserved in at least two out of four samples (WT_1, WT_2, PR_1, PR_2) and not overlapping with backlisted regions were counted as real signals. The number of sequencing reads at these ATAC-seq signals were calculated using the summarizedOverlaps function with the "mode="Union" parameter (GenomicAlignments package). These raw read counts were normalized and differentially accessible regions were calculated using DESeq2. These normalization factors were used to generate normalized coverage tracks (bigWigs) using deepTools bamCoverage function. Heatmaps and average plots were generated identical to CUT&RUN data. The fGSEA package (Korotkevich et al., 2021) was used to generate the "GSEA". Here the ATAC-seq signals that were associated with significant deregulated genes (mRNA-seq) using ChIPseeker (Yu et al., 2015) and used as the "pathways" and the Wald's t-

test (DESeq2) for all ATAC-seq signals were used as the “ranked gene list”.

4.6 *Xenopus laevis* studies led by Stefanie Gossen at Annette Borcher’s Laboratory at the University of Marburg

All studies involving Xenopus laevis eggs or embryos were conducted in the Laboratory of Prof. Dr. Anette Borchers at the Institute for Molecular Embryology.

4.6.1 *Whole-mount RNA in situ hybridization*

To visualize the expression of a gene of interest with the whole mount in tissues, *in situ* RNA hybridization can be performed to specifically stain a targeted mRNA as described in (Harland, 1991).

Primers to generate probes are listed on page 31. Expression patterns were characterized using albino embryos; sense controls were analyzed for all documented stages. 50 µm sections were prepared using a Leica VT1000S vibratome and mounted in Mowiol.

4.6.2 *Hmg20a depletion in developing Xenopus laevis tadpoles*

To specifically inhibit *hmg20a* mRNA translation, morpholinos control MO, 5′-CCTCTT ACCTCAGTTACAATTTATA-3′ and *hmg20a* translation *hmg20a* MO, 5′- TGCAGAGGCTG TGCTTTCCATCTAG-3′ (Gene Tools, LLC) were microinjected along with capped sense *lac*Z mRNA (tracer) into a blastomere at the two-cell stage.

4.6.3 *Collagen II staining in Xenopus laevis tadpoles*

To assess defects in cartilage formation upon treatment, whole-mount collagen II stain can be used and cartilage sizes can be compared between the manipulated and untreated sides of *Xenopus laevis* tadpoles. Collagen II staining was performed in whole mount as described in (Harland, 1991). The quantification of cartilage sizes was done using the ImageJ polygon function. The ratio between the relative surface area of the Morpholino-injected side and the control side was calculated and plotted in a box plot diagram. For phenotypical and immunofluorescence documentation, a Nikon stereo microscope (SMZ18) with a DS-Fi3 Nikon camera and the NIS-Elements imaging software was used.

5 Results

5.1 HMG20A interacts with PRTH, NuRD and BHC complexes

To investigate, if HMG20A is involved in PWWP2A/H2A.Z regulated processes, I clarified that indeed, HMG20A binds to H2A.Z containing nucleosomes, as it was previously reported (Pünzeler et al., 2017). For that, I extracted nuclei from GFP and GFP-H2A.Z.1 expressing cells (stable overexpressing cell lines), digested their chromatin with MNase to generate mononucleosomes, before I immunoprecipitated GFP or GFP-H2A.Z.1 and stained for interacting proteins with immunoblot. I included verification of the co-immunoprecipitation of the known strong H2A.Z binding proteins PWWP2A and BRD2 (Draker et al., 2012; Link et al., 2018; Pünzeler et al., 2017). To control for proper stability of H2A.Z-containing nucleosomes in lysates, co-immunoprecipitation of H3 was tested additionally. When considering the different expression levels of GFP-H2A.Z.1 and GFP-H2A, HMG20A effectively bound H2A.Z-containing nucleosomes over H2A-containing nucleosomes (Figure 5).

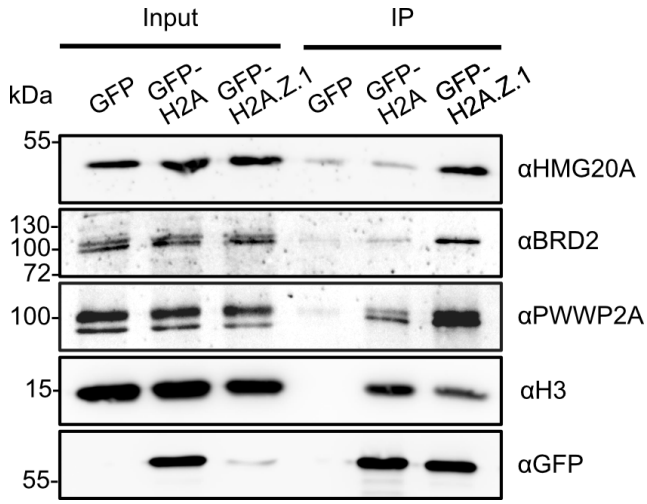


Figure 5: HMG20A preferentially binds H2A.Z-containing nucleosomes
Immunoblots of GFP, HMG20A, BRD2 and PWWP2A as well as H3 upon GFP, GFP-H2A and GFP-H2A.Z.1 mononucleosome immunoprecipitations using GFP-TRAP nanobody technology (Chromotek). Immunoblots of BRD2, PWWP2A served as positive control for immunoprecipitation of H2A.Z.1 mononucleosomes

5.1.1 *Generation of Hela cell lines expressing GFP-PRTH*

After I confirmed the interaction of HMG20A with H2A.Z-containing nucleosomes, I sought to analyze its proteome-wide interactions. To archive consistency, I applied the same cell lysis and affinity purification protocols as for the analysis of the H2A.Z and PWWP2A interactomes (Link et al., 2018; Pünzeler et al., 2017) and used the established commercially available GFP-TRAP nanobody technology (Chromotek). As for H2A.Z and PWWP2A done previously, I created Hela cell lines that stably overexpress GFP-PHF14, GFP-HMG20A, and GFP-RAI1 fusion proteins. Although I created GFP-TCF20 plasmids, I was unable to achieve successful (over-)expression of this construct. As expected, GFP-fusion proteins localized to nuclei (Figure App 1A, Appendix) and all cell clones were pure, without contaminating WT cells (Figure App 1B-D, Appendix).

5.1.2 *HMG20A interacts with complexes that modify chromatin*

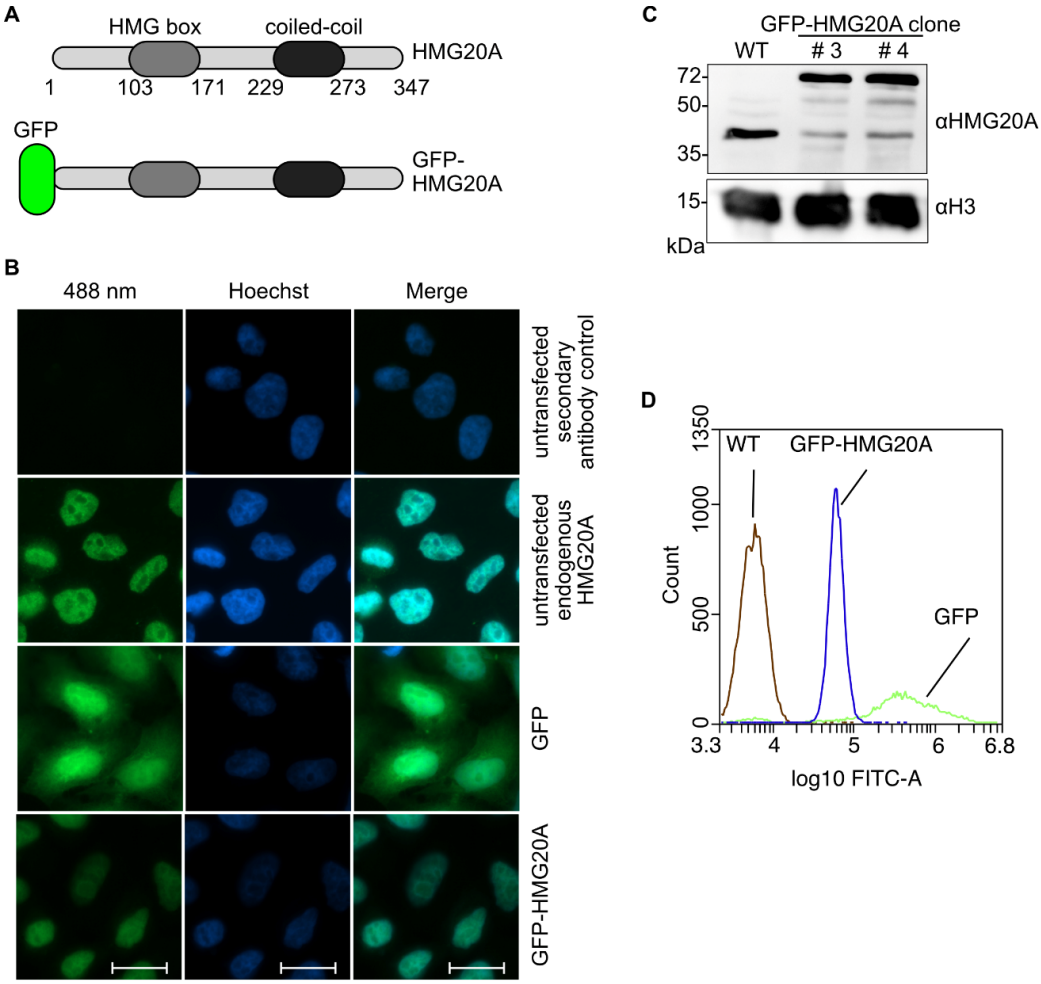


Figure 6: GFP-HMG20A localizes to nuclei with expression levels comparable to endogenous HMG20A in Hela
(A) Schematic representation of the predicted functional domains of HMG20A (top) with the added GFP tag (bottom). The numbers indicate the positions of the amino acids within the protein. (B) Fluorescent microscopy images of GFP, GFP-HMG20A and endogenous HMG20A (488 nm, green) in Hela cells, DNA was stained with Hoechst (blue). Scale bar: 20 µm. (C) Immunoblot analysis of HMG20A and GFP expression in wild-type, GFP and GFP-HMG20A expressing Hela cells. (D) Flow cytometry analysis of wild-type Hela, GFP and two individual clones of GFP-HMG20A expressing Hela cells.

HMG20A is predicted to have two functional domains, the High mobility group box (HMG box) that gives its name and a coiled-coil domain (Figure 6A, top). Ectopically expressed GFP-HMG20A (Figure 6A, bottom) was located primarily in nuclei, with patterns comparable to

endogenous HMG20A (Figure 6B), and has adequate molecular weight (Figure 6C). GFP-HMG20A expressing Hela cells remained pure, without contamination by wild-type (WT) cells or loss of expression during propagation (Figure 6D). Note that overexpression of GFP-HMG20A consistently reduced the protein level of endogenous HMG20A protein (Figure 6C), indicating a tight regulation of HMG20A protein levels in Hela cells.

After successful digestion of chromatin by micrococcal nuclease to enrich for mononucleosomes (Figure 7A) of GFP and GFP-HMG20A-expressing Hela nuclear extracts and GFP immunoprecipitation via GFP-TRAP, interacting proteins were subjected to tryptic digestion. The peptides were quantified by label free quantitative mass spectrometry in collaboration with Alexander Reim from the Matthias Mann Laboratory (Max Plank Institute for Biochemistry, Munich) (Figure 7B, C).

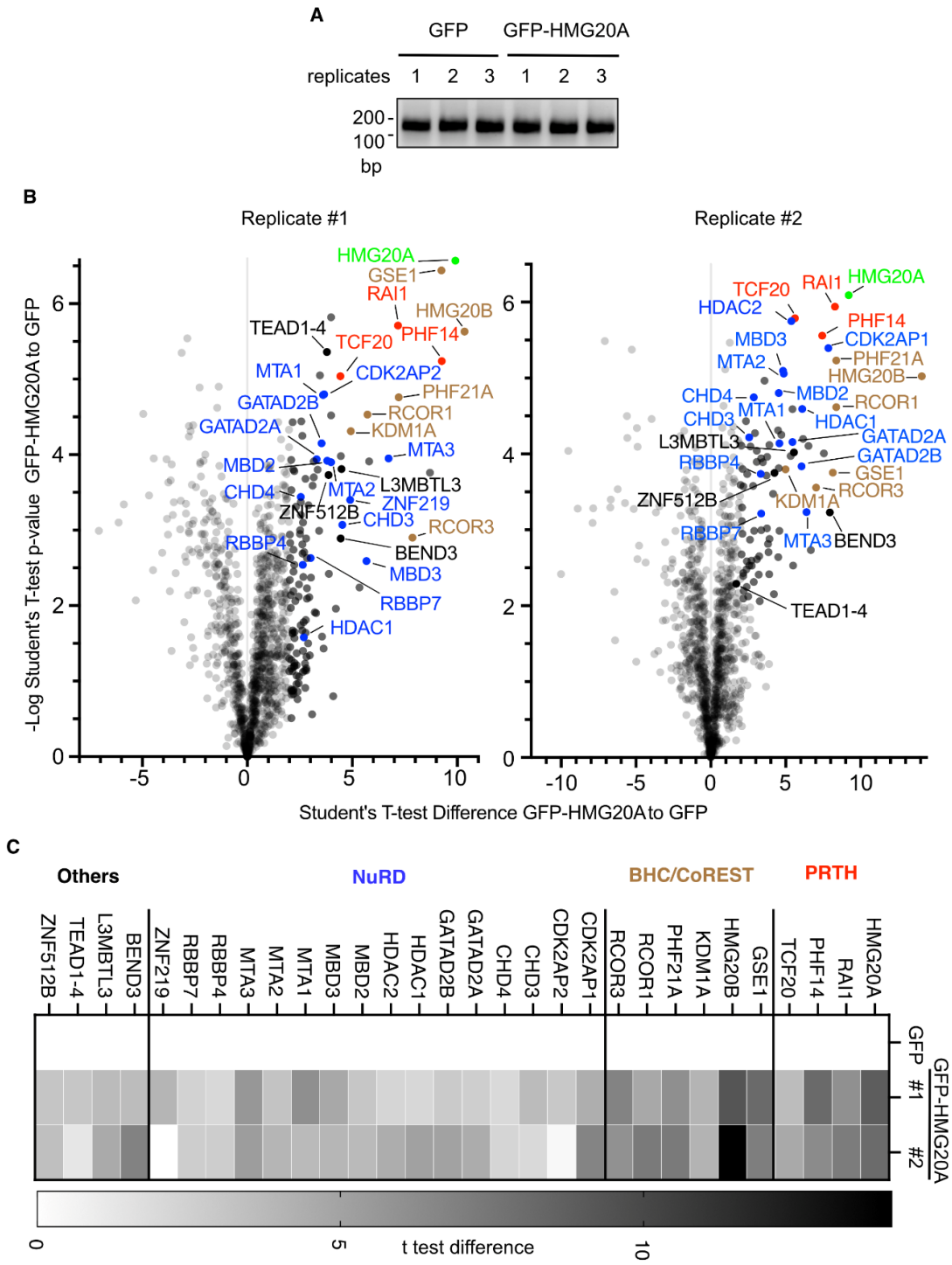


Figure 7: HMG20A interacts with protein (complexes) that modify chromatin

(A) Agarose gel of purified DNA from micrococcal nuclease-digested chromatin. (B) Volcano plot of interaction partners of GFP-HMG20A-associated mononucleosomes. Significantly enriched proteins over GFP-associated mononucleosomes are shown in the upper right. t-test differences were obtained by two-sample t-test. HMG20A is highlighted in bright green, PRTH members are highlighted in red, BHC/CoREST members are highlighted in brown, NuRD members are highlighted in blue, other proteins are highlighted in black, and background binding proteins are highlighted in gray.

(C) Heatmap of significant outliers from two independent GFP-HMG20A pulldowns analyzed by label-free quantitative mass spectrometry normalized to GFP. Scale bar: log2-fold differences of the t test.

HMG20A repeatedly showed interaction with already documented binders such as PRTH members PHF14, RAI1, TCF20 and the BHC/CoREST proteins HMG20B, Genetic Suppressor Element 1 (GSE1), PHD Finger Protein 21A (PHF21A), Lysine Demethylase 1A (KDM1A), REST Corepressor 1 and 3 (RCOR1 and RCOR3) (Eberl et al., 2013; McClellan et al., 2019; Wynder et al., 2005). In addition to those, I identified HMG20A interaction TEF-1 and abaA domain transcription factor family (TEAD1-4), BEN Domain Containing 3 (BEND3), Lethal(3)Malignant Brain Tumor-Like Protein 3 (L3MBTL3), as well as the known H2A.Z and PWWP2A interactor Zinc finger protein 512B (ZNF512B) (Link et al., 2018; Pünzeler et al., 2017), and the NuRD complex (Figure 7B and C). The interaction with ZNF512B, the BHC/CoREST member and homolog of HMG20A, HMG20B, and the NuRD components HDAC2 and MBD2 were validated in an independent immunoprecipitation experiment by Immunoblot (Figure 8).

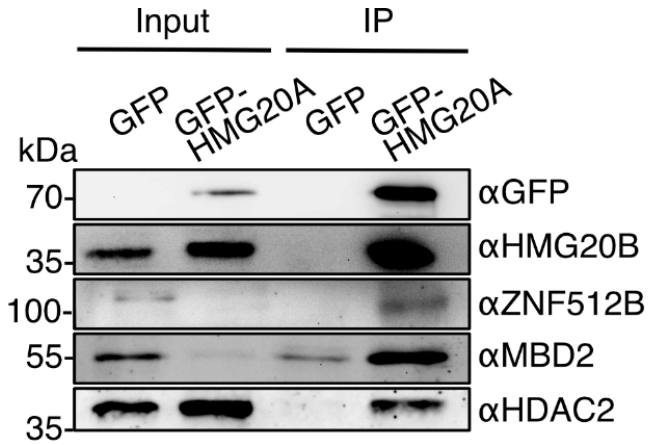


Figure 8: HMG20A binds to NuRD, PRTH, BHC, and ZNF512B
Immunoblots of GFP and GFP-HMG20A mononucleosome immunoprecipitation detecting endogenous members of the complexes BHC/CoREST (HMG20B) and NuRD (MBD2, HDAC2) complexes as well as the ZNF512B protein.

HMG20A did not interact exclusively with the M1HR-subunit of NuRD, as PWWP2A does (Link et al., 2018; Low et al., 2020; Zhang et al., 2018), but rather with both the acetyltransferase and the remodeling subunits of the complex.

5.1.2.1 *HMG20A is a new NuRD-interacting protein.*

Since HMG20A was identified as NuRD interacting protein in the context of H2A.Z and PWWP2A I sought out to analyze whether HMG20A has a similar paralogue bias. To gain first insight into which member of the NuRD complex could be bound directly by HMG20A, I collaborated with Hakimeh Moghaddas Sani in the laboratory of Prof. Dr. Joel Mackay (School of Life and Environmental Sciences, Sydney, Australia). She co-transfected GFP-HMG20A together with key components of the NuRD complex in different combinations in HEK293T cells (Figure 9)

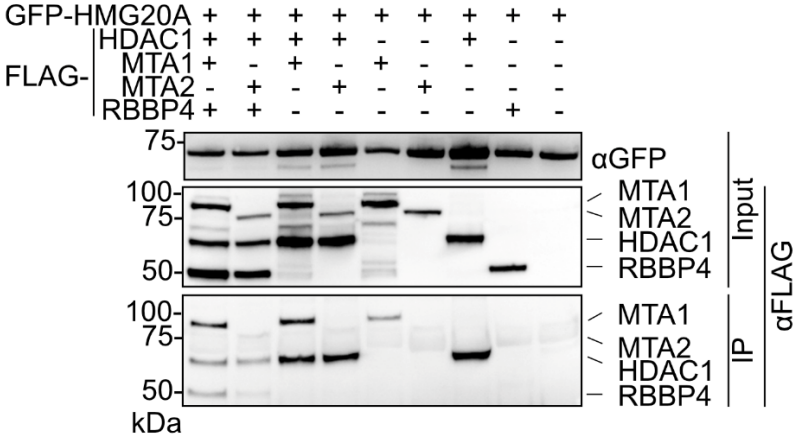


Figure 9: HMG20A preferentially binds to MTA1 and might compete with RBBP4 to bind to HDAC1.
(A) Anti-GFP immunoprecipitations of HEK293 cell extracts co-transfected with FLAG-HDAC1, -MTA1, -MTA2 and -RBBP4 and GFP-HMG20A. Proteins were detected by Immunoblotting with anti-FLAG or anti-GFP antibodies.

Indeed, similar to PWWP2A (Link et al., 2018; Low et al., 2020; Zhang et al., 2018), HMG20A preferred binding to MTA1 over MTA2 (Figure 9, lane 1 vs. 2 and 3 vs. 4). When MTA1 was replaced by MTA2, the interaction of HMG20A with RBBP4 was reduced (Figure 9, lane 1 vs. 2). Presumably, because less RBBP4 was bound to HMG20A via (endogenous) MTA1, since HMG20A did not appear to bind to RBBP4 directly (Figure 9, lane 8). The HMG20A interaction with HDAC1 was unchanged regardless of co-transfection of MTA1 or MTA2. Indicating that HMG20A additionally bound to HDAC1 in an NuRD-independent manner, possibly via BHC/CoREST members (Gocke and Yu, 2008; Song et al., 2020; Yin et al., 2014). When RBBP4 was co-transfected, the interaction of HMG20A with HDAC1 appeared to be weaker (Figure 9, lane 1 and 2 vs 3 and 4), suggesting a competition between HMG20A and RBBP4 to bind to HDAC1.

For HMG20A, two different protein-protein binding events have been reported. For interaction with the BHC/CoREST complex, the coiled-coil domain of HMG20A is needed (Rivero et al., 2015). Binding to PHF14 was reported to be mediated by the HMG box (Käsper et al., 2021). To investigate which part of the HMG20A protein is needed or the interaction with NuRD (with MTA1 as a representative), truncations of the fusions of GFP-HMG20A were generated (Figure 10A) and subjected to co-immunoprecipitation experiments in HEK293 cells showing that HMG20A bound MTA1 via its C-terminus (Figure 10B).

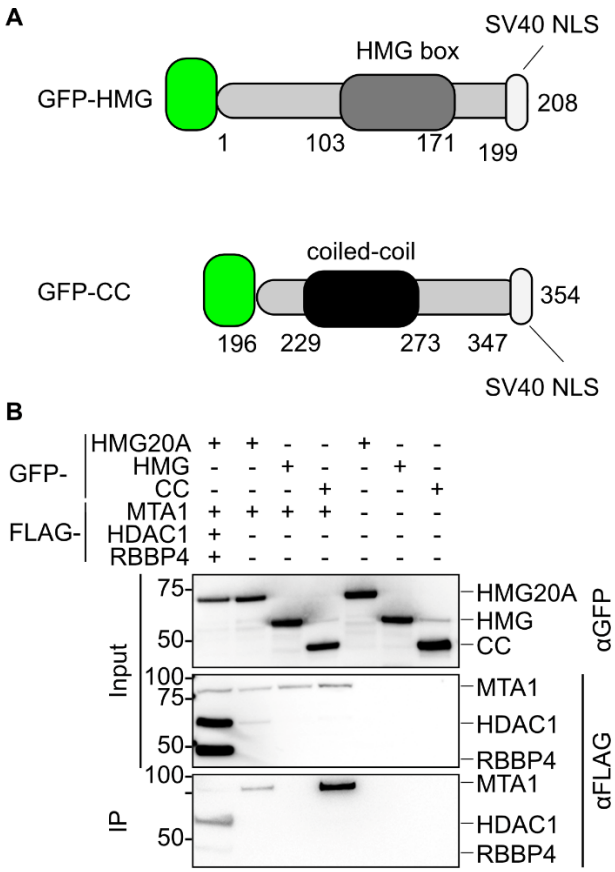


Figure 10: HMG20A binds to MTA1 through its C-terminus containing the coiled-coil domain terminus
(A) Schematics depicting HMG20A deletion constructs. (B) Anti-GFP immunoprecipitations of HEK293 cell extracts co-transfected with GFP-HMG20A or its deletions (HMG, CC) and with NuRD members (FLAG-MTA1, HDAC1-FLAG, FLAG-RBBP4). Proteins were detected by Immunoblotting with anti-FLAG or anti-GFP antibodies.

As HMG20A appeared to have a paralogue bias with respect to MTA proteins (Figure 9), we wondered if this bias also exists for components of the NuRD deacetylase cassette. To address this, we decided to repeat GFP-HMG20A co-transfections, with MBD2/3, GATA2DA, and CHD4 as well as CHD4 truncations (Figure 11). Here, HMG20A did not show any paralogue bias for MBD2 or MBD3, (Figure 11, lane 1 and lane 2). The binding of GATAD2A to HMG20A did not change irrespective of co-transfected components (Figure 11, lanes 1-3). Additionally, HMG20A did not bind to the N-terminal part of CHD4 (CHD-N), only weakly to the C-terminal part of CHD4 (CHD4-C), and most strongly to its middle part

(CHD-M), that harbors the Translocase, CDs and PHDs (Zhong et al., 2022). Taken together, these data suggest that HMG20A bound to NuRD with its C-terminus, containing its coiled-coil domain, via MTA1 and presumably to the middle region of CHD4.

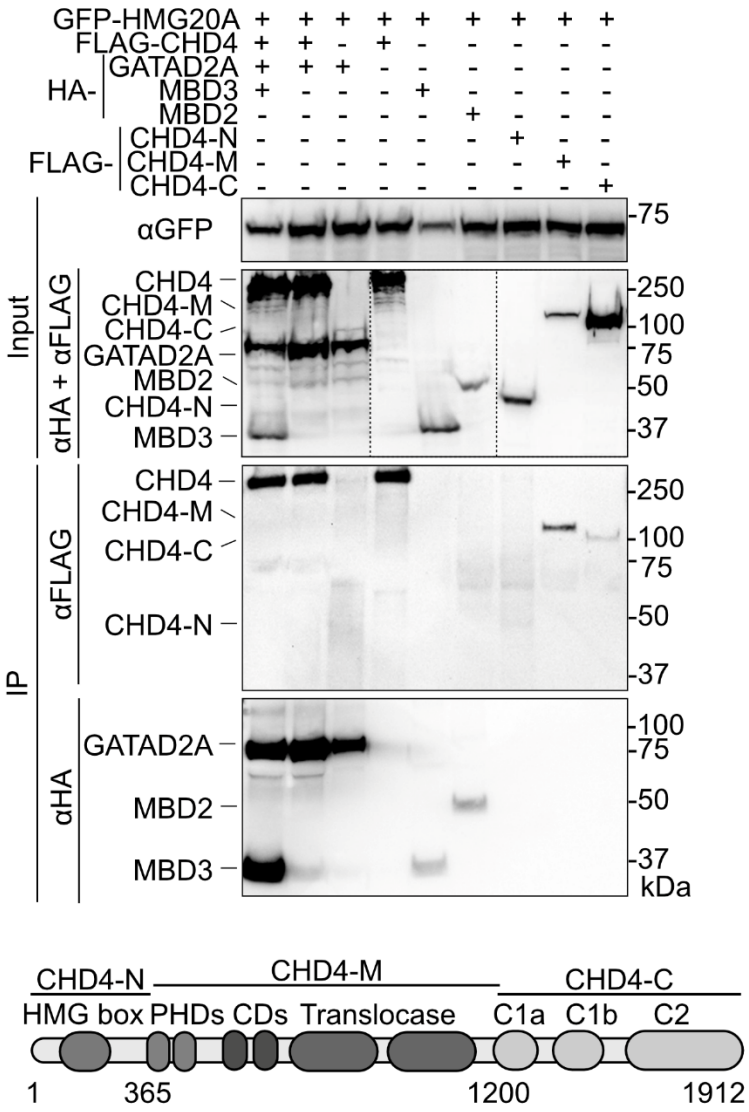


Figure 11: HMG20A as no paralogue bias for MBD2 or MBD3 and binds to the middle region of CHD4

Top: Anti-GFP immunoprecipitations of HEK293 cell extracts co-transfected with GFP-HMG20A and FLAG-CHD4 (CHD4), -CHD4-N-terminus (CHD4-N), -CHD4-middle domain (CHD4-M), -CHD4-C-terminus (CHD4-C) and HA-GATAD2A, -MBD2 and -MBD3. Proteins were detected via Immunoblot with anti-FLAG and anti-HA or anti-GFP antibodies. Bottom: schematic depiction of CHD4 deletion constructs. PHDs: Plant Homeodomain type zinc fingers, CDs: Chromodomain

5.2 HMG20A binds to DNA

Most of the binding abilities of the HMG20A protein to identified binding partners are mediated by its coiled-coil domain (Figure 7, Figure 10, (Wynder et al., 2005)). That is why I wondered if the predicted HMG box at the N-terminus of HMG20A actually binds to DNA. Together with Dr. Jörg Leers (Staff scientist at the Institute for Genetics, Justus Liebig University Giessen), I expressed FLAG-HMG20A, FLAG-HMG, FLAG-CC) fusion proteins in Sf9 cells (Figure 12A, B and C). Whole cell extracts were prepared and used to perform electrophoretic mobility shift assays (EMSA) of a PCR generated Cy5-labeled probe containing random nucleotide sequences (Figure 12C).

Indeed, HMG20A bound DNA. Binding was transmitted by its N-terminal part, which contains the predicted HMG box. With a sufficient amount of protein, the HMG20A N-terminus even bound single-stranded DNA oligomers (Figure 12D). The C-terminal part of HMG20A did not bind directly to DNA, but reduced and stabilized the interaction between DNA and proteins. Deletion of this part resulted in a stronger (see more reduction in the free probe) but less stable DNA-protein interaction during the time of migration through the gel, resulting in DNA smearing instead of a distinct shifted band, as observed in FLAG-HMG20A EMSA. The migration of Cy5-labeled DNA was not altered by extracts of the uninfected Sf9 extract, arguing that, in fact, the presence of FLAG-HMG20A constructs in the whole cell extracts of Sf9 is essential for the reduced migration of labeled DNA probes in EMSA (Figure 12D).

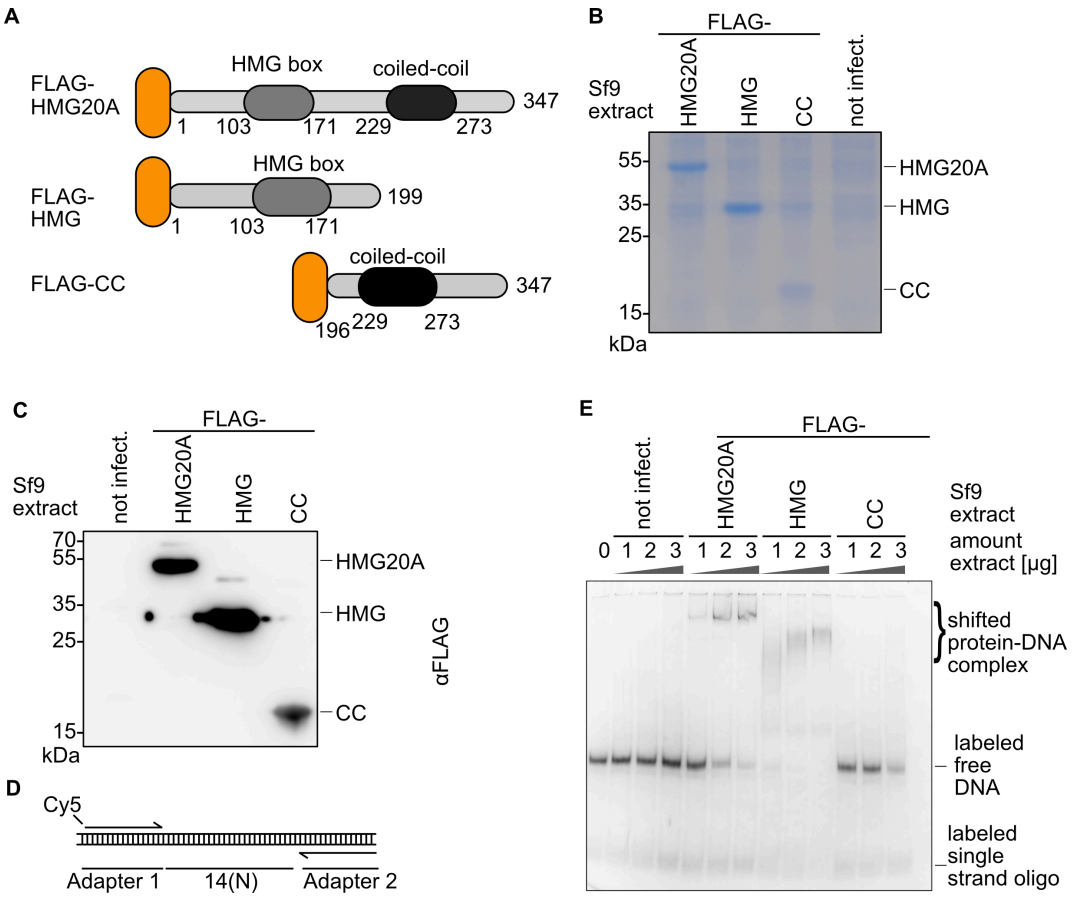


Figure 12: HMG20A binds DNA *in vitro* through its N-terminus
(A) Schematic depiction of HMG20A constructs tagged with N-terminal FLAG. The orange areas represent the FLAG tag. Light grey areas HMG boxes and black areas coiled-coil domains. The numbers indicate amino acid positions. Coomassie staining (B) or immunoblotting (C) of cell extracts from Sf9 wild-type cells (control), or Sf9 cells expressing FLAG-HMG20A, -HMG and -CC. FLAG fusion proteins were detected by anti-FLAG staining in (C). (D) Cy5-labeled random DNA sequence. The probe was generated by PCR including Cy5-labeled primer. (E) Electrophoretic mobility shift assay (EMSA) of the Cy5 labeled probe together with increasing amounts of purified recombinant FLAG-HMG20A WT and deletion proteins (see above). EMSA with whole cell extracts of uninfected Sf9 cells served as negative control.

In summary, the results indicate that HMG20A is indeed a H2A.Z nucleosome-associated protein. It associated with the entire NuRD complex through its C-terminus, which houses a predicted coiled-coil domain, while its N-terminus, which contains a predicted HMG box, bound to DNA.

5.3 HMG20A binds to chromatin in an H2A.Z-dependent and –independent manner

Having revealed the interaction of HMG20A with chromatin-modifying complexes (Figure 7) and its DNA binding capabilities (Figure 12), I sought to map its genomic location by chromatin immunoprecipitation of GFP-HMG20A with anti-GFP antibody followed by next generation sequencing in Hela cells (ChIP-seq). Before ChIP samples were used to generate sequencing libraries, I performed qPCR with them, to quantify enrichment of HMG20A at likely binding sites. To control for technical errors, I performed anti-GFP ChIP in Hela cells expressing GFP-H2A.Z, using the same GFP antibody. To ensure the feasibility of the method, performed ChIP-qPCR of the well-established H2A.Z site downstream of the Prolyl-TRNA Synthetase 2 (PARS2) gene TSS (+1 nucleosome) (Link et al., 2018) (Figure 13A). I found that shearing of crosslinked chromatin and binding of the GFP antibody were successful (Figure 13B and C). As expected, both genomic sites generated strong signals for H3K4me, and GFP-H2A.Z signal was roughly 10 times smaller for the H2A.Z-weak site in the gene body of Ribosomal Protein L11 (RPL11), which acted as negative control for unspecific GFP antibody binding, than for PARS2. Although I was unable to detect enrichment of GFP-HMG20A at PARS2 over RPL11 (Figure 13C), I decided to sequence GFP-HMG20A precipitated DNA as all controls produced sufficient results.

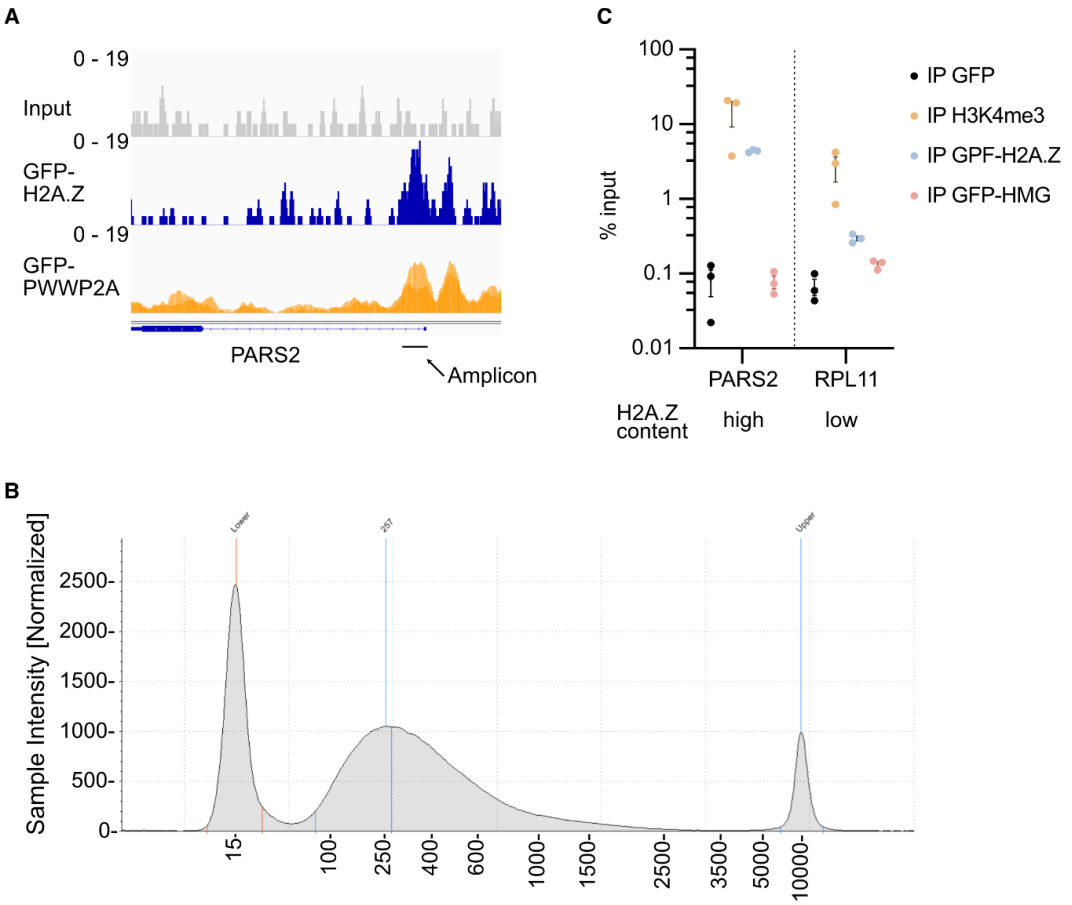


Figure 13: The GFP antibody successfully precipitates H2A.Z-bound DNA. (A) Genome browser snapshot of the H2A.Z and PWWP2A bound region of the PARS2 promotor. (B) Representative DNA track of the tape station after crosslinking and shearing of chromatin derived from GFP-HMG20A expressing Hela cells. (C) Quantitative PCR of immunoprecipitated DNA by indicated antibodies and constructs. Error bars represent SEM of three technical replicates. PARS2 is a known H2A.Z occupied site, RPL11 is a known region without/reduced presence of H2A.Z nucleosomes (Link et al., 2018). Precipitation with H3K4me3 antibody served as positive control for the applied ChIP protocol.

All the following analyses on the genomic localization of GFP-HMG20A identified by ChIP-seq were performed in conjunction with Prof. Dr. Marek Bartkuhn (Biomedical Informatics and Systems Medicine, Giessen, Germany) and Tobias Friedrich (group of Prof. Dr. Tilman Borggrefe, Institute for Biochemistry, Giessen, Germany). We analyzed, if there is an overlap of H2A.Z, PWWP2A and HMG20A in their localization on chromatin. For that, we combined ChIP-seq data from Hela cells stably overexpressing GFP-HMG20A (see above) and from

ChIP-seq data from Hela cells stably over expressing GFP-H2A.Z and GFP-PWWP2A (Link et al., 2018; Pünzeler et al., 2017). To estimate, if a given ChIP-signal lies within a promoter or enhancer region, and if said region can be considered to be active, we overlayed them with H3K4me3, H3K4me1 and H3K27ac Chip-seq data from Hela cells provided by the ENCODE consortia. For a representative snapshot of the genome browser, see Figure 14.

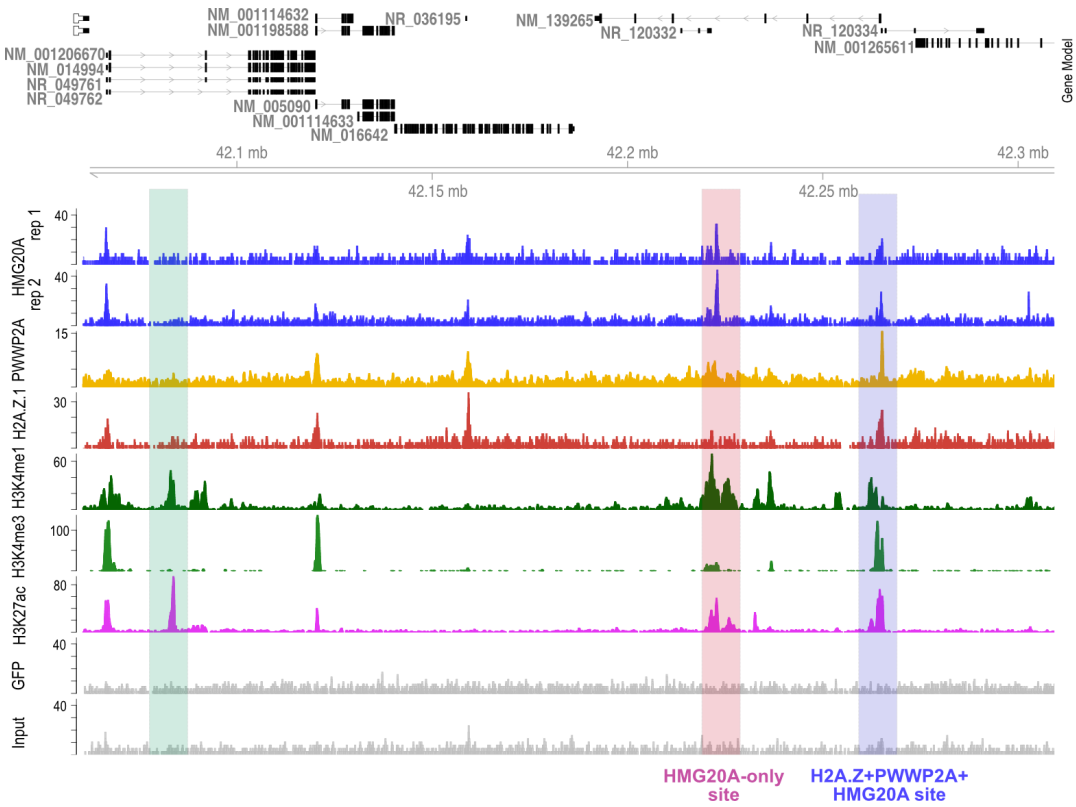


Figure 14: HMG20A binds to specific chromatin sites
Genome browser snapshot of ChIP-seq results from two replicates of GFP-HMG20A (blue tracks), GFP-PWWP2A (orange track), GFP-H2A.Z.1 (red track), H3K4me1 (dark green track), H3K4me3 (light green track), H3K27ac (magenta track), GFP and input material (grey tracks). Green bar highlights a site without H2A.Z/PWWP2A/HMG20A binding, pink bar highlights an HMG20A site devoid of H2A.Z/PWWP2A (HMG20A-only site), blue bar highlights site containing H2A.Z/PWWP2A/HMG20A (H2A.Z+PWWP2A+HMG20A-site).

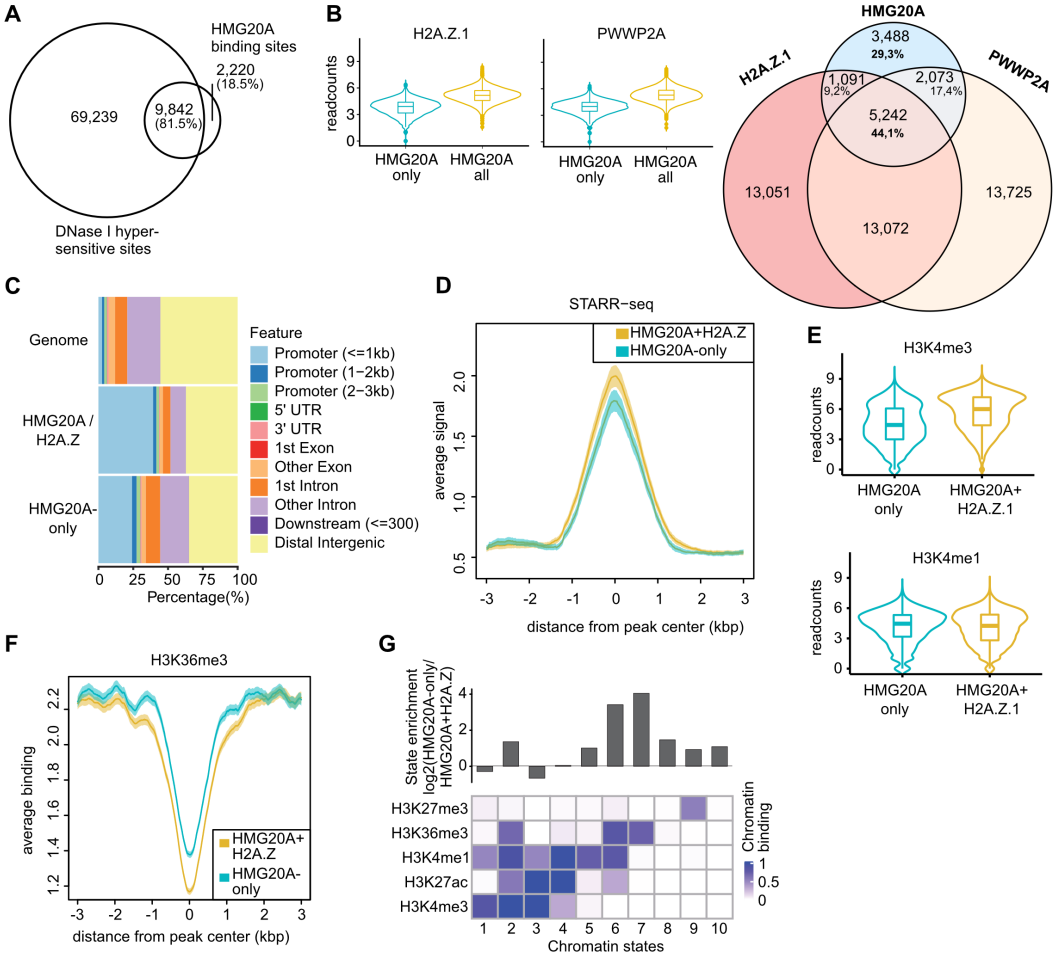


Figure 15: HMG20A binds to H2A.Z- and PWWP2A-occupied promoters and intronic enhancers

(A) Venn diagram showing overlap of HMG20A ChIP-seq binding sites compared to published DNase I-hypersensitive sites in Hela cells (https://www.encodeproject.org/files/ENCFF526VFR). (B) Left: Violin plots of ChIP-seq read counts of GFP-H2A.Z (left) and GFP-PWWP2A (right) at GFP-HMG20A binding sites. Right: Venn diagram showing the numbers of GFP-H2A.Z, -PWWP2A and -HMG20A ChIP-seq binding sites and their overlap. (C) Feature plot depicting distribution of HMG20A+H2A.Z and HMG20A-only binding along genomic features. (D) Distribution of publicly available STARR-Seq signal along HMG20A+H2A.Z and HMG20A-only sites (Muerdter et al., 2018). I Violin plots showing read counts of publicly available H3K4me1 or H3K4me3 at HMG20A+H2A.Z and HMG20A-only binding sites. (F) Average enrichment plot of ENCODE H3K36me3-containing regions over HMG20A-only (yellow) and HMG20A+H2A.Z (blue) ChIP-seq sites. (G) ChromHMM (Ernst and Kellis, 2012; Ernst and Kellis, 2017)-based enrichment of chromatin states (due to enriched post-translation histone modification sites) of GFP-HMG20A-only compared to HMG20A+H2A.Z-containing genomic regions

In general, HMG20A was found to be bound to accessible, Dnase I-hypersensitive sites (Figure 15A) (provided by the ENDCODE consortia). Surprisingly, HMG20A bound to chromatin in two different modes. Firstly, it bound to regions that are less occupied by H2A.Z and PWWP2A (HMG20A-only) compared to all HMG20A binding sites (Figure 15B, top), making up approximately 30% of the identified HMG20A binding sites. Secondly, it was found together with H2A.Z and PWWP2A (H2A.Z+HMG20A sites) at approximately 44% of the HMG20A binding sites (Figure 15B, bottom). Although H2A.Z+HMG20A sites were prominently located in promoters, HMG20A-only sites showed an enrichment on promoter as well as on intronic regions (Figure 15C). An enrichment of HMG20A bound sites in regulatory regions was independently confirmed by a strong correlation of HMG20A-bound regions with published STARR-seq data (Muerdter et al., 2018). These two different binding patters were also reflected in a higher co-occurrence with the promotor mark H3K4me3 at H2A.Z and HMG20A sites, as well as a higher co-occupancy with the H3K4me1 enhancer, and the actively transcribed gene body mark H3K36me3 at HMG20A-only sites (Figure 15E and F). To increase the confidence out these results and to better characterize genomic HMG20A binding regions, we performed a more powerful comparison between our data sets and the chromatin states defined by training a 10-state model on ENCODE data stets using ChromHMM (Ernst and Kellis, 2012; Ernst and Kellis, 2017), showing that, in fact, HMG20A-only sites were more associated with chromatin state 6 and 7 (active enhancers in transcribed gene bodies and gene bodies in general), while being less prominent active promoters (state 1 and 3) (Figure 15G).

We have shown HMG20A and H2A.Z sites to be primarily associated with promoters (Figure 15). Hence, we wondered whether HMG20A binds to

nucleosome-depleted regions of transcriptional start sites (TSS) or to the well-positioned surrounded TSS nucleosomes themselves. Unlike PWWP2A, HMG20A seemed not to bind to the H2A.Z containing +1 and -1 nucleosomes, but rather bound to DNA within nucleosome depleted regions (NDRs) (Figure 16).

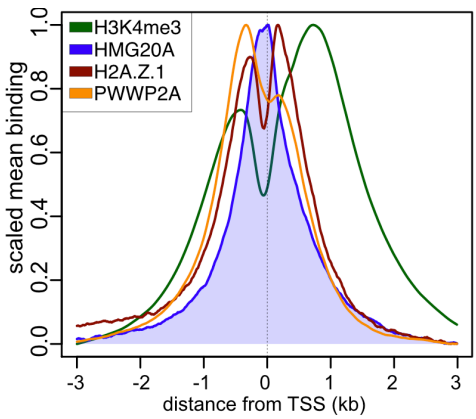


Figure 16: HMG20A binds to nucleosome-depleted regions at transcriptional start sites
Average binding profiles at 500 transcriptional start sites of GFP-HMG20A (blue), -H2A.Z.1 (red), -PWWP2A (orange) and H3K4me3 (green); mean coverage signals at TSS of expressed genes in ChIP-seq experiments.

Since we showed that the N-terminus containing the HMG box bound DNA. We wondered whether this dependency is true in ChIP, too. For that, I, together with Felix Diegmüller, a Bachelor student I supervised, we created Hela cell lines stably expressing GFP-HMG and GFP-CC proteins (Figure 10A) and performed ChIP-qPCR with them. Since neither GFP-HMG nor GFP-CC precipitated DNA efficiently compared to GFP-HMG20A at representative binding sites, it is not clear whether the HMG-box of HMG20A is sufficient for chromatin binding.

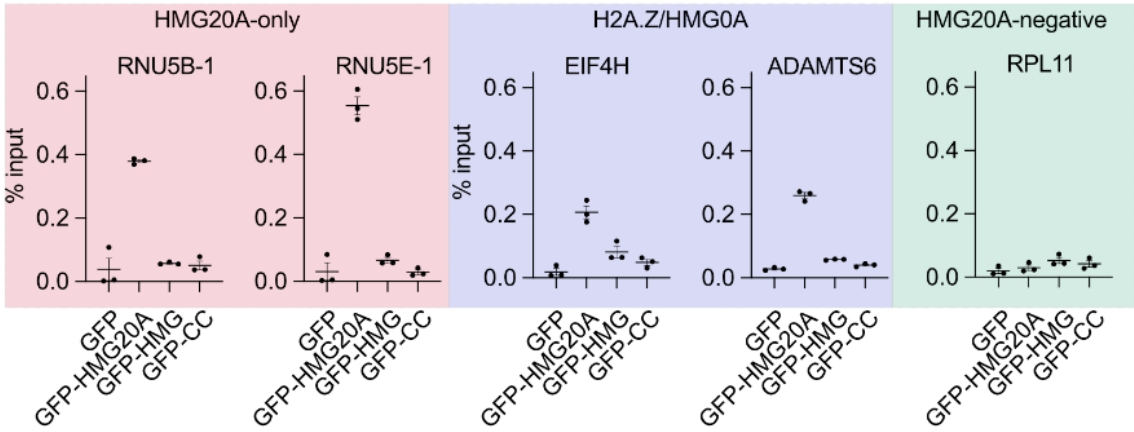


Figure 17: Combination of both functional HMG20A protein domains is required to efficiently precipitate DNA
ChIP-qPCR at selected loci. Shown is percent precipitated DNA compared to input of three biological replicates of GFP, GFP-HMG20A, -HMG or -CC ChIP-qPCR of HMG20A-only sites (red: RNUB-1 and RNUE-1downstream; see red bar in Figure 14 as example), HMG20A+H2A.Z.1-positive sites (blue: EIF4H promoter and ADAMTS3 gene body; see blue bar in Figure 14 as example) and an HMG20A+H2A.Z.1-negative site (green: RPL11 gene body; see green bar in Figure 14 as example) as negative control. Data is presented as mean ± SEM of three biological replicates.

Together, the data presented here show that HMG20A bound 1) to mainly promoter regions occupied by H2A.Z and PWWP2A and 2) independently of H2A.Z or PWWP2A at intronic enhancers of actively transcribed genes. It remains elusive if HMG20A's chromatin interaction is mediated by its HMG-box or its coiled-coil domain, as corresponding truncations fail to precipitate chromatin.

As described above, HMG20A binds to chromatin in two different modes (Figure 14 and Figure 15). We wondered if this is due to different DNA motifs that are represented within these groups. MEME-ChIP motif detection (meme-suite.org) reports Jun Proto-Oncogene / Fos Proto-Oncogene (FOS/JUN), Krueppel-Like Factor (KLF) transcription factor family, and One Cut Domain Family Member 3/ Forkhead Box E1 (ONECUT3/FOXE1) motifs as the most similar motifs in HMG20A+H2A.Z binding sites (Figure 18, top), while DNA motifs in HMG20A-only sites

seem to be more similar to FOS/JUN and Specificity Protein (SP) transcription factor binding motifs (Figure 18, bottom). One cannot conclude that differences in HMG20A binding are caused by different DNA-motifs in combined HMG20A and H2A.Z sites, and HMG20A-only binding sites, since DNA motifs in both binding modes are very similar, although MEME analysis calculates different most similar motifs for them.

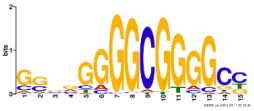
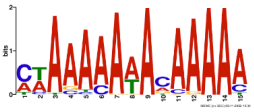
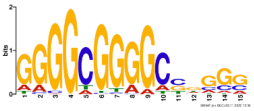
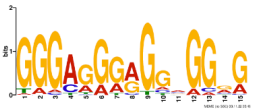
HMG20A-only					
Motif Logo	**Motif Source**	**Width**	**Ratio**	**E-value**	**Most similar motifs**
	MEME	8	1213/5976 (20.30%)	6.6E-110	FOSB::JUNB (MA1135.1) FOSL2 (MA0478.1) FOSL2::JUNB (MA1138.1)
	MEME	15	890/5976 (14.89%)	3.1e-085	KLF15 (MA1513.1) PATZ1 (MA1961.1) KLF12 (MA0742.2
	MEME	15	72/5976 (1.2%)	1.9e-027	ONECUT3 (MA0757.1) ONECUT3_DBD FOXE1 (MA1487.2)
HMG20A+H2A.Z					
Motif Logo	**Motif Source**	**Width**	**Ratio**	**E-value**	**Most similar motifs**
	MEME	15	853/6086 (14.01%)	1.5E-68	SP2 (MA0516.3) SP1 (MA0079.5) KLF12 (MA0742.2)
	MEME	10	710/6086 (11.67%)	5.6E-45	FOSL2::JUN (MA1130.1) FOS::JUND (MA1141.1) Jun (MA0489.2)
	MEME	15	480/6086 (7.89%)	4.6E-36	ZNF263 (MA0528.2) SP5 (MA1965.1) Sp4_primary (UP00002_1)

Figure 18: DNA binding motifs of FOS/JUN, KLF SP and FOXE1 transcription factors are enriched in HMG20A binding sites in Hela cells
Table summarizing top-enriched DNA motifs in HMG20A binding sites, generated by MEME-Chip analysis using the meme-suite.org web interface. Top: DNA motifs enriched in HMG20A-only binding sites. Bottom: DNA motifs enriched in HMG20A+H2A.Z binding sites.

5.4 Loss of HMG20A has little effect on transcriptional regulation in differentiated human carcinoma cell lines

On the one hand, it is reasonable to assume that HMG20A depletion will lead to deregulation of transcriptional programs, as it bound to regulatory genomic regions and major chromatin modifiers. On the other hand, reports on HMG20A's function related it mainly to neural developmental processes and diabetes (Garay et al., 2016; Lorenzo et al., 2021; Mellado-Gil et al., 2018; Rivero et al., 2015; Wynder et al., 2005). To analyze whether HMG20A has an impact on transcriptional regulation in differentiated cancer cell lines, I established HMG20A CRISPR/Cas9-mediated knock out and HMG20A degron (auxin inducible, a kind gift from Prof. Dr. Lienhard Schmitz, Institute for Biochemistry, Giessen, Germany) in diploid HCT116 cells. These HCT116 cells express Transport Inhibitor Response 1 (TIR1) from *Oryza sativa*, that recruits endogenous E2 ubiquitin-conjugating enzymes to auxin-inducible-degron (AID) tagged target protein, in this case HMG20A, leading to its degradation via the proteasome (Natsume et al., 2016). This recruitment is induced by adding synthetic auxin-derived 1-Naphthaleneacetic acid (NAA) to the culture medium.

As depicted in Figure 19A, HMG20A expression could be prohibited by excision of the first ten codons within the *HMG20A* locus. Applying AID, HMG20A was depleted within 2 hours. Note that the AID tag resulted in a change of HMG20A's molecular weight compared to the untagged wild type protein (Figure 19B). Removal of NAA rescues the expression of HMG20A after approximately 12 hours (Figure 19C).

Auxin-independent degradation of AID-HMG20A can be prevented by adding freshly prepared TIR1 inhibitor compound auxinole (a kind gift

from Prof. Dr. Wibke Diederich (Institute of Pharmaceutical Chemistry, Marburg, Germany) and Prof. Dr. Alexander Brehm (Institute of Molecular Biology and Tumor Research, Biomedical Research Center, Marburg, Germany)) to the culture medium (Yesbolatova et al., 2019) (Figure 19D). Since the addition of auxinole for several days has a significant influence on cell morphology (data not shown), I decided not to use this system for functional studies of HMG20A.

As an alternative approach, I performed RNAi by siRNA transfection in Hela cells and performed mRNA sequencing of two of the three independent biological replicates shown here, after I tested the RNA integrity of isolated total RNA a native agarose gel (Figure 19E). After mRNA sequencing of those samples, data was analyzed together with Tobias Friedrich (Prof. Dr. Tilman Borggrefe, Institute for Biochemistry, Justus Liebig University Giessen, Germany).

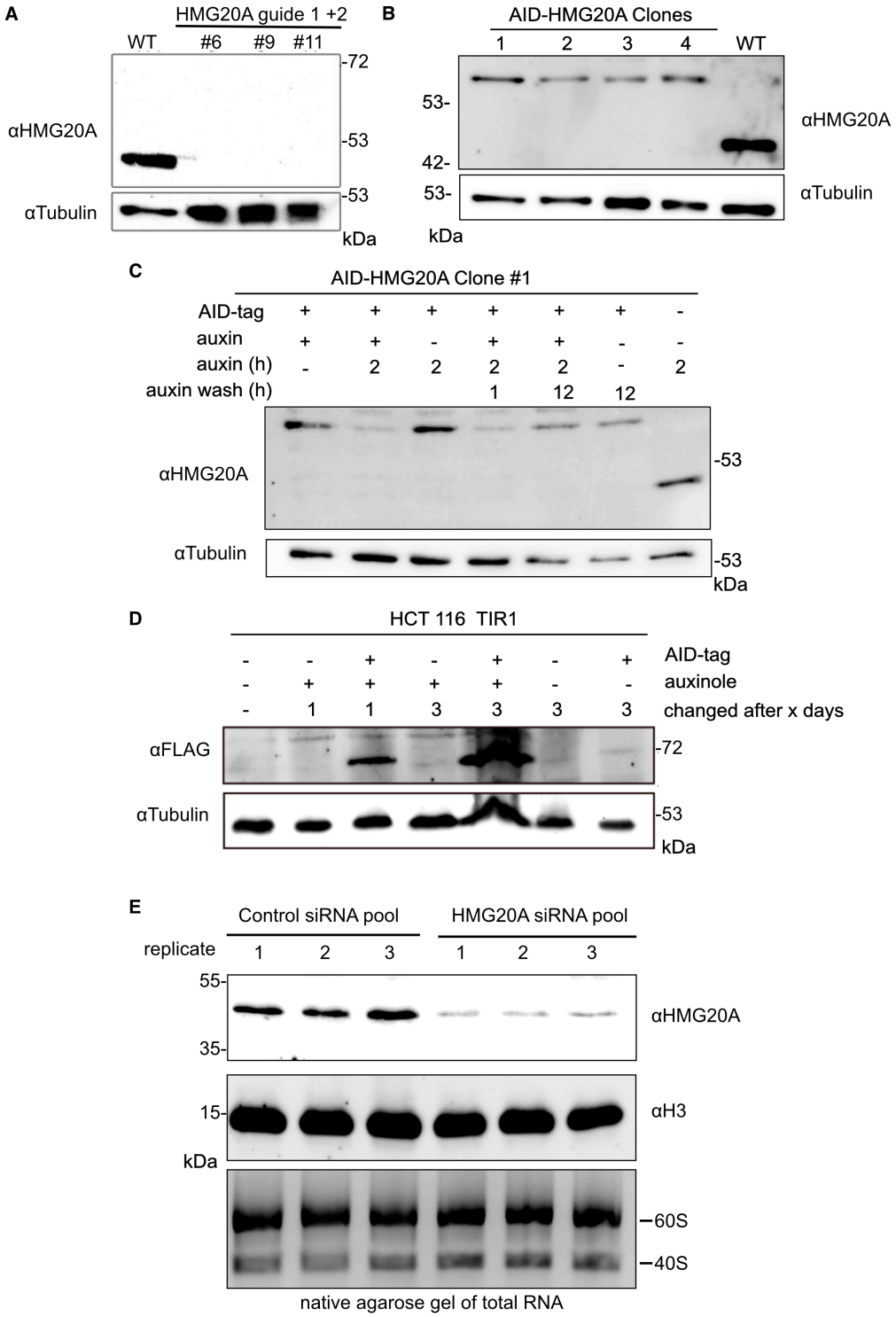


Figure 19. Successful HMG20A depletion in Hela and HCT116 cell lines
(A)-(E) Immunoblots of HMG20A in indicate depletion scenarios, Immunoblots of H3 or Tubulin acted as loading controls. (A) HMG20A was depleted by CRISPR/Cas9 in HCT116 cells. (B) HMG20A is endogenously tagged with a mAID domain (mini auxin inducible degron) domain. (C) AID-HMG20A is depleted in an auxin-dependent manner in HCT116 cells expressing TIR1. (D) Auxin independent depletion of HMG20A can be rescued by adding a fresh auxinole compound. (E) HMG20A is depleted by HMG20A targeting siRNA pool.

The integrity of isolated total RNA from the indicated samples was analyzed by native agarose gel electrophoresis (bottom).

Indeed, as suggested in the literature, HMG20A depletion had only a small effect (170 differentially regulated genes) on the transcriptome of already differentiated Hela cells (Figure 20A).

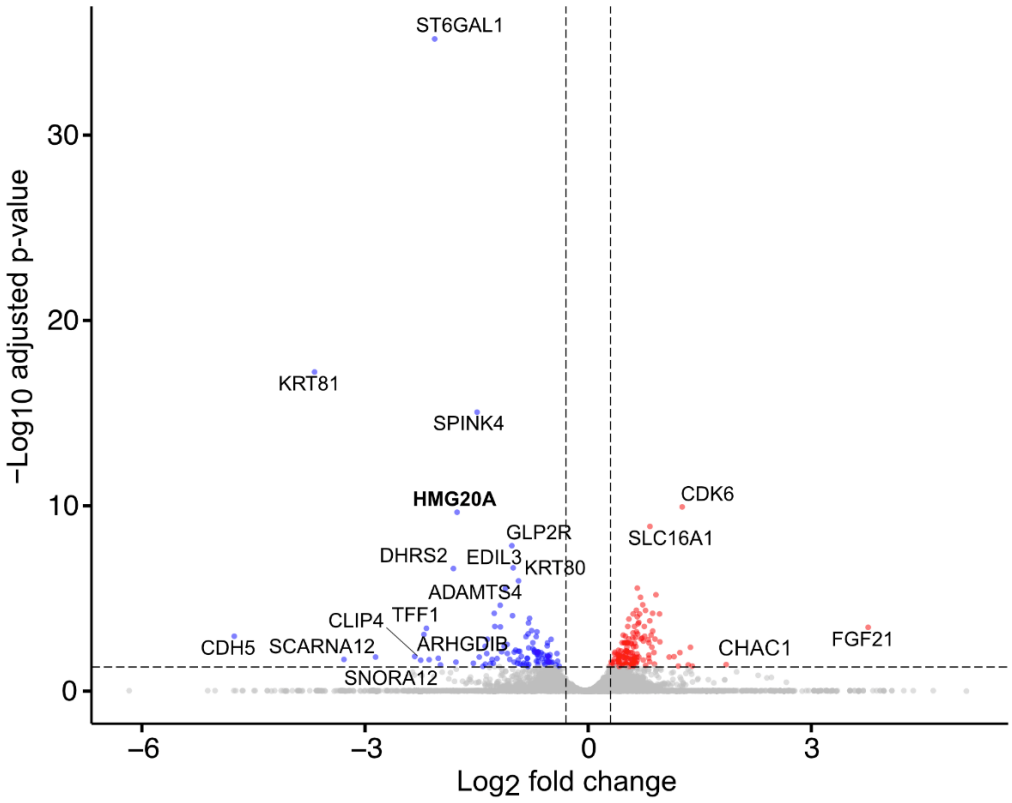


Figure 20: HMG20A depletion does not lead to major transcriptional deregulation in Hela cells
Scatter plot of significantly deregulated (log2 fold change <-0.3 and >0.3, adjusted p-value <0.05, Wald test) mRNAs from two independent siRNA-mediated HMG20A depletion experiments analyzed by mRNA sequencing. Red: up-regulated transcripts; blue: down-regulated transcripts.

Next, I set out to validate the changes in expression levels of some of the genes (FGF21 and CHAC1) by RT-qPCR. Unfortunately, the via mRNA-seq determined differences in transcription upon HMG20A-depletion could not be validated (Figure 21). These results suggest that in Hela cells HMG20A is not required for transcriptional regulation.

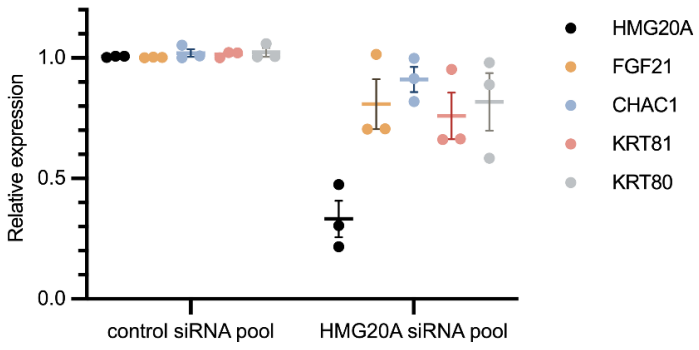


Figure 21: Deregulated genes after HMG20A depletion in mRNA sequencing cannot be validated in independent experiments. Quantitative PCR of reverse-transcribed mRNAs of significantly deregulated genes identified in (Figure 17) normalized to HPRT expression. Expression values are for three independent experiments each. The error bars represent SEM, n=3.

Since the protein-protein interaction data from HMG20A presented above, supported the idea that it forms a protein complex with PHF14, RAI1 and TCF20 (Eberl et al., 2013), I intended to analyze, whether they regulate expression of similar genes. As for HMG20A, depletion of PHF14, TCF20 and RAI1 deregulates only 136, 276 and 13 genes respectively (Figure 22A and B). Furthermore, there was only little overlap of genes deregulated in all PRTH RNAi scenarios (two biological replicates per RNAi pool) (Figure 22C).

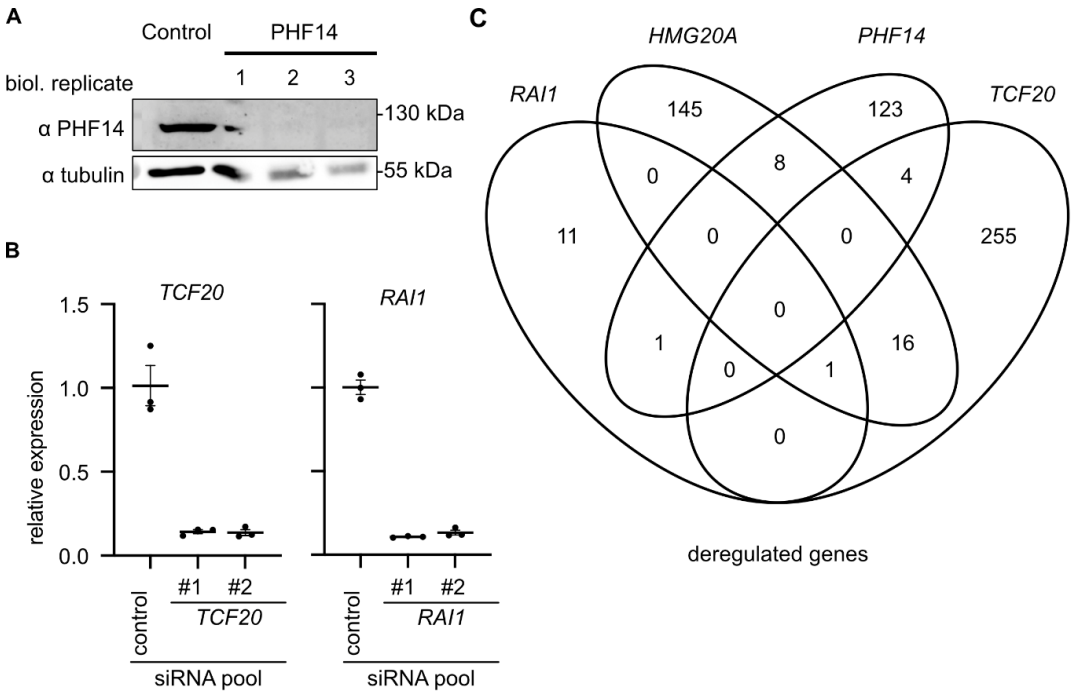


Figure 22: PHF14, RAI1, TCF20, and HMG20A do not regulate the same transcriptional programs in Hela cells
(A) PHF14 is depleted by RNAi mediated by PHF14-specific siRNA pools. Proteins were detected by Immunoblotting with anti-PHF14 and anti-tubulin antibodies. (B) TCF20 (right) and RAI1 (left) mRNA is depleted upon RNAi mediated by gene-specific siRNA pools, respectively. The relative abundance was detected by quantitative polymerase chain reaction of TCF20 and RAI1 mRNA, normalized to HPRT expression. Error bars indicate SEM of three technical replicates (n=3). (C) Venn diagram depicting the number and overlap of deregulated genes in HMG20A, PHF14, TCF20, and RAI1 RNAi mRNA sequencing experiments.

Although Hela cells reduced endogenous HMG20A levels, when GFP-HMG20A was overexpressed, implying a dosage depended effect (Figure 6C), HMG20A was associated with chromatin modifiers (Figure 7) and bound to regulatory regions in chromatin (Figure 14, Figure 15 and Figure 16), its depletion had only little effect on transcriptional regulation.

These data, as well as published reports, suggest that the main biological role of HMG20A lies not in differentiated cells, but rather in developmental processes (Ceballos-Chávez et al., 2012; Gómez-Marín et al., 2022; Rivero et al., 2015; Wynder et al., 2005).

5.5 HMG20A regulates the formation of the neural crest and heart

HMG20A, in particular its predicted HMG box and coiled-coil domain, are highly conserved from *Homo sapiens* over *Mus musculus* to *Xenopus laevis* (Figure 23). This is why I decided to cooperate with Stefanie Gossen from the laboratory of Prof. Dr. Annette Borchers (Department of Biology, Molecular Embryology, Marburg, Germany) to study the influence of hmg20a on the easy-to-manipulate embryonic development of *Xenopus laevis*.

```
                1         10        20        30        40        50        60        70        80        90        100
H. sapiens      ENLMTSSTLPPLFADEDGSKESNDLATTGLNHPEVPYSSGAT-------SSTNNPEFVEDLSQGQ-LLQS-ESSNAAEGNEQRHEDEQRSKRGGWSKGRKRKKPL
M. musculus     ESLMASSTLPPLFADEDGSKESNDLATSGLTHPEGPYGSAAT-------S-TTNPEFVEDLSQGQ-LLQS-EASNAVEGNEQRPEDEQRSKRGGWSKGRKRKKPL
X. laevis       E--STASAVPPSSED---------LVADTKENNQPPFCGTTVSGSSQAPLHPHSPTLQQDEREELTLHQSGEQQLGNSGELRQEEELPKARRGGWNKGRKRKRSP
                *   ::*::**   *         *.:    : : *: . :.           .* : :*  :   * ** * . . .*: :: *:  :::****.******:
                                                  HMG box
                    110       120       130       140       150       160       170       180       190       200       210
H. sapiens      RDSNAPKSPLTGYVRFMNERREQLRAKRPEVPFPEITRMLGNEWSKLPPEEKQRYLDEADRDKERYMKELEQYQKTEAYKVFSRKTQDRQKGKSHRQDAARQATH
M. musculus     RDSNAPKSPLTGYVRFMNERREQLRAKRPEVPFPEITRMLGNEWSKLPPEEKQRYLDEADRDKERYMKELEQYQKTEAYKVFSRKTQDRQKGKSHRQDAARQATH
X. laevis       RDNNAPKAPLTGYVRFMNERREQLRTERPDVPFPEITRIVGSEWSKLPAHEKQHYLDEAEKDKERYTKELQKYQNTDAYQTYSRKAKSRQKGRQQRQEGVRGVPS
                **.****:*****************:.**:********:.*.****** .***:*****:.***** ***:.**:*:**:.:***:.:****:.:**:..* .
                                                  coiled-coil
                        220       230       240       250       260       270       280       290       300       310
H. sapiens      DHEKETEVKERSVFDIPIFTEEFLNHSKAREAELRQLRKSNMEFEERNAALQKHVESMRTAVEKLEVDVIQERSRNTVLQQHLETLRQVLTSSFASMPLPGSGET
M. musculus     DHEKETEVKERSVFDIPIFTEEFLNHSKAREAELRQLRKSNMEFEERNAALQKHVESMRTAVEKLEVDVIQERSRNTVLQQHLETLRQMLTSSFASMPLPGSGEI
X. laevis       NTEKESILKERPIFDIPIFTEEFLNHSKAREAELRQLRKSNMEFEERNAALQKHVESMRSAVQRLEAELSQEHERNSLLQQHLQSVRQALTHCLQSVPVPGTTET
                : ***: :*** :**************************************************:**:.**.:: **:.**:.*****:::** ** .: *:*:**: *
                    320       330       340       350
H. sapiens      PTVDTIDSYMNRLHSIILANPQDNENFIATVREVVNRLDR
M. musculus     PTVDTIDSYMNRLHSIILANPQDNENFIATVREVVNRLDR
X. laevis       PTLETIDLYMSRLQNAVLTHPKESEVIISGVREVLSQLEG
                **::*** **.**:. :*::*::.* :*: ****:.:*:
```

Figure 23: The HMG20A protein is highly conserved in *H. sapiens*, *M. musculus* and *X. laevis*
HMG20A amino acid sequences of indicated species aligned by Clustal Omega (Madeira et al., 2022). Blue letters indicate the predicted HMG box, purple letters the predicted coiled-coil domain, asterisks fully conserved residues, colons conservation between groups of strongly similar properties, and periods conservation between groups of weakly similar properties.

In order to analyze hmg20a expression patterns in *Xenopus laevis* embryos during development, we applied RNA *in situ* hybridization of the *hmg20a.L* gene. Maternally expressed *hmg20a* was detected in early

stages of cleavage and blastula (Figure App 2C-F, Appendix), while high and ubiquitous zygotic expression occurs in stages of gastrula (Figure App 2B and G, Appendix). At neurula stages, *hmg20a.L* was expressed in neural folds and cells of the cranial neural crest (Figure App 2 K-N, Appendix), followed by expression in cells of the migratory cranial neural crest, the brain, and the eyes in the stages of tadpole (Figure App 2O-R, Appendix). Taken together, we identified *hmg20a.L* to be specifically expressed in neural and neural crest progenitor cells in neurula, already implying that NCC differentiation might be affected by hmg20a depletion.

To investigate the effect of hmg20a on development, and to analyze, if NCC differentiation is indeed perturbed by loss of hmg20a, its translation was blocked by injection of anti-hmg20a morpholinos. Developmental defects of morphants revealed its role in cartilage formation due to the abnormal pattern of twist-expressing neural crest cells, cartilage formation, heart anatomy and contractility, caused by disturbed *mhcα* expression, (Figure App 3, Appendix).

As hypothesized, due to *hmg20a* expression patterns documented by RNA *in situ* hybridization and loss of function phenotypes of HMG20A interacting proteins PWWP2A and RAI1 (Pünzeler et al., 2017; Tahir et al., 2014), hmg20a did indeed regulate NCC differentiation, resulting in cartilage and craniofacial defects. Surprisingly, heart formation was affected as well, a phenotype that was not described for PWWP2A and RAI1.

5.5.1 *Depletion of HMG20A in mouse embryonic stem cells affects neural crest differentiation in vitro*

The developmental defects of *Xenopus laevis* tadpoles encouraged me to further analyze the effects of Hmg20a depletion on a molecular level. For that I switched to mouse embryonic stem cells (mESCs), where I investigated HMG20A's influence on differentiation of NCC and cardiomyocytes (CM) specifically.

I inserted, in alliance with Dr. Jie Lan (Postdoctoral fellow at the Institute for Genetics, Justus Liebig University Giessen), a cassette of a respective 'selector gene' into both *Hmg20a* alleles. I used the *mCherry* gene for the first allele and a resistance gene against puromycin for the second allele. The selector gene was followed by a tandem of transcriptional terminator sequences directly downstream of the *Hmg20a* start codon (Figure 24A), resulting in the depletion of its RNA and protein (Figure 24B, C). Notice that although there was no HMG20A protein detectable, varying amounts of mRNA remained verifiable in individual clones.

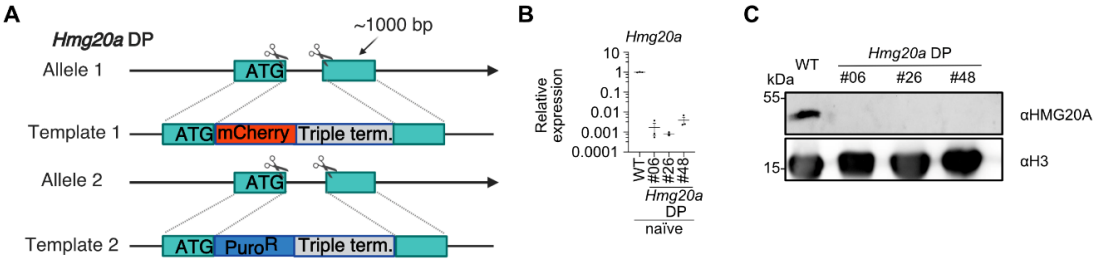


Figure 24: CRISPR/Cas9-based depletion of HMG20A in mouse embryonic stem cells
(A) Schematic representation of the generation of *Hmg20a* depletion (DP) mouse embryonic stem cells by introducing mCherry followed by triple terminator sites and puromycin resistance followed by triple terminator site into both *Hmg20a* alleles directly after the start codon using a CRISPR/Cas9-based approach. (B) Right: Relative *Hmg20a* mRNA expression measured by RT-qPCR of isolated mRNA from naïve state wild-type cells and three individual *Hmg20a* DP cell clones normalized to *Hprt* expression. Error bars indicate the SEM of three technical replicates. (C) HMG20A immunoblot of whole cell

extracts from naïve state wild type mESCs and three *Hmg20a* DP cell clones. Anti-H3 antibody staining served as loading control.

With HMG20a depleted mESCs established, I analyzed their capability to differentiate to NCC compared to WT mESCs (Figure 25A). As expected, pluripotency factor Octamer-Binding Protein 4 (*Oct4*) (Schöler et al., 1990) becomes repressed, while the key markers of neural crest Snail Family Transcriptional Repressor 2 (*Slug*) (Nieto et al., 1994), Twist Family basic helix-loop-helix (BHLH) Transcription Factor (*Twist*) (Soo et al., 2002) and *Cadherin* 2 (*Cdh2*) (Nakagawa and Takeichi, 1998) become expressed (Figure 25B). In line with the RNA *in situ* hybridization data in *Xenopus laevis* (Figure App 2, Appendix), *Hmg20a* expression increased in NCCs (Figure App 2K-N, Appendix and Figure 25B). When HMG20A was depleted, neural crest differentiation was strongly impaired, as reflected by the reduced expression of neural crest marker genes like Paired Box 3 (*Pax3*) (Conway et al., 1997), *Twist*, *Slug*, *Cdh2* and *Cadherin* 1 (*Cdh1*) (Nakagawa and Takeichi, 1998) at Day9 of the neural crest differentiation protocol (Figure 25C left). Note that there are clonal effects where marker gene expression correlates with residual *Hmg20a* expression levels in individual DP clones (Figure 25C right), indicating that even small amounts of HMG20A are sufficient to promote neural crest differentiation.

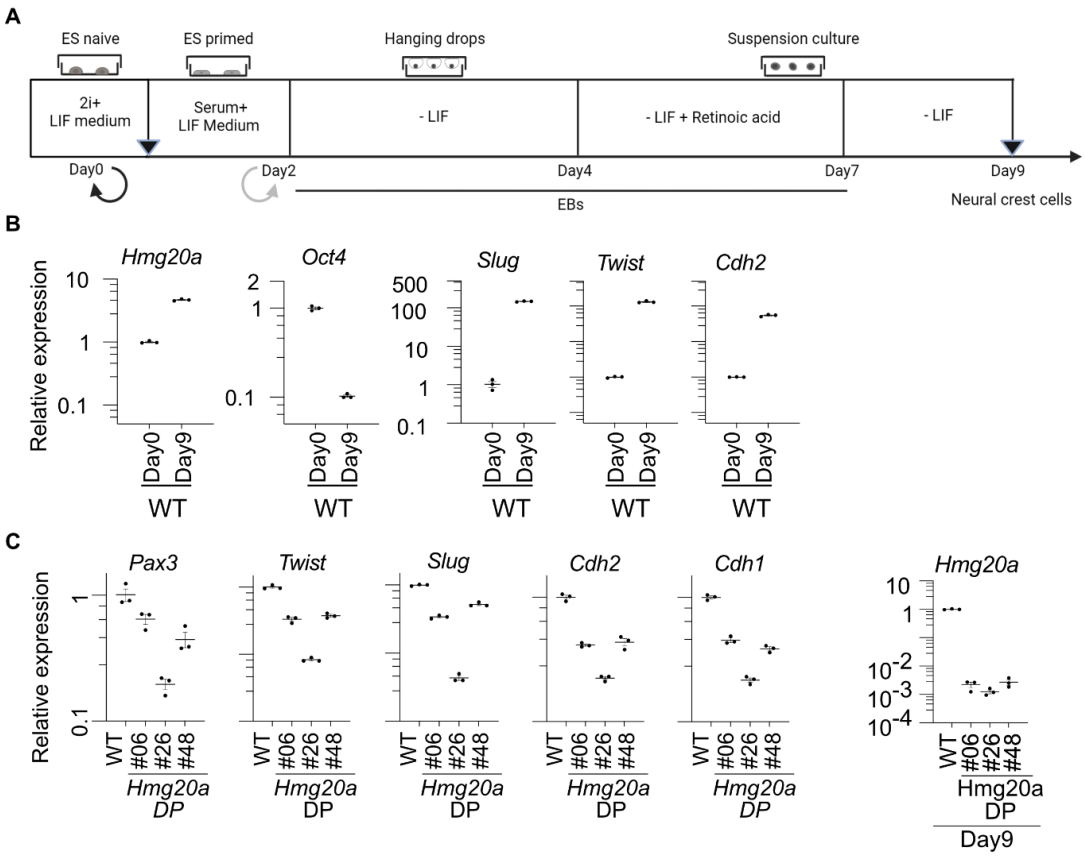


Figure 25: Hmg20a-depleted MESCs do not differentiate into neuronal crest cells

(A) mESCs neural crest differentiation scheme. (B-C) Relative expression measured by RT-qRCR of the *Oct4* pluripotency marker and the NCC marker genes *Slug*, *Twist* and *Cdh2*, in WT cells on Day0 and Day9 of the neural crest differentiation protocol (B), of *Pax3*, *Twist*, *Slug Cdh2* and *Chd1* in WT cells and 3 individual *Hmg20a* DP clones at Day9 (C,left) and *Hmg20a* (C, right) on Day9. Expression was normalized to *Hprt*, 16S RNA, and *Gapdh* expression. Error bars indicate SEM of three technical replicates.

A hallmark of neural crest cells is their ability to migrate, as they distribute throughout the growing embryo to precise target destinations. To analyze migration capabilities of Hmg20a depleted NCCs in embryoid bodies. Embryoid bodies (EBs) were seeded on galantine coated Petri dishes overnight. The next day spreading, i.e., migration of cells away from the embryoid bodies, was monitored by phase contrast microscopy (Figure 26, top). Characterization and quantification of migration was done by Tim M. Wunderlich (PhD student at the Institute for Genetics, Justus Liebig University Giessen) in a blinded manner (Figure 26, bottom).

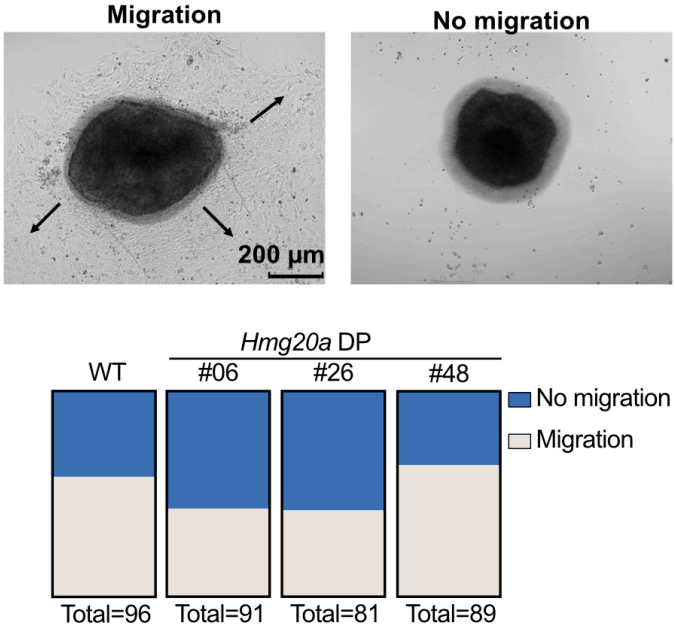


Figure 26: Depletion of Hmg20a leads to a slight decrease in migration ability of neural crest cells.
Top: Representative microscopy pictures of EBs at Day9 of migrating (right) cells. Bottom: Quantification of migration capability of cells from wild type and three *Hmg20a* DP embryoid body (EB) cell clones based on visual inspection (see top pictures).

There is a little effect on the neural crest migration capability upon HMG20A depletion. Like in X. *laevis*, Hmg20a depleted NCCs tended to migrate less than WT NCCs. But since only about 60% of WT EBs started to migrate these results are hard to compare to generated data in X. *laevis*. And should be interpreted with caution.

5.5.2 *HMG20A is essential for in vitro cardiomyocyte differentiation*

Since morphants not only showed anatomical heart malformations, but also reduced heart muscle contractility (Figure App 3, Appendix), I, together with Dr. Jie Lan, decided to analyze the role of Hmg20a in cardiomyocyte differentiation, since cardiomyocytes mediate heart muscle contraction.

To examine the involvement of Hmg20a in cardiomyocyte formation, we initiated differentiation by adding vitamin C (ascorbic acid) while cells were grown in hanging drops (Figure 27A). Notice that the cardiomyocyte marker genes Actin Alpha 2 (*Acta2*) and GATA Binding Protein 2 (*Gata2*) were induced, while, unlike during neural crest differentiation, the expression of *Hmg20a* was hardly changed (Figure 27B).

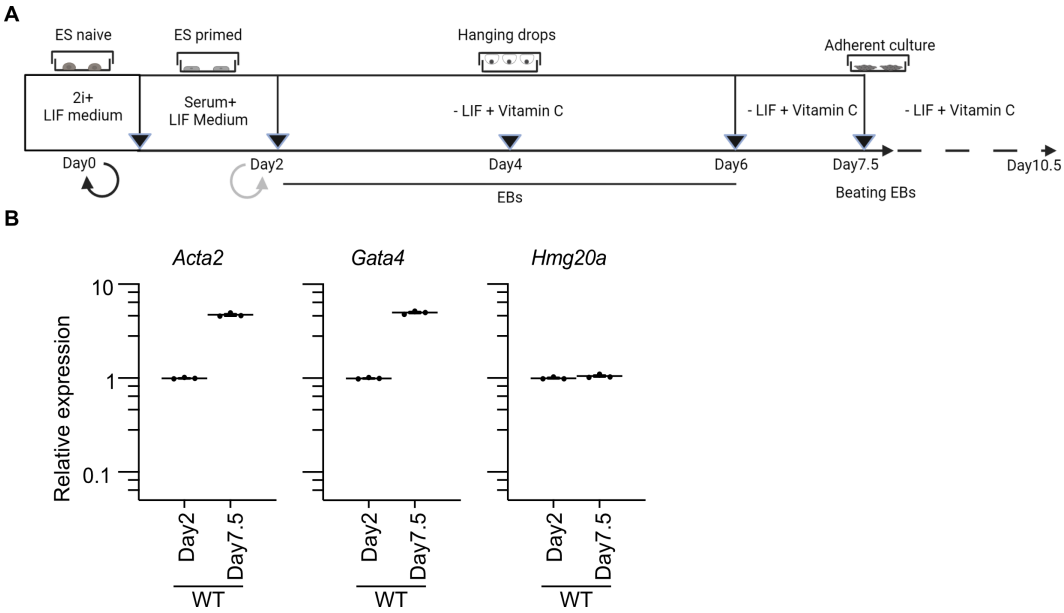


Figure 27: Mouse embryonic stem cells differentiate to cardiomyocytes in hanging drops with vitamin C supplementation.
(A) Mouse embryonic stem cell cardiomyocyte differentiation scheme. (B) Relative expression of cardiomyocyte marker genes *Acta2*, *Gata4*, and *Hmg20a* on Day2 and Day7.5 of cardiomyocyte differentiation measured by RT-qPCR (left). Expression was normalized to *Hprt* and *Gapdh* expression. Error bars indicate SEM of three technical replicates.

Although *Hmg20a* expression was not changed i.e., elevated during cardiomyocyte differentiation, it proved to be essential for this very process. *Hmg20a* DP cells not only grew slower, but also often failed to form large colonies of beating cardiomyocytes (Figure 28A and intermediate state Day6 quantified in Figure 28B).

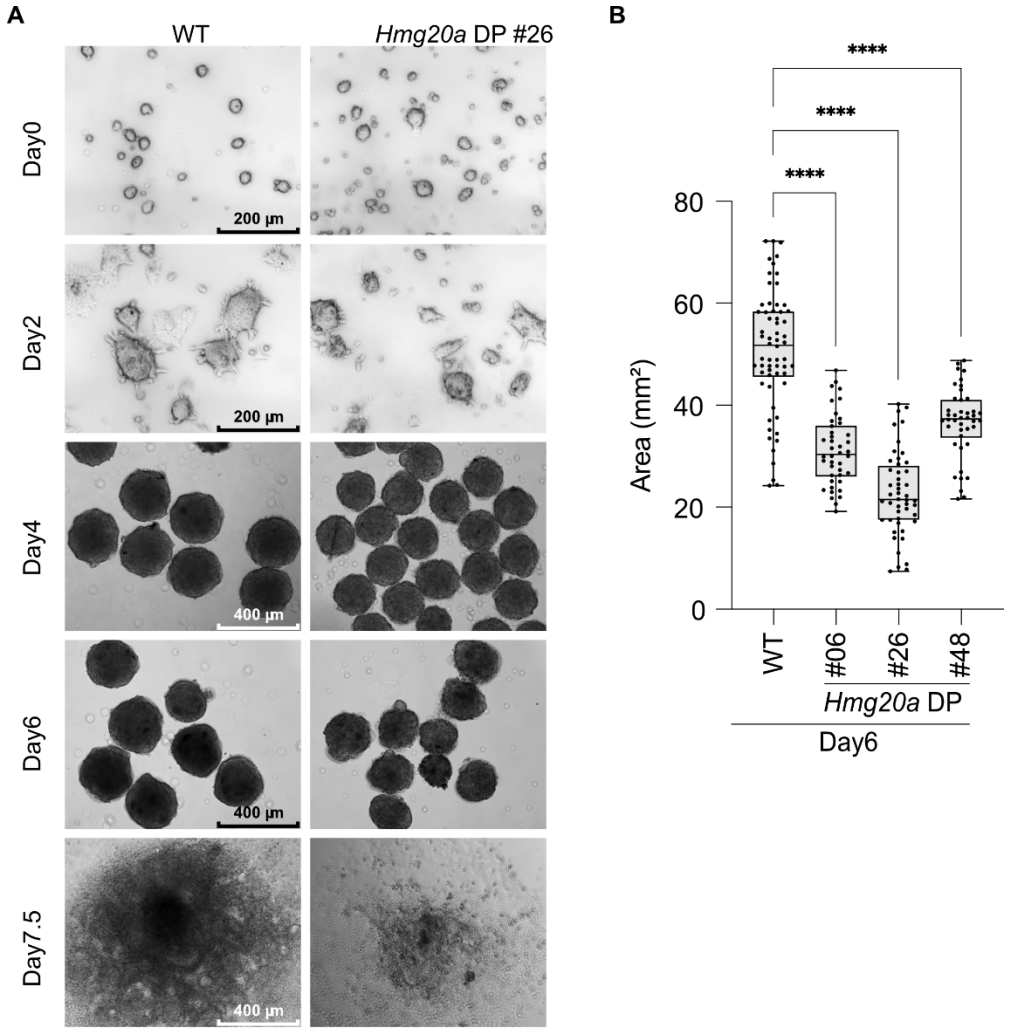


Figure 28: *In vitro* cardiomyocyte differentiation is slowed-down in *Hmg20a* DP cells
(A) Representative phase-contrast microscopy images of wild-type and *Hmg20a* DP mouse embryonic stem cells during cardiomyocyte differentiation. Scale bar: 200 or 400 µm. (B) Sizes of EBs of WT and three individual *Hmg20a* DP cells on Day6 of the CM differentiation protocol. Number of EBs measured indicated above, **** $p < 0.001$, (two tailed Mann Whitney test).

While 98% percent of WT cell colonies did start beating on Day7.5 colonies formed by *Hmg20a* DP cells did only beat to roughly 22 %,

3.5 % and 4.5 %, respectively (Figure 29A, top). When culturing of adherent colonies was prolonged for three additional days until Day10.5 (Figure 27A), WT cells started to die while more colonies from *Hmg20a* DP clones started to beat (58.5 %, 10 % and 81.5 %, respectively) (Figure 29A, bottom), implying a delay of cardiomyocyte differentiation of *Hmg20a* DP cells rather than a total loss of it. Again, similar to the expression of the neural crest marker genes, downregulation of the key cardiomyocyte maker genes T-Box Transcription Factor 5 (*Tbx5*), Myocyte Enhancer Factor 2C (*Mef2c*) (Akerberg et al., 2019) and *Acta2* as well as severity of delayed beating closely correlated with the residual expression of Hmg20a on Day7.5 (Figure 25 C, right Figure 29B). Further indicating, that even residual amounts of Hmg20a might promote differentiating processes.

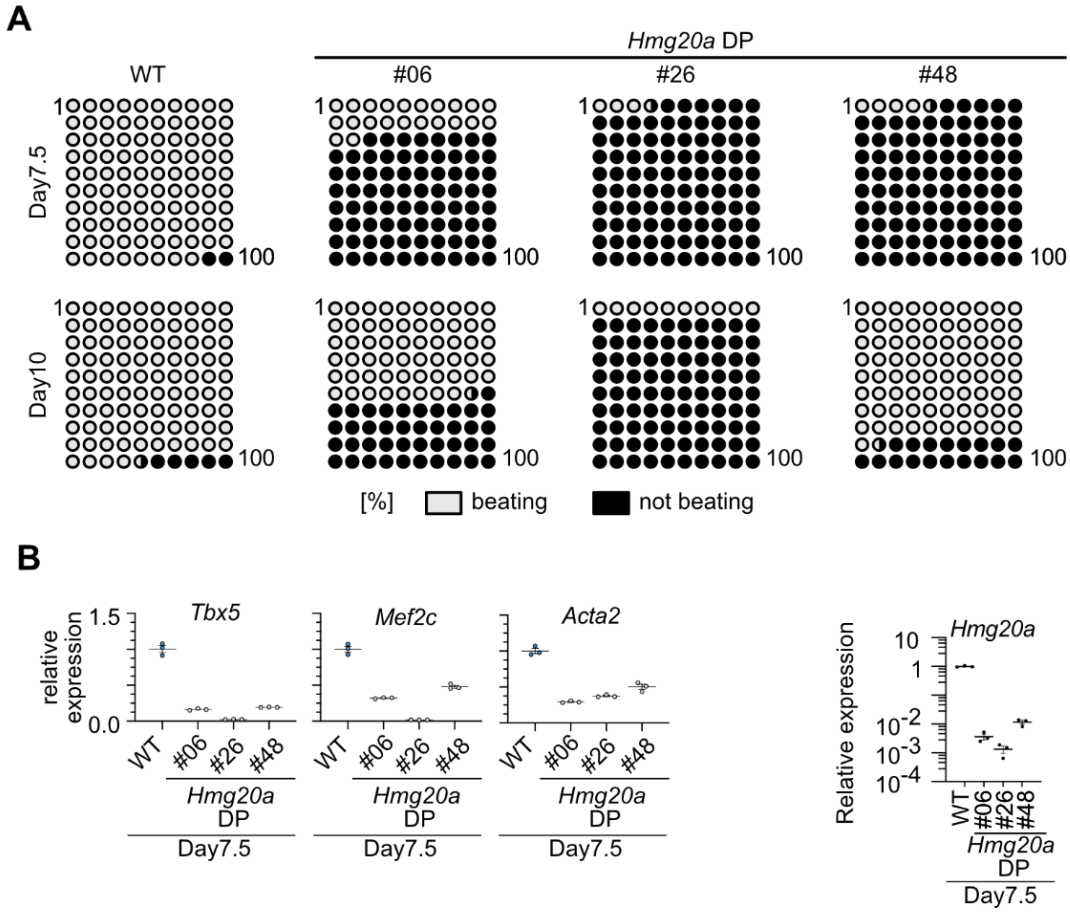


Figure 29: Cardiomyocyte differentiation is delayed in *Hm20a* DP cells

(A) Depiction of percent of beating colonies (gray) or non-beating colonies (black) on Day7 (top) or Day10 (bottom) of the cardiomyocyte differentiation procedure. (B) Relative expression measured by RT-qPCR of Tbx5, Mef2c, Acta2 (left), and *Hmg20a* in WT and three independent *Hmg20a* DP cell clones on Day2 and Day7.5 of the cardiomyocyte differentiation protocol. normalized to *Hprt*, 18S RNA and *Gapdh* expression. The error bars indicate the SEM of three technical replications.

To understand how HMG20A controls cardiomyocyte differentiation mechanistically, I sought out to analyze HMG20A's influences on transcription in cardiomyocyte differentiation over time. For that, I repeated the CM differentiation protocol with *Hmg20a* DP clone #26, as this one showed the lowest residual *Hmg20a* expression and the strongest intensity of phenotypes (Figure 25, Figure 27, Figure 28, Figure 29), isolated total RNA at indicated time points Day0, Day2, Day4, Day6 and Day7.5 (arrow heads in Figure 28A) and performed mRNA sequencing. Data analysis was performed in conjunction with Tobias Friedrich (Prof. Dr. Tilman Borggrefe, Institute for Biochemistry, Justus Liebig University Giessen, Germany).

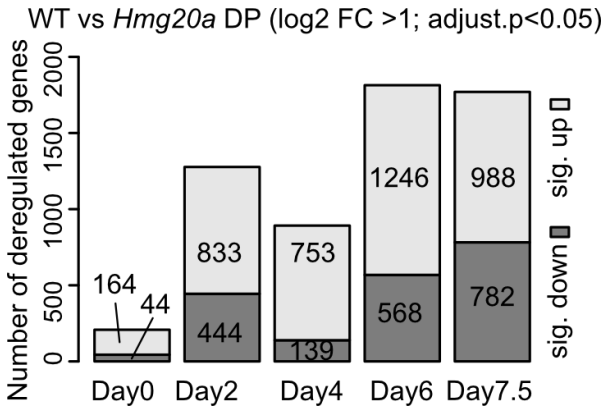


Figure 30: HMG20A regulates genes expression during cardiomyocyte differentiation
Stacked Bar plot of numbers of significantly up (log2 fold change >1) and down (log2 fold change <-1) regulated genes (adjusted p<0.05) during indicated time points of cardiomyocyte differentiation of wild type and *Hmg20a* DP clone #26 as identified by mRNA-seq with DEseq2 (Love et al., 2014) (see for Figure 27A for cardiomyocyte differentiation scheme).

Depletion of Hmg20a had, similar to fully differentiated human Hela cells (Figure 20), only a minor effect on transcriptional regulation in naive stem cells (Day0). Interestingly, HMG20As influence on transcription was greatly increased after a differentiation stimulus (removal of MEK inhibitor (MEKi) and GSK3 inhibitor (2i) and addition of fetal calf serum). The number of deregulated genes increased dramatically, with more genes being up-regulated than down-regulated upon loss of HMG20A (Figure 30) from this time on.

5.5.2.1.1 *Transcriptional repression of Hmg20a correlates with chromatin accessibility in Hmg20a DP*

Because Day2 of the CM differentiation protocol was the earliest timepoint with increased transcriptional deregulation (Figure 30), and both NCC and CM phenotypes could be explained by an deregulation at this time point, since both protocols have the same treatment at that time see Figure 25A and Figure 27A). I focused from now on Hmg20a role in transcriptional regulation at this timepoint in differentiation.

Because, depletion of HMG20A caused mostly upregulation of genes (Figure 30), and I identified its interaction to transcriptional repressors, especially NuRD in Hela cells (Figure 8), I wondered, if transcriptional deregulation correlated with the chromatin accessibility of those genes. I went on to measure genome-wide differences in chromatin accessibility by ATAC-Seq.

Here, chromatin was probed with a hyperactive mutant Tn5 transposase that inserts sequencing adapters into open regions of the genome, causing fragmentation, so-called tagmentation, of genomic DNA. These fragments were purified, amplified by PCR, and sequenced. Mapping back the resulting sequence information to a reference genome allowed

the reconstruction of DNA accessibility - or, put differently, density of nucleosomes - genome-wide (Buenrostro et al., 2013). Data analysis was performed together with Tobias Friedrich (Prof. Dr. Tilman Borggrefe, Institute for Biochemistry, Justus Liebig University Giessen, Germany).

Results from this ATAC-seq experiment showed that, firstly the majority of accessible sites does not change and secondly, that some sites gain, while fewer sites lose accessibility upon loss of HMG20A (Figure 31A). Gene set enrichment analysis of differentially accessible regions with deregulated genes in *Hmg20a* DP cells revealed a correlation of changes in DNA accessibility with gene transcription (Figure 31B and C), leading to the conclusion that transcriptional deregulation by HMG20A depletion is indeed tightly correlated to the accessibility of the chromatin landscape.

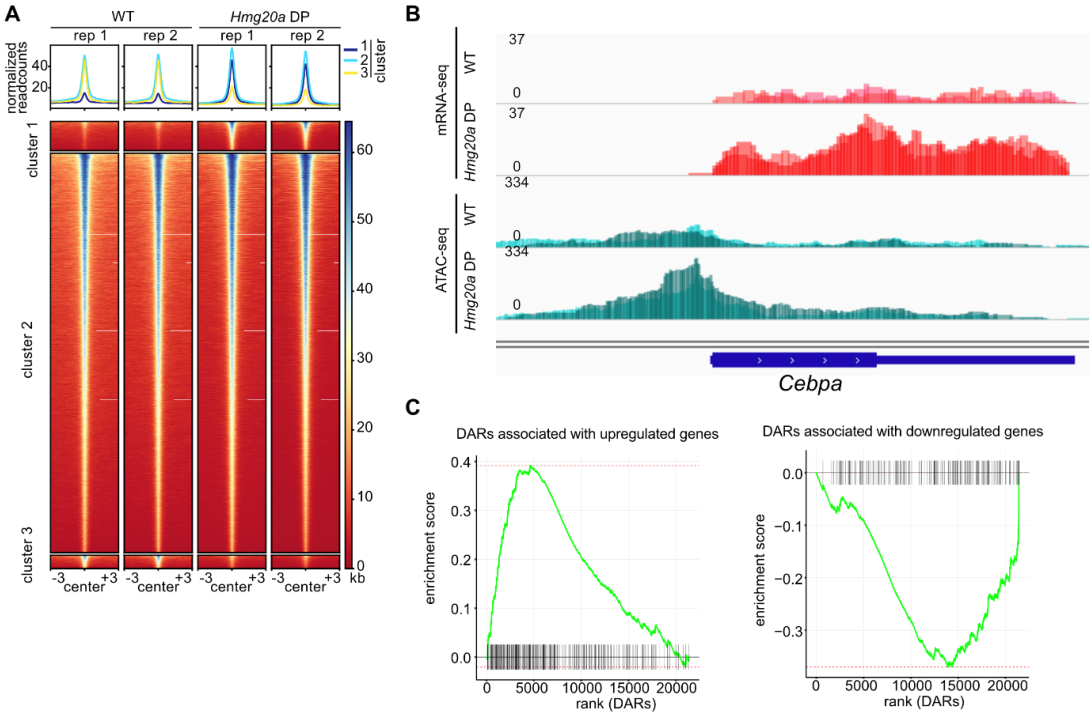


Figure 31: HMG20A regulates transcription by predominantly repressing DNA accessibility
(A) Density heatmap of ATAC-seq sites (two replicates) that become more open (cluster 1), remain unaffected (cluster 2) or become more closed (cluster 3) upon HMG20A depletion at Day2 of mESC differentiation. (B) Representative

IGV browser snapshot of two independent mRNA- and ATAC-seq signals, dark and light colors each) in Day2 mESCs in WT and *Hmg20a* DP cells. (C) Gene set enrichment plot of genes associated with differentially accessible regions after HMG20A depletion correlated to gene expression. Notice that more open accessibly sites correlate with increase in gene transcription (top; adjusted p-value = 5.502e-16; NES = 2.299), while more inaccessible sites correlate with reduction in gene expression (bottom; adjusted p-value = 1.620e-04; NES = -1.664) (calculated as in (Korotkevich et al., 2021)) DARs: differently accessible regions.

5.5.2.1.2 *Loss of HMG20A alters transcriptomic programs of cardiomyocyte differentiation*

To see whether the delay in cardiomyocyte differentiation observed before (Figure 29A) is also detectable throughout the transcriptome, it is necessary to compare similarities and/or differences between individual mRNA-seq data sets, rather than relying on the deregulated genes themselves. Principal component analysis (PCA) allowed us to compare such data sets with each other in reduced complexity. PCA showed that replicates of each sample were very similar (sometimes even overlapping at the presented resolution), while during differentiation, data sets became more and more different to those of earlier time points, revealing a stage-dependent gene expression trajectory. Interestingly, the mRNA-seq data sets of Day7.5 from *Hmg20a* DP were more similar to the mRNA-seq data sets of Day6 from WT cells, than their Day7.5 counterpart. This showed once again, that there is a delay in CM differentiation progression, even on transcriptome wide scale.

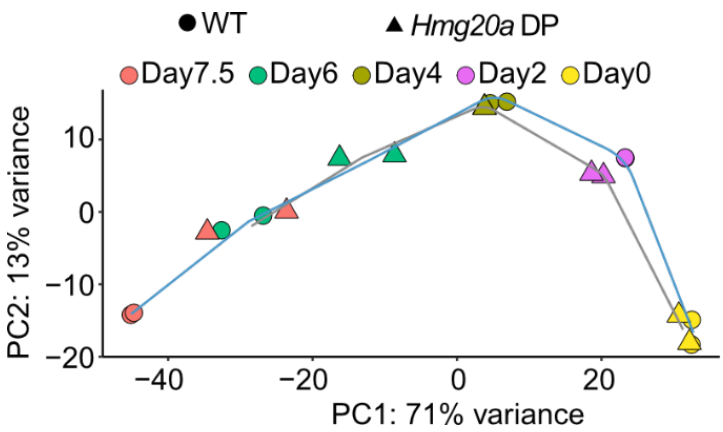


Figure 32: Delay in cardiomyocyte differentiation of *Hmg20a* DP cells is confirmed transcriptome-wide
Principal component Analysis (PCA) of RNA-seq data of two replicates of wild type (circle) and *Hmg20a* DP clone#26 (triangle) at Day0 (yellow), Day2 (magenta), Day4 (olive), Day6 (green) and Day7.5 (red) differentiation time points (see for Figure 27A for cardiomyocyte differentiation scheme).

To understand how HMG20A regulates transcriptional programs during cardiomyocyte differentiation, we analyzed how certain transcriptional programs behave over time, by calculating their z-scaled expression values in differentiation, and divided them into 10 clusters. Next, we compared the z-scaled expression values of those clusters with expression in *Hmg20a* DP cells.

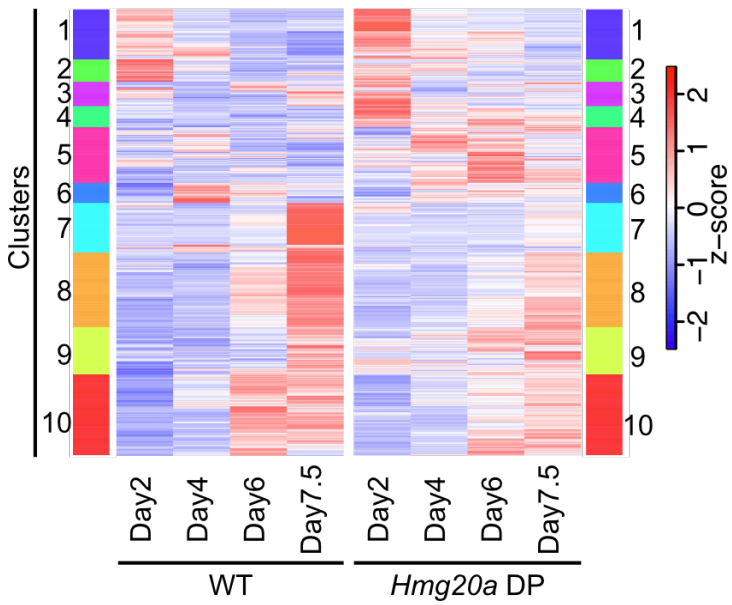


Figure 33: Transcriptional programs of cardiomyocyte differentiation are perturbed by HMG20A depletion
Heat map showing the z-scaled expression values from all significant deregulated genes comparing the differentiation steps (Day 2 vs. 4; 4 vs. 6; 6 vs. 7.5). Genes are clustered according to the Euclidean distance by an unsupervised agglomerative hierarchical approach. Shown are the mean z-scales of two replicates for each day for wild type (left panel) or *Hmg20a* DP (right panel) cells.

We observed that basically all 10 clusters showed differences in their expression behavior (Figure 33). When performing a gene ontology (GO) analysis on genes within those clusters, it becomes apparent that many different biological pathways and processes were affected (Figure 34). Notice that as expected, genes regulating heart and muscle development (Cluster 7) were induced between Day6 and Day7.5 in wild type cells, while *Hmg20a* DP cells failed to do so. Furthermore, the expression of genes related to amebodial-type cell migration, cartilage/organ formation, and skeletal development/morphogenesis (Cluster 4) peaked strongly on Day2 while being higher expressed in general, again indicating that very early in differentiation critical transcriptional programs are not properly induced.

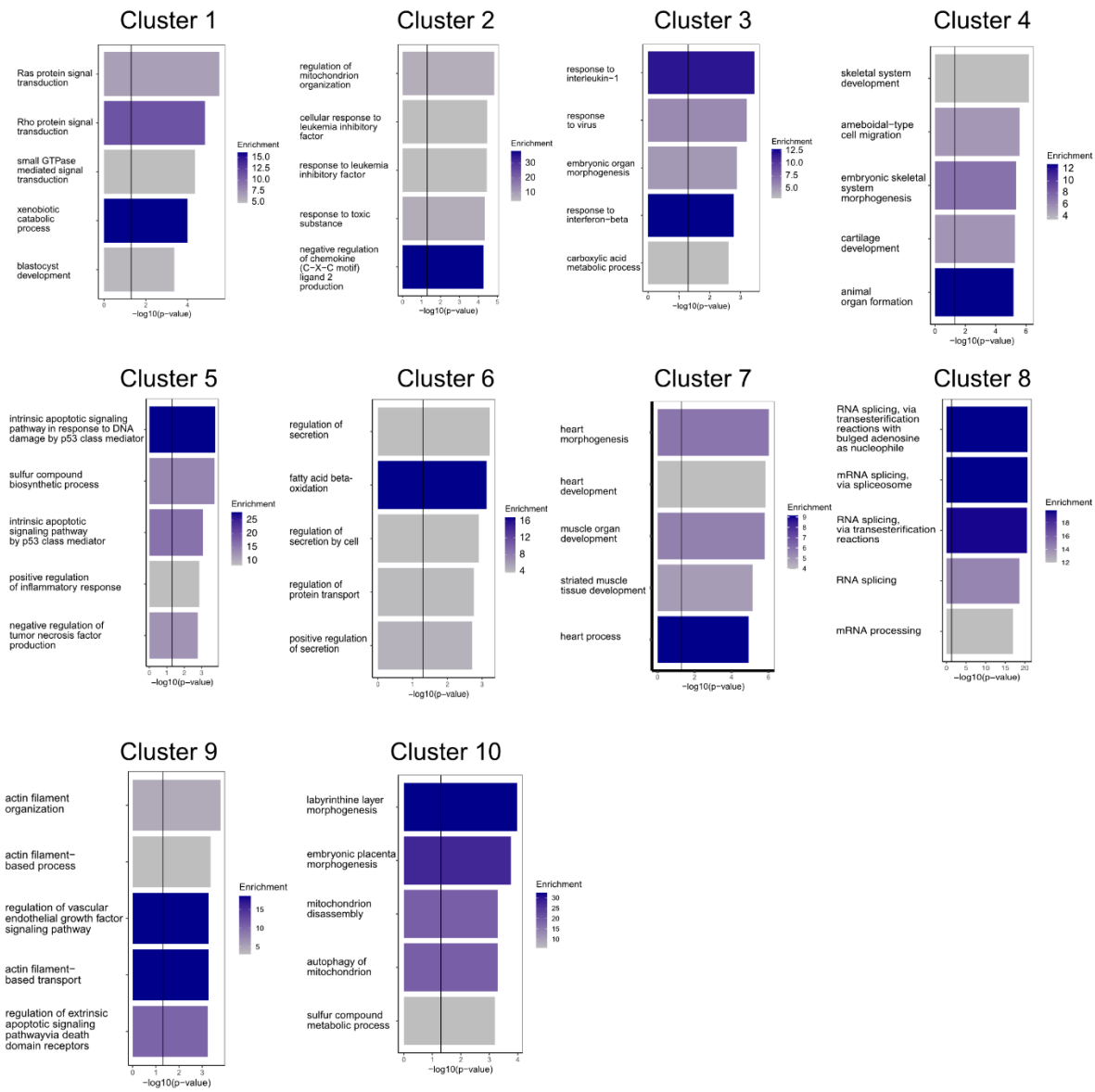


Figure 34: HMG20A regulates a multitude of transcriptional programs in cardiomyocyte differentiation
Depiction of the Top 5 gene ontology (GO) terms of Euclidian clusters defined in Figure 33) from mRNA-seq data during CM differentiation at indicated timepoint. revealed by metascape analysis (Zhou et al., 2019)

5.5.2.1.3 *Pioneer transcription factors and master regulators of differentiation are deregulated in Day2 Hmg20a DP cells*

Given the facts that differentiation of NCCs and CMs from mESCs was disturbed and that this treatment (removal of 2i and addition of fetal calf serum) was the last shared condition of these differentiation protocols

(Figure 25A and Figure 27A) while at the same time depletion of HMG20A caused deregulation of > 6 times more genes than in naive conditions (Figure 30), it is tempting to hypothesize that, at this time point, the factors determining cell fate are deregulated and therefore *Hmg20a* DP cells fail to initiate proper differentiation.

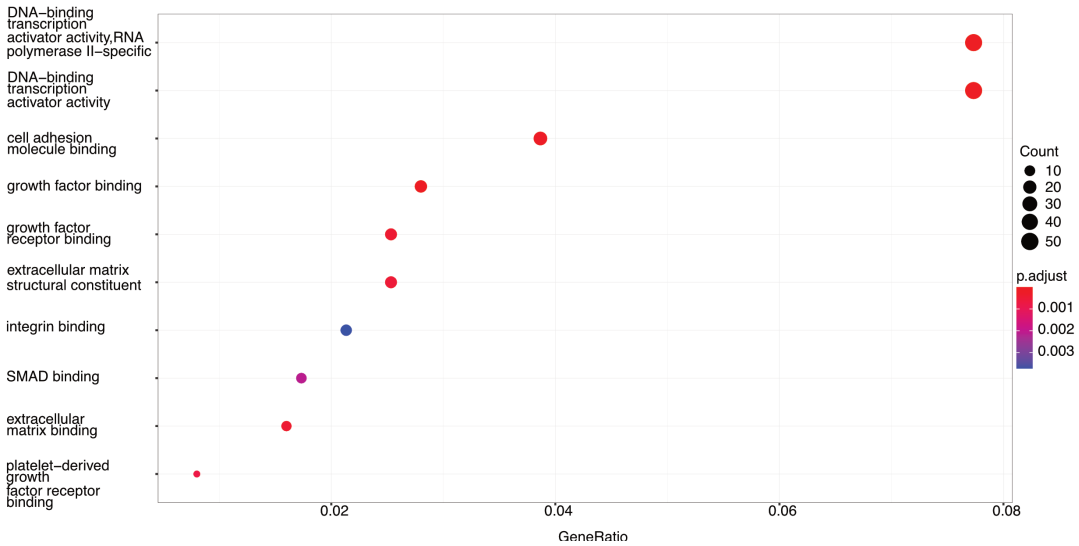


Figure 35: Transcriptional activators, cell adhesion/migration factors and SMAD binding proteins are upregulated in *Hmg20a* DP cells at Day2 of *in vitro* differentiation
Gene ontology term analysis of upregulated genes upon loss of HMG20A at Day2 of differentiation protocols (see Figure 30, Day2) Gene ontology terms GO:0001228 (DNA-binding transcription activator activity, RNA polymerase II-specific) and GO:0001216 (DNA-binding transcription activator activity) are most enriched.

In fact, developmental transcriptional activators, especially those regulating RNA Polymerase II (Figure 35), such as, among others, the highly conserved Sine Oculis Homeobox (*Six*) and Homeobox Protein (*Hox*) transcription factor families (Carnesecchi et al., 2018; Kumar, 2008; Yu et al., 2020), pioneering factors such as *Fox transcription factors*, CCAAT Enhancer Binding Protein Alpha (*Cebpa*), *Gata1/2/3*, Neuronal Differentiation 1 (*Neurod1*), were significantly up-regulated in Day2 *Hmg20a* DP over WT cells (Figure 36). All of which are reported to be involved in cell fate reprogramming (Mayran and Drouin, 2018). In

addition to that, genes involved in cell-to-cell signaling and cell migration, such as growth factor (receptors) and integrin binders, as well as components of extracellular matrixes, were up-regulated in Day2 *Hmg20a* DP cells. Furthermore, in accordance with a recent publication, we found that SMAD binding proteins (e.g. Transforming Growth Factor Beta Induced Factor Homeobox 1 like 1 (*Tgfb1i1*), Transforming Growth Factor Beta Receptor 2 and 3 (*Tgfbr2* and *Tgfbr3*), mediators of TGFB signaling) were deregulated after loss of HMG20A (Gómez-Marín et al., 2022).

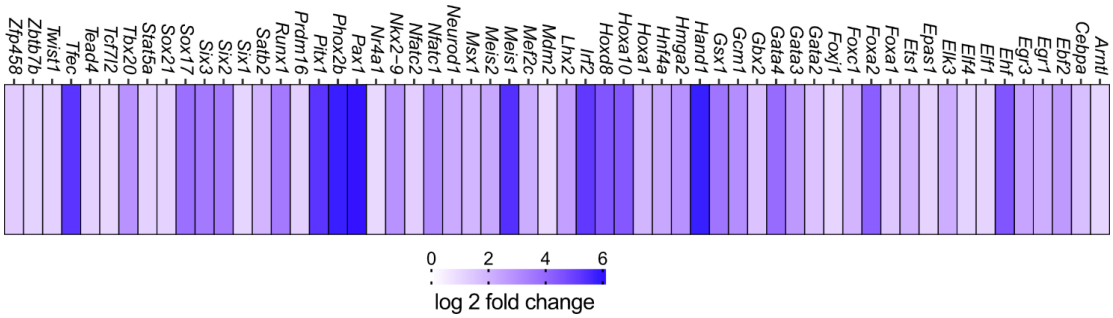


Figure 36: RNA Polymerase II activating transcription factors are significantly upregulated in Day2 *Hmg20a* DP cells
Heatmap of log2 fold changed expression of Transcriptional activators including *Six*, *Hox*, *and* pioneering transcription factors in Day2 *Hmg20a* DP cells, assessed by mRNA-seq. All depicted genes were significantly up-regulated (adjusted p-value, Wald test <0.05).

Having identified master regulators of transcriptional programs, such as pioneer factors or TGFB signaling components, to be repressed by HMG20A in Day2 cells, I wondered whether this repression is caused by HMG20A binding to those genes, and possibly tethering transcriptional repressors (as identified in Hela cells (Figure 8)), or if activation of these factors is a downstream effect of Hmg20a depletion.

5.6 HMG20A localized to promoters and enhancers in Day2 mESCs regulating chromatin organization and embryonic development.

In order to identify HMG20A target genes, it is necessary to map the genome-wide HMG20A chromatin binding. For that I performed 'cleavage under targets and release using nuclease followed by sequencing' (CUT&RUN) in Day2 cells. The resulting data was analyzed in conjunction with Tobias Friedrich (Prof. Dr. Tilman Borggrefe, Institute for Biochemistry, Justus Liebig University Giessen, Germany).

CUT&RUN is a cost-effective alternative to ChIP-seq is. Here, a MNase-Protein A-fusion protein is tethered to a protein of interest by a specific antibody in isolated nuclei. Protein A will bind to the heavy chain of the antibody, bringing the MNase in close proximity to genomic DNA. As a result of the addition of calcium chloride to the assay, MNase is activated and cleaves DNA up- and downstream of the protein of interest. The generated small DNA fragments are released from the nucleus by increasing the temperature to 37 °C and are then purified, before they are used to generate sequencing libraries. After sequencing, the information is mapped back to a reference genome to learn about genome-wide localization of the protein of interest. To control for CUT&RUN background fragmentation of chromatin I performed CUT&RUN in *Hmg20a* DP Day2 cells as well.

5.6.1 HMG20A *dampens chromatin accessibility of transcriptionally active genes*

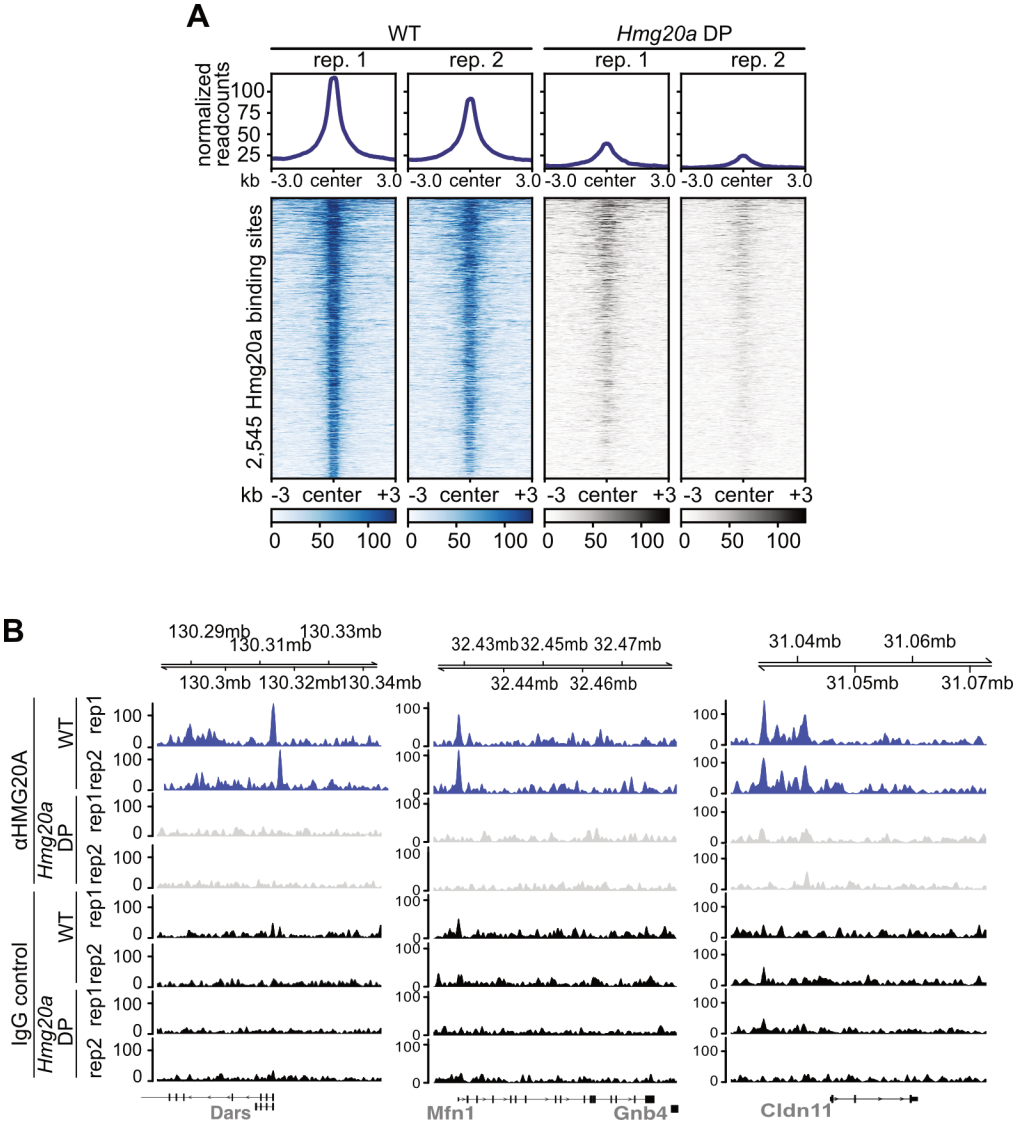


Figure 37: Hmg20a binds to chromatin in primed mouse embryonic stem cells
(A) Density heatmap of 2,545 HMG20A binding sites detected in CUT&RUN. Color intensity represents normalized and globally scaled tag counts. (B) Snapshot of the genome browser of representative Hmg20a-binding regions in Day2 mouse embryonic stem cells.

Applying CUT&RUN of HMG20A in Day2 mESCs, we identified 2,545 bona fide binding sites (Figure 37A, B) corresponding to 2,094 genes. Binding sites were characterized as specific by the presence of signal

peaks (αHMG20A antibody versus IgG) in wild type cells and their absence in the *Hmg20a* DP cell clone #26.

Similar to GFP-HMG20A in human Hela cells, endogenous HMG20A was located mainly in accessible chromatin regions (Figure 15A and Figure 38A). Interestingly, when HMG20A was depleted, those sites opened up even more (Figure 38B), indicating that HMG20A limits their accessibility to a certain extent. In line with that, we discovered a correlation between HMG20A binding intensity and expression level of a given gene (Figure 38C).

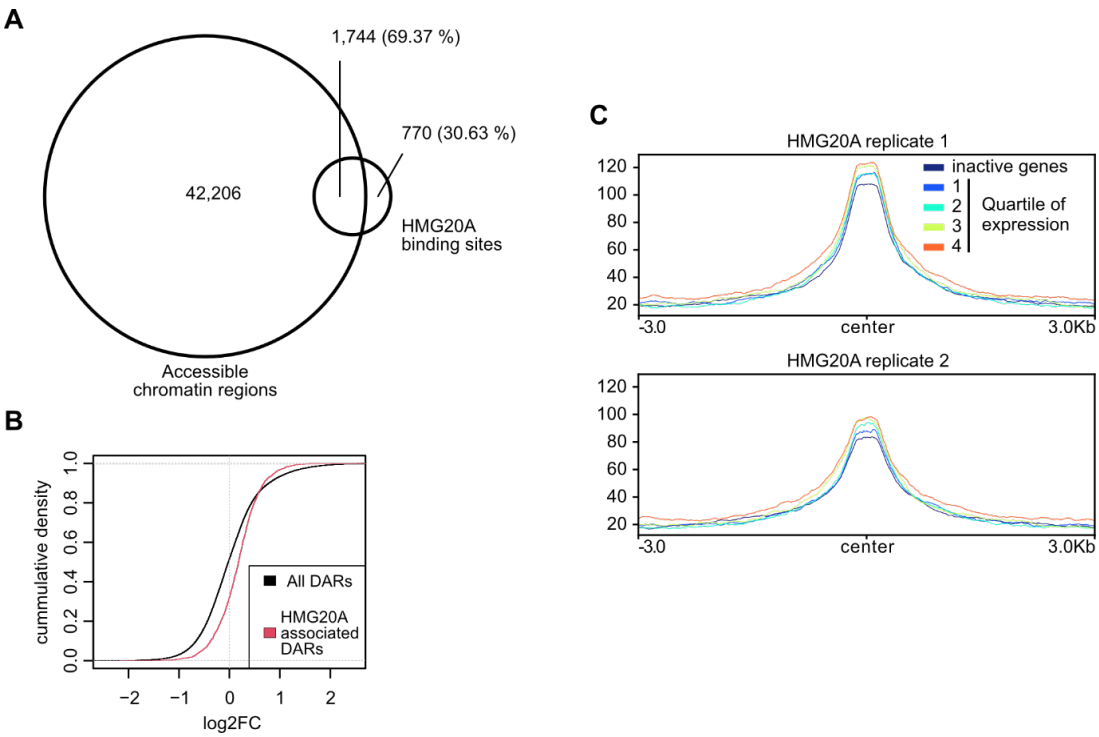


Figure 38: HMG20A regulates accessibility of actively transcribed genes in Day2 cells
(A) Venn diagram depicting overlay of CUT&RUN identified HMG20A binding sites with ATAC-seq identified accessible chromatin regions. (B) Cumulative density plot showing the distribution of the observed changes in chromatin accessibility (shown as log2FC (DP/WT)) for all ATAC-seq signals (black) and for those ATAC-seq signals overlapping with HMG20A (red). DARs: differentially accessibly regions. (C) Average binding plots of both replicates of CUT&RUN identified HMG20A binding. Line colors reflect the average binned expression levels of associated genes.

Since HMG20A seems to act as a rheostat of transcription, I wondered, which genes are directly bond by HMG20A, and are upregulated upon its depletion.

5.6.2 *Genes involved in developmental processes and cell migration are directly repressed by HMG20A*

To analyze, which target genes of HMG20A were upregulated in *Hmg20a* DP compare to WT cells, I combined CUT&RUN and mRNA-seq from Day2 cells to extract genes, that are both, HMG20A bound upregulated in *Hmg20a* DP cells, and performed GO analysis with the resulting gene list.

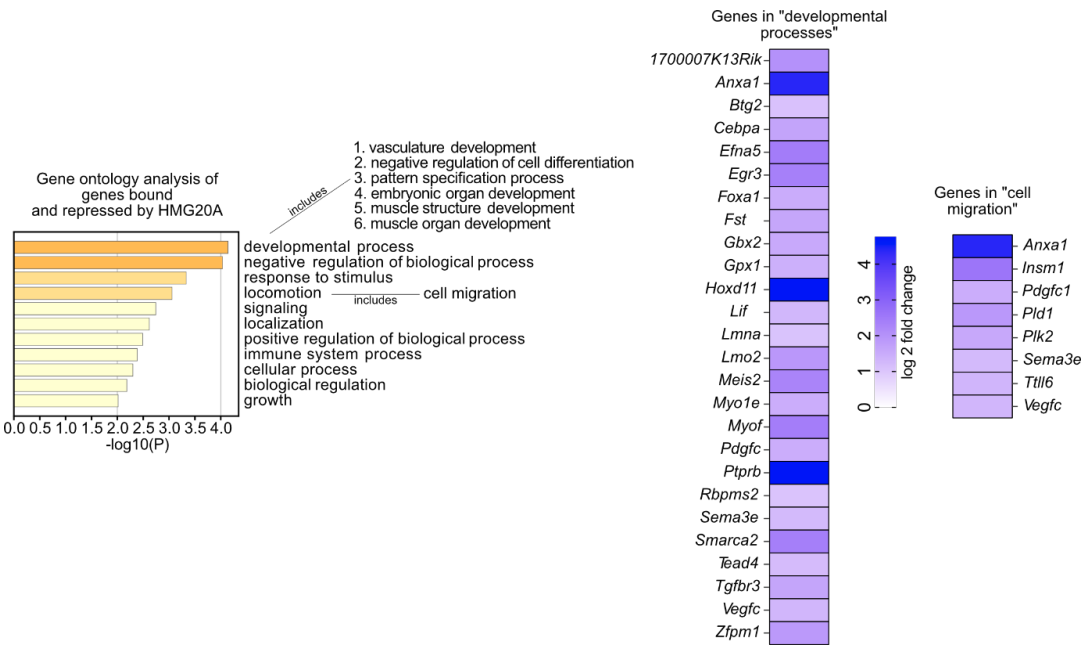


Figure 39: Genes regulating developmental processes and cell migration are directly repressed by HMG20A
Parental, and indicated daughter gene ontologies revealed in GO analysis of HMG20A bound and repressed genes identified in CUT&RUN and mRNA-seq of Day2 cells applying metascape (Zhou et al., 2019) bar color indicates hierarchy of enrichment of indicated GO term (left), right: Heatmap of log2 fold changed expression of HMG20A bound genes as identified in CUT&RUN of Day2 cells with the GO term 'developmental processes' and 'cell migration' in *Hmg20a* DP Day2 cells, assessed by mRNA-seq. All depicted genes were significantly up-regulated (adjusted p-value, Wald test <0.05).

HM20A primarily bound and repressed genes regulating 'developmental processes' (Figure 39), including some of the pioneer factors mentioned above (*Cebpa*, *Foxa1*). Note that genes that regulate 'cell migration' were also bound and repressed in Day2 cells by HMG20A, possibly affecting cell travel in early development.

5.6.3 *HMG20A colocalizes with NuRD and LSD1/BHC complex to H2A.Z occupied promotors and H2A.Z independent enhancers*

To understand how HMG20A facilitates its biological function in development, with respect to its physical interaction with H2A.Z-containing nucleosomes, Tobias Friedrich (Prof. Dr. Tilman Borggrefe, Institute for Biochemistry, Justus Liebig University Giessen, Germany) and I used publicly available ChIP-seq data of H2A.Z in mESCs. Again, similar to human Hela cells, we found in mESCs HMG20A partially overlapped with H2A.Z sites occupied by the promoter mark H3K4me3 (HMG20A+H2A.Z, cluster 1), while HMG20A binding sites that were not located at H2A.Z nucleosomes (HMG20A-only, cluster 2) were not positive for H3K4me3, but rather H3K4me1-associated (Figure 14 and Figure 40). Publicly available ChIP-seq data sets from mESCs of MTA1, CHD4, and LSD1 (KDM1A) I identified to be HMG20A interacting proteins in Hela cells (Figure 8) (Burgold et al., 2019; Luo et al., 2015; Whyte et al., 2012) were also found to bind chromatin at HMG20A binding sites, regardless of H2A.Z occupation (Figure 40), supporting these data on HMG20A protein-protein interaction.

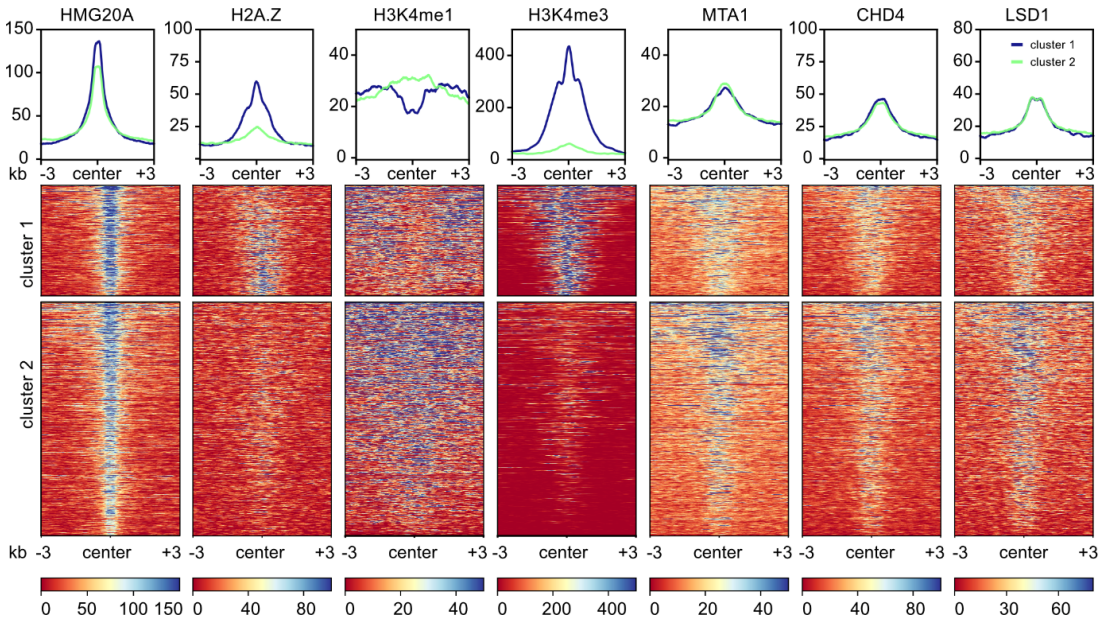


Figure 40: HMG20A colocalizes with H2A.Z at H3K4me3 and binds independent of H2A.Z at H3K4me1 positive sites
Density heatmap of HMG20A binding sites detected in CUT&RUN compared to publicly available H2A.Z, H3K4me1, H3K4me3, MTA1, CHD4 and LSD1/KDM1A chromatin immunoprecipitation sequencing data from mouse embryonic stem cells, computationally (k-means) separated into two clusters: cluster 1 (blue line top): HMG20A+H2A.Z and cluster 2 (green line top): HMG20A-only sites. Color intensity represents normalized and globally scaled tag counts.

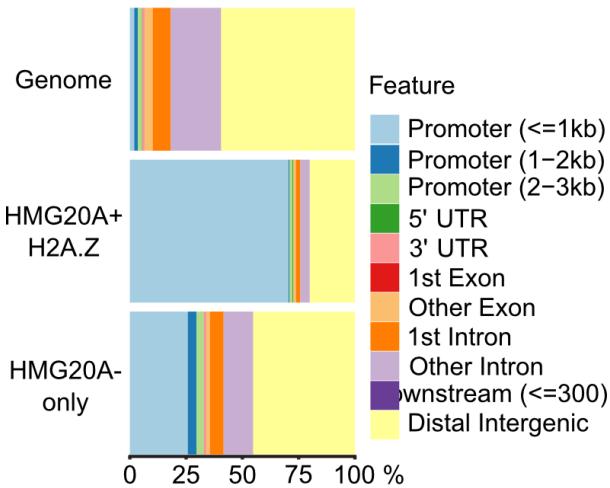


Figure 41: HMG20A binds almost exclusively to H2A.Z-occupied promotors, while only a quarter of HMG20A-only sites reside there
Feature plot depicting distribution of HMG20A+H2A.Z and HMG20A-only binding along genomic features.

In line with the co-localization of HMG20A+H2A.Z with H3K4me3 (Figure 40, cluster 1) these binding sites were highly enriched in annotated promotors. Compared to them, HMG20A-only sites (Figure 40, cluster 2) were found less often in promoters, but more in distal intergenic and intronic regions. HMG20A's chromatin binding in mESCs showed the same general behavior as observed in human Hela cells, indicating conservation of HMG20A's function from mice to humans (Figure 15C and Figure 41).

5.6.4 *HMG20A regulates either genes involved in chromatin organization or embryonic development, depending on its co-localization to H2A.Z*

Having identified sites, that are occupied by HMG20A and H2A.Z (HMG20A+H2A.Z) and sites, that are occupied by HMG20A but not by H2A.Z (HMG20A-only), I marveled, whether said binding modes are associated with different biological processes. As HMG20A depletion led to two different phenotypes (NCC differentiation and CM differentiation delay (Figure 25 and Figure 29) and HMG20A target genes regulate developmental processes as well as cell migration (Figure 39), Tobias Friedrich (Prof. Dr. Tilman Borggrefe, Institute for Biochemistry, Justus Liebig University Giessen, Germany) and I analyzed whether these phenotypes were rooted in the HMG20A+H2A.Z and HMG20A-only binding modes.

To identify genes that are directly regulated by either binding mode, we combined mRNA expression changes with HMG20A's chromatin localization at both binding site subsets HMG20A+H2A.Z and HMG20A-only.

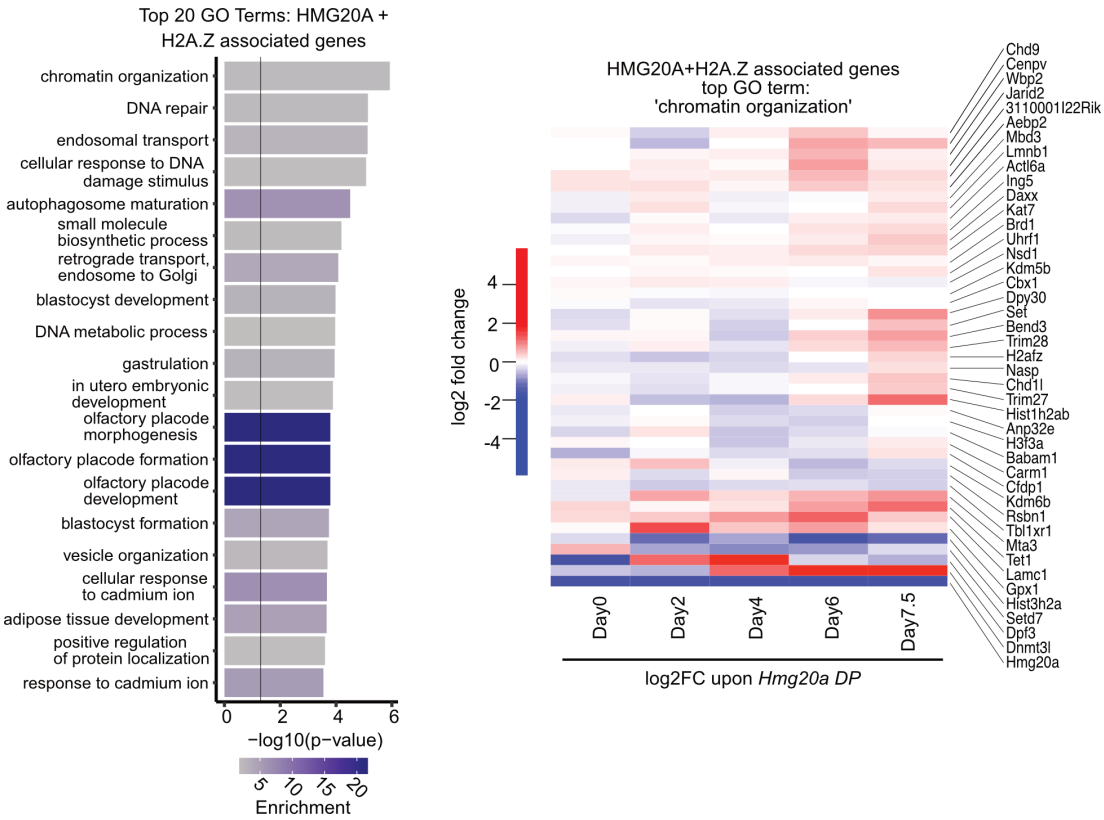


Figure 42: HMG20A and H2A.Z regulate 'chromatin organization' in cardiomyocyte differentiation
Left: depiction of the Top 20 GO terms of genes associated with HMG20A+H2A.Z binding sites (Zhou et al., 2019) Right: Heatmap of expression changes of HMG20A+H2A.Z bound genes associated with the GO term 'chromatin organization' in *Hmg20a* DP cells.

GO analysis of genes regulated by HMG20A+H2A.Z showed that they are mainly associated with DNA-based processes such as 'chromatin organization' and 'DNA repair', but also with genes associated with initial development such as blastocyst development/formation and gastrulation. Genes involved in vesicle organization and retrograde transport of endosomes to the Golgi apparatus were also identified as directly bound by HMG20A+H2A.Z (Figure 42, left). The chromatin organization genes mentioned above tended to be repressed in later stages of cardiomyocyte differentiation (starting from Day4), indicating, that *Hmg20a* DP cells actually fail to reprogram transcription and its underlying chromatin reorganization (Figure 42, right). Note that the *de novo* DNA methylation cofactor *Dnmt3l* was highly up-regulated in later

cardiomyocyte differentiation. *Dnmt3l* is highly expressed in mouse embryonic stem cells and is repressed over time in development (Neri et al., 2013). *Hmg20a* DP cells appear not to be able to repress this gene, potentially inflicting major deregulation of the DNA methylation pattern.

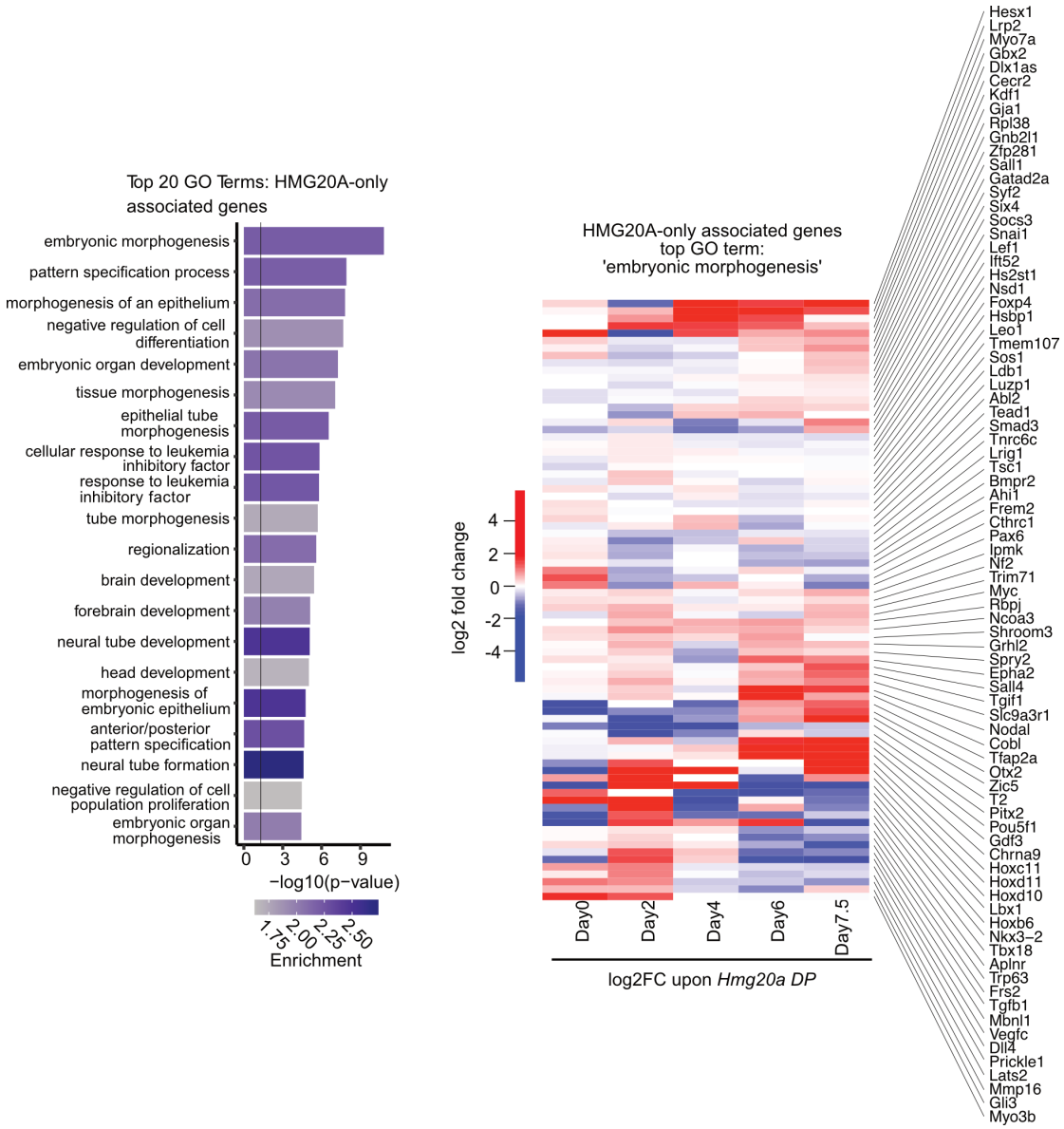


Figure 43: HMG20A-only binding regulates genes involved in embryonic development and morphogenesis
Left: depiction of the Top 20 gene ontology (GO) terms of genes associated with HMG20A-only binding sites revealed by metascape analysis (metascape.org). Heatmap of expression changes of HMG20A+H2A.Z bound genes associated with the GO term 'embryonic morphogenesis in *Hmg20a* DP cells.

HMG20A-only binding, on the other hand, was associated with genes involved mainly in development, such as embryonic morphogenesis, pattern specification, embryonic organ development, and head development (Figure 43, left). Here, HMG20A regulated those genes earlier in differentiation (starting on Day2), and the deregulation was also more severe, in positive and negative directions, in *Hmg20a* DP cells.

Taken together the data presented here strongly support that HMG20A is a master regulator of head/neural crest and heart development in vertebrates. It probably stabilizes expression of genes involved in modulating chromatin structure by tethering transcriptional repressors to transcriptionally active promoters (HMG20A+H2A.Z), and to enhancers within active genes (HMG20A-only) of genes involved in development and cell migration. This way HMG20A affects specific developmental gene expression programs, securing proper differentiation and cell fate commitment.

6 Discussion

This study sheds light on HMG20A's function as transcriptional repressor in early differential processes. Its interactome contains reported interactors BHC/CoREST and PRTH, as well as all factors of the NuRD complex, to which its binding was determined to be mediated by HMG20A's C-terminus, harboring the coiled-coil domain, while the N-terminus, containing an HMG box, binds DNA directly. HMG20A's genomic binding patters are conserved in human and mice, it preferentially binds to open chromatin regions, particularly to nucleosome depleted regions of H2A.Z occupied promotors and to intronic enhancers of actively transcribed genes. HMG20A depletion results in more open chromatin regions and conserved phenotypes in the development of the neural crest and the heart.

6.1 HMG20A binds transcriptionally repressive complexes

Usage of label-free quantitative mass spectrometry revealed the HMG20A interactome. HMG20A was identified to bind BHC/CoREST, known to be HMG20A associated (Rivero et al., 2015), the complete NuRD complex, and heterochromatic readers L3MTBL3 and BEND3 (Arai and Miyazaki, 2005; Sathyan et al., 2011) and the TEAD transcription factors. In addition, it binds to proteins that are also associated with H2A.Z/PWWP2A, namely ZNF512B, members of the PRTH complex (PHF14, RAI1 and TCF20) and the NuRD core components MTA1, HDAC, and RBBP (M1HR) (Eberl et al., 2013; Gómez-Marín et al., 2022; Käsper et al., 2021; Link et al., 2018; Pünzeler et al., 2017).

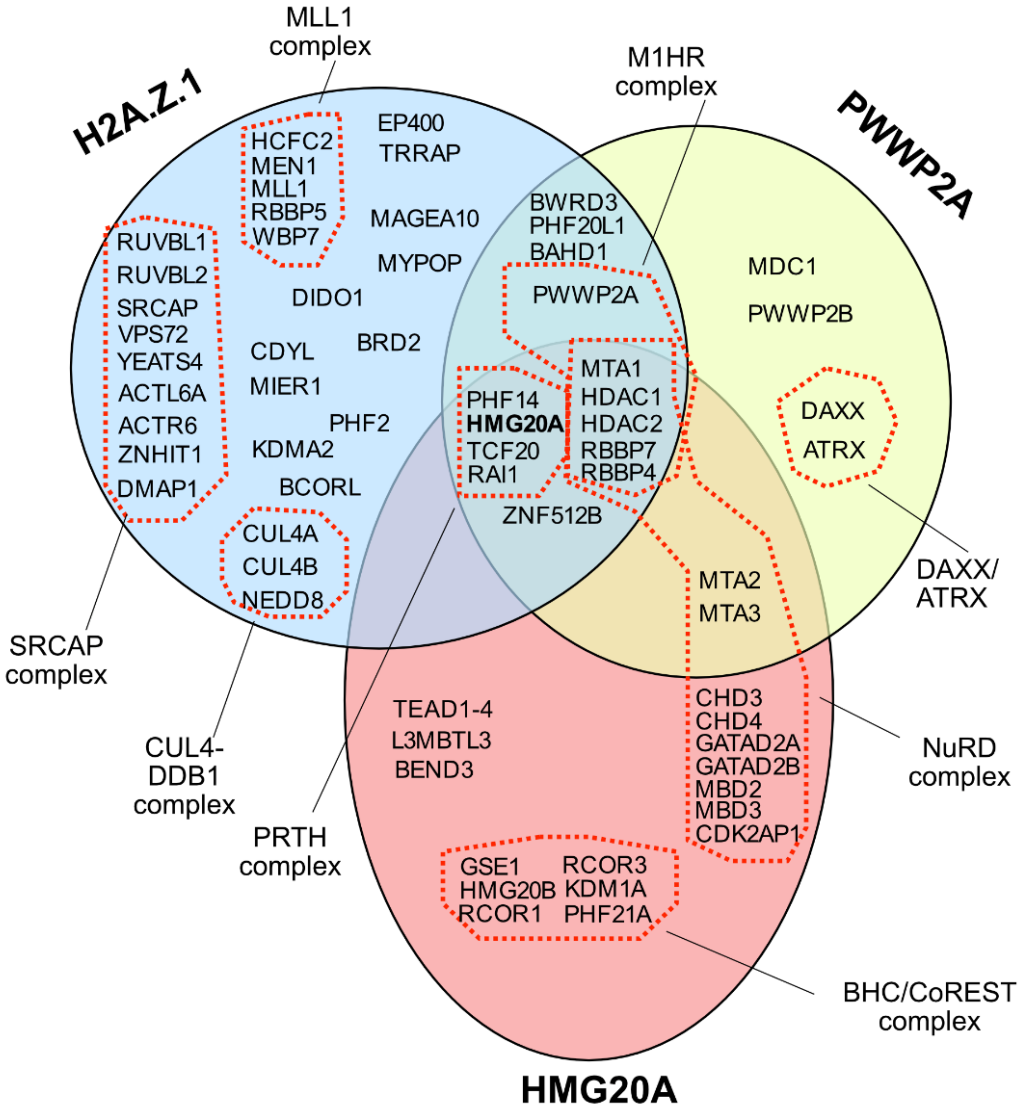


Figure 44: Summary of the interactomes of HMG20A, PWWP2A and H2A.Z
Venn diagram of protein names identified as interactors of H2A.Z.1 (blue background, (Pünzeler et al., 2017)), PWWP2A (yellow background, (Link et al., 2018), and HMG20A (red background). Characterized complexes are indicated by red dotted lines.

6.1.1 *HMG20A is part of an intricate network of interactions of components of the NuRD complex.*

To analyze HMG20A's interaction to the NuRD complex in more detail, GFP-HMG20A was co-transfected with different combinations of NuRD components. Note that these experiments were performed in HEK293T cells with endogenous NuRD expression (Reid et al., 2023; Zhong et al., 2022), making it difficult to conclude about direct interactions or exclude NuRD-independent interactions of HMG20A with single NuRD components. Furthermore, the expression levels of the transfected constructs can vary from one assay to another; even when performed in parallel. Together with semiquantitative Immunoblot readouts, conclusions regarding binding affinities are estimations rather than measurements.

Nevertheless, HMG20A's coiled-coil domain containing C-terminal part seems to bind MTA1 but not MTA2 (Figure 9, Figure 10). HMG20A does not interact with RPPB4, but if it is co-transfected with HDAC1, it inhibits HMG20A and HDAC1 binding, indicating a competition between HMG20A and RBBP4 to interact with HDAC1 (Figure 9). However, differences in HDAC1 co-precipitation can be caused by different expression levels in each assay, which are difficult to precisely monitor with Immunoblot readouts (see HDAC1 signal in input, Figure 9).

Unlike for MTA1 and MTA2, HMG20A binds to all tested paralogues of the NuRD remodeling cassette independent of other co-transfected complex members (Figure 11). Surprisingly, HMG20A appears to interact with the middle part of CHD4 (CHD-M) that contains the translocase domain (Figure 11). To clarify, whether HMG20A indeed binds to CHD-M

directly protein-protein binding assays without background of endogenous NuRD components need to be conducted. Functionally, binding of HMG20A to CHD-M could potentially have implications on CHD4 activity, by either preventing it to act on its designated substrate, as this CHD4 construct is capable to remodel recombinant mononucleosomes or by promoting its action by influencing binding of CHD4's C1a autoinhibitory domain (Zhong et al., 2022). In case HMG20A binds to the CDs or PHD fingers, presents in CHD-M, it might influence CHD4's histone or DNA binding near the nucleosome dyad, the DNA entry and exit point of the nucleosome (Nodelman et al., 2017; Schindler et al., 1993; Sims et al., 2005).

In summary, HMG20A binds to/associates with NuRD, potentially in a multivalent manner, possibly with a preference for the M1HR sub-complex via MTA1. The data presented here are consistent with previous reports suggesting that the H2A.Z target regions are regulated by NuRD subcomplex M1HR (Link et al., 2018; Low et al., 2020; Zhang et al., 2018).

Since HMG20A is bound by chromatin associated interactors, it is reasonable to assume that the HMG box of HMG20A enhances the DNA binding of the complex as its binding DNA. Although HMG20A has been reported to inhibit the function of BHC/CoREST (Rivero et al., 2015), the influence of HMG20A on other complexes identified in this study is still elusive.

6.1.2 *DNA sequence-specific binding of* **HMG20A**[Office1]

In general, there are two types of HMG box proteins described. (Štros et al., 2007). One, with that binds to specific DNA motifs, while the other binds DNA in an unspecific manner. In line with Gómez-Marin and colleagues, who report unspecific DNA binding of HMG20A (Gómez-

Marín et al., 2022), this study showed binding of HMG20A to DNA of random sequence (Figure 12). While Gómez-Marin and Käsper showed increased binding of HMG20A to four-way junction (4WJ) DNA (Gómez-Marín et al., 2022; Käsper et al., 2021), motif analysis of HMG20A chromatin binding assays in this study revealed enrichment for DNA motifs similar to ONECUT3 and FOXE3 DNA binding motif AAANAAANAAANAAA, in HMG20A binding sites (Figure 18). Interestingly, this motif is quite similar to the binding motif AGAACAAAGAA of the *Schizosaccharomyces pombe* HMG box containing Transcription factor ste11 (Ste11) (Beest et al., 2000) (Figure 18).

To exam if HMG20A does, in addition to general DNA binding, confer specific binding to this motif, Jörg Leers (Staff scientist at the Institute for Genetics, Justus Liebig University Giessen) and I decided to test whether DNA of GGAAANAAANAAANAAAGG (GG is added 5'-end and 3'-end to improve proper hybridization) sequence will bind to FLAG-HMG20A or FLAG-HMG over a DNA sequence converted to TTCCTCCCTCCCCTT. Therefore, we purified FLAG-HMG20A and FLAG-HMG from Sf9 extracts using a heparin column (HiTRAP Heparin HP, Cytiva) (Figure 45A), pooled fraction 7+8 and 8 + 9, respectively, and performed competitive EMSA on both DNA sequences mentioned above. We determined that FLAG-HMG binds stronger to GGAAANAAANAAANAAAGG than to TTCCTCCCTCCCCCCTT while no binding of FLAG-HMG20A was detected. Again, the presence of HMG20A's C-terminus seemed to inhibit its DNA binding.

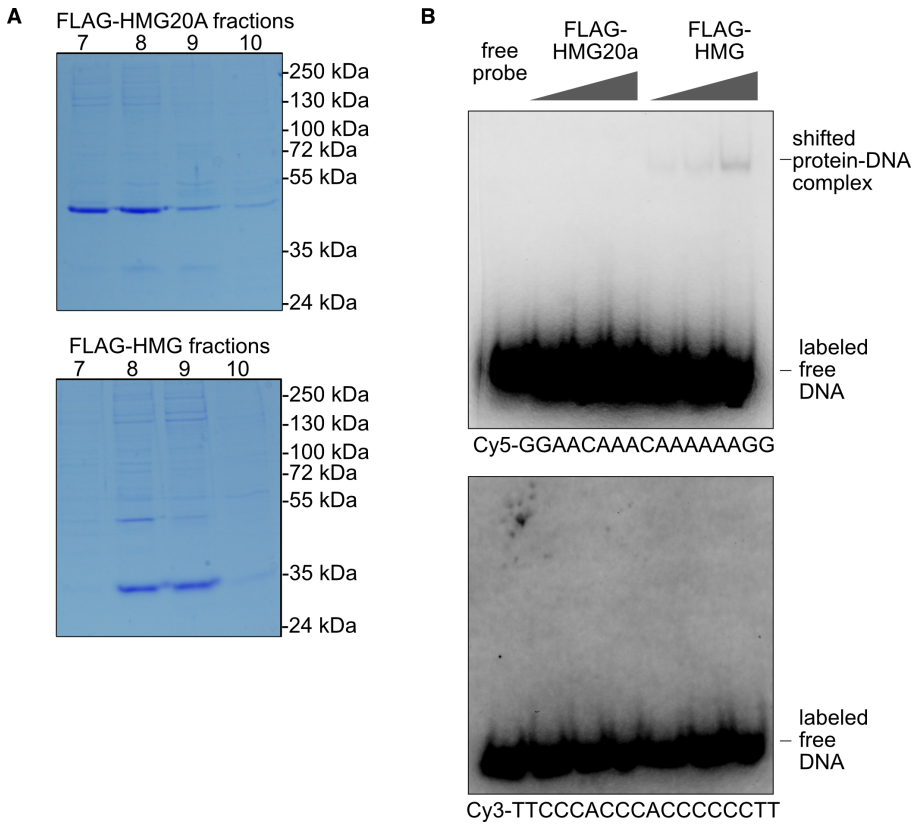


Figure 45: HMG20A preferentially binds to the ONECUT3/FOXE3 motif over its inverted counterpart
(A) Coomassie staining of the heparin column purified FLAG-HMG20A protein (top), FLAG-HMG protein (bottom) from the Sf9 cell extract. (B) Competitive EMSA of FLAG-HMG20A and FLAG-HMG protein on DNA labeled with Cy5-GGAAANAAANAAANAAAGG (top) and Cy3-TTCCTCCCTCCCCTT (bottom).

This data indicates that there is at least small amount of sequence specificity in HMG20A DNA binding, together with the DNA binding data from others, I propose, that HMG20A can bind specific (AT-rich) sequences, as well as 4WJ DNA (Gómez-Marín et al., 2022; Käsper et al., 2021).

Combining data on the protein-protein interaction of HMG20A (Figure 7), its ability to bind DNA (Figure 12, Figure 45), and the co-localization with MTA1 and CHD4 on chromatin (Figure 40) (Burgold et al., 2019; Luo et al., 2015; Whyte et al., 2012). I propose a model where HMG20A interacts with the NuRD complex, potentially in a multivalent way, influencing its remodeling activity, and increasing its DNA binding

ability. Since PWW2A occupied H2A.Z nucleosomes are more enriched for the M1HR NuRD subunit than the complete NuRD complex one can assume, that MTA1 is the dominant protein interaction partner of HMG20A. To test this hypothesis, I performed chromatin immunoprecipitation experiments in HMG20A DP mESCs. I was unable to precipitate any DNA applying antibodies against MTA1 or CHD4 after shearing chromatin from cross-linked mESCs in WT and *Hmg20a* DP conditions (data not shown). Before these experiments can be conducted, ChIP protocols for NuRD proteins must be carefully adapted.

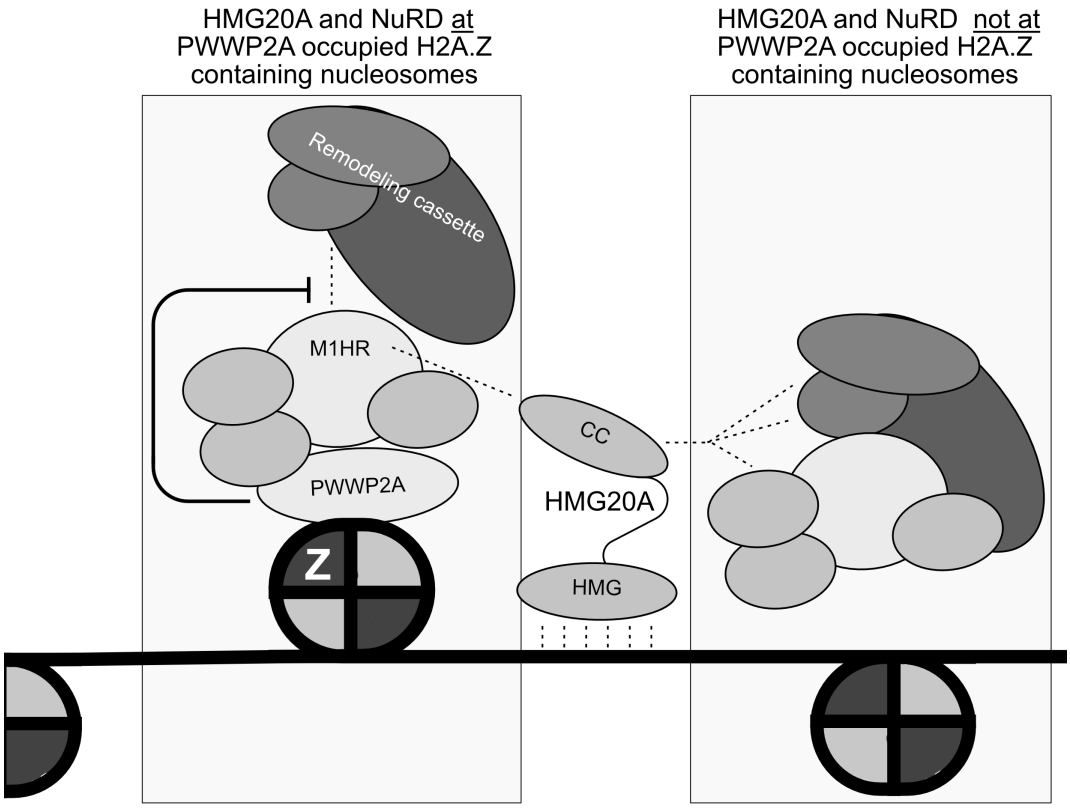


Figure 34: HMG20A might bind to the NuRD complex in a multivalent manner.
Proposed model, of how HMG20A might interact with NuRD complexes. CC: coiled-coil domain, HMG: HMG box, Z: H2A.Z

6.2 HMG20A binds to open, regulatory regions of the genome, limiting their expression

Chromatin binding studies of HMG20A in human and murine cells reported conserved binding to accessible regulatory regions (Figure 15, Figure 38A, Figure 40, Figure 41). Consistent with the data on physical interaction with NuRD and BHC/CoREST in human cells (Figure 7), mESC ChIP-seq data from Burgold et al., Lou et al. and Whyte et al. showed NuRD and BHC/CoREST to localize the the same genomic sites as HMG20A (Figure 40) (Burgold et al., 2019; Luo et al., 2015; Whyte et al., 2012). In line with that, HMG20A depletion caused an increase in accessibility there (Figure 31A, Figure 38B). In consequence, HMG20A loss leads to more up- than downregulated genes (Figure 21). Interestingly, HMG20A binding intensity correlates with the expression level of a given target gene (Figure 38), implying a rheostat-like function. A similar function was previously described for HMG20A interactors NuRD, RCOR1 of the BHC/CoREST complex and PWWP2A (Bornelöv et al., 2018; Link et al., 2018; Rivera et al., 2022; Zhang et al., 2018).

Since HMG20A depletion caused general derepression of genes, its overexpression should repress genes more than normal. In fact, GFP-HMG20A overexpression in Hela cells leads to a reduction of endogenous HMG20A protein (Figure 7C). ChIP-seq and CUT&RUN data showed that HMG20A binds close to its own promoter in Hela (Figure 46A) and mESCs (Figure 46B), presumably promoting binding of repressive complexes to it and inhibiting its own expression in this way, ensuring stable HMG20A protein levels in the cell.

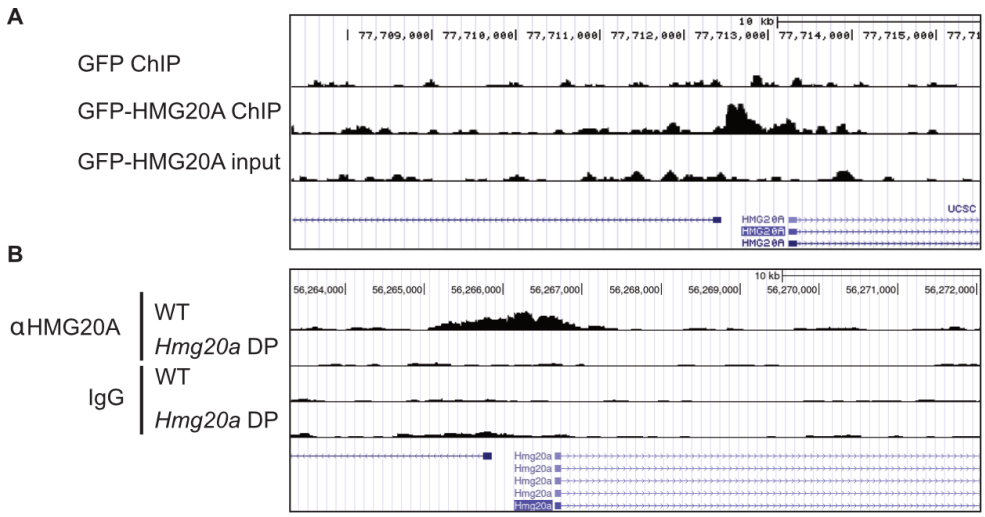


Figure 46: HMG20A binds to its own promoter in mouse embryonic stem cells and human Hela cells, potentially enforcing a negative feedback loop
USCS genome browser snapshot of HMG20A promoter regions in GFP, and GFP-HMG20A ChIP-seq (A), and mESC HMG20A CUT&RUN (B) experiments.

6.3 Do small amounts of HMG20A enforce its function?

There could be two reasons why HMG20A depletion in Hela cells hardly influences transcription (Figure 20), while in mESCs the effects were drastic (Figure 30). The first reason could be that HMG20A's regulates genes involved in 'developmental process', 'cell migration', 'chromatin organization' and 'embryonic development' (Figure 39, Figure 42, Figure 43) and in differentiated cells its function is either not needed, or redundant and compensated by other factors.

Second, only a small portion of HMG20A protein present in the cell could be required for regulation. After RNAi in Hela there was still protein detectable with Immunoblot (Figure 19E), therefore there might still be enough residual protein to actually act on chromatin and HMG20A-regulated genes are hardly affected. Keeping that in mind, residual *Hmg20a* mRNA in mESCs correlated with the severity of monitored phenotypes and marker gene expression (Figure 24B, Figure 25C, Figure 28B, Figure 29).

6.4 Loss of hmg20a skews neural crest differentiation towards melanocytes and perturbs cardiomyocyte differentiation

In general, HMG20A depletion promoted the expression of transcriptional activators in Day2 mESCs, including pioneer transcription factors, which could have major implications on a multitude of transcriptional downstream effects, making it difficult to speculate how HMG20A specifically perturbs neural crest differentiation or cardiomyocyte differentiation (Figure 39).

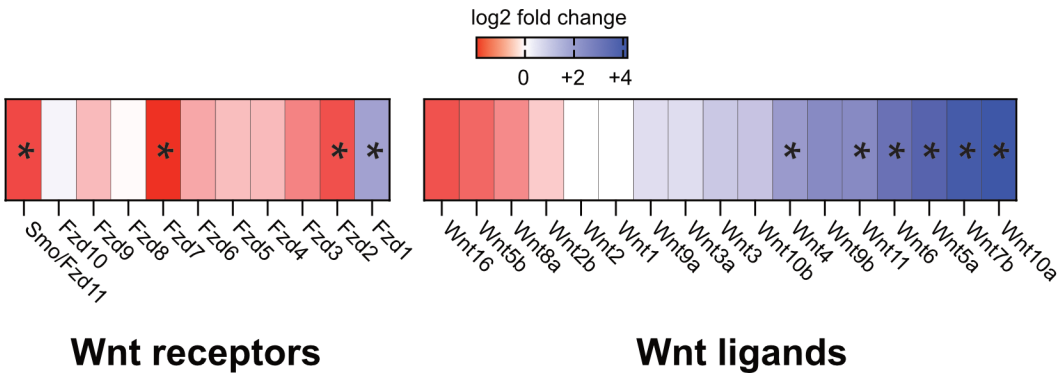


Figure 47: Wnt signaling pathway is deregulated in *Hmg20a* Day2 cells
Heatmap of log2 fold changes in gene expression of Wingless-related integration site (Wnt) ligands and receptors in *Hmg20a* DP Day2 cells compared to WT cells. Asterix indicate adjusted p-value <0.05 (Wald test).

The only transcriptome-wide data related to the applied neural crest differentiation protocol is the mRNA-seq data set of Day2 mESCs. These cells are believed to resemble cells that are in transition from inner cell mass to epiblast (Wang et al., 2021b), while neural plate border cells, from which NCCs emerge, are determined at the beginning of neurulation. However, a recent study showed that epigenetic variability and intensity of growth factor and Wingless-related integration site (Wnt) signaling, as observed for *Hmg20a* DP Day2 cells (Figure 35 and

Figure 47), can dramatically determine the potential of stem cells to differentiate (Ortmann et al., 2020).

It should be noted that two biological processes essential for neural crest development and migration were identified to be related with HMG20A in genome- and transcriptome-wide studies of Day2 mESCs. First, promoters of genes responsible for 'olfactory placode formation', 'vesicle organization' and 'endosome transport' are all found to be bound by HMG20A+H2A.Z (Figure 43, left). Note that cranial facial NCCs, which are responsible for cartilage and bone formation, mainly in the future face (Santagati and Rijli, 2003), are "chasing" the migrating olfactory placode cells to find their final destination for terminal differentiation and that cell-to-cell communication via vesicles promotes directional NCC migration (Daniele et al., 2022; Gustafson et al., 2022; Kulesa and Fraser, 2000; McKinney et al., 2011; Scarpa and Mayor, 2016). Second, genes involved in 'cell migration' were generally up-regulated in Day2 *Hmg20a* DP cells (Figure 39, left). Genes involved in orchestrating 'Ras and Rho GTPase cascades', which trigger EMT and NCC delamination and are the drivers of their migration, were repressed by HMG20A in Day2 cells specifically (Cluster 1 in Figure 33 and Figure 34) (Casado-Medrano et al., 2019; Scarpa and Mayor, 2016; Tripathi and Garg, 2018), again implicating that cell migration and, in particular, NCC migration could be disturbed and/or misguided at the end of neurula. This could then result in a shift in the distribution of their successor cell types.

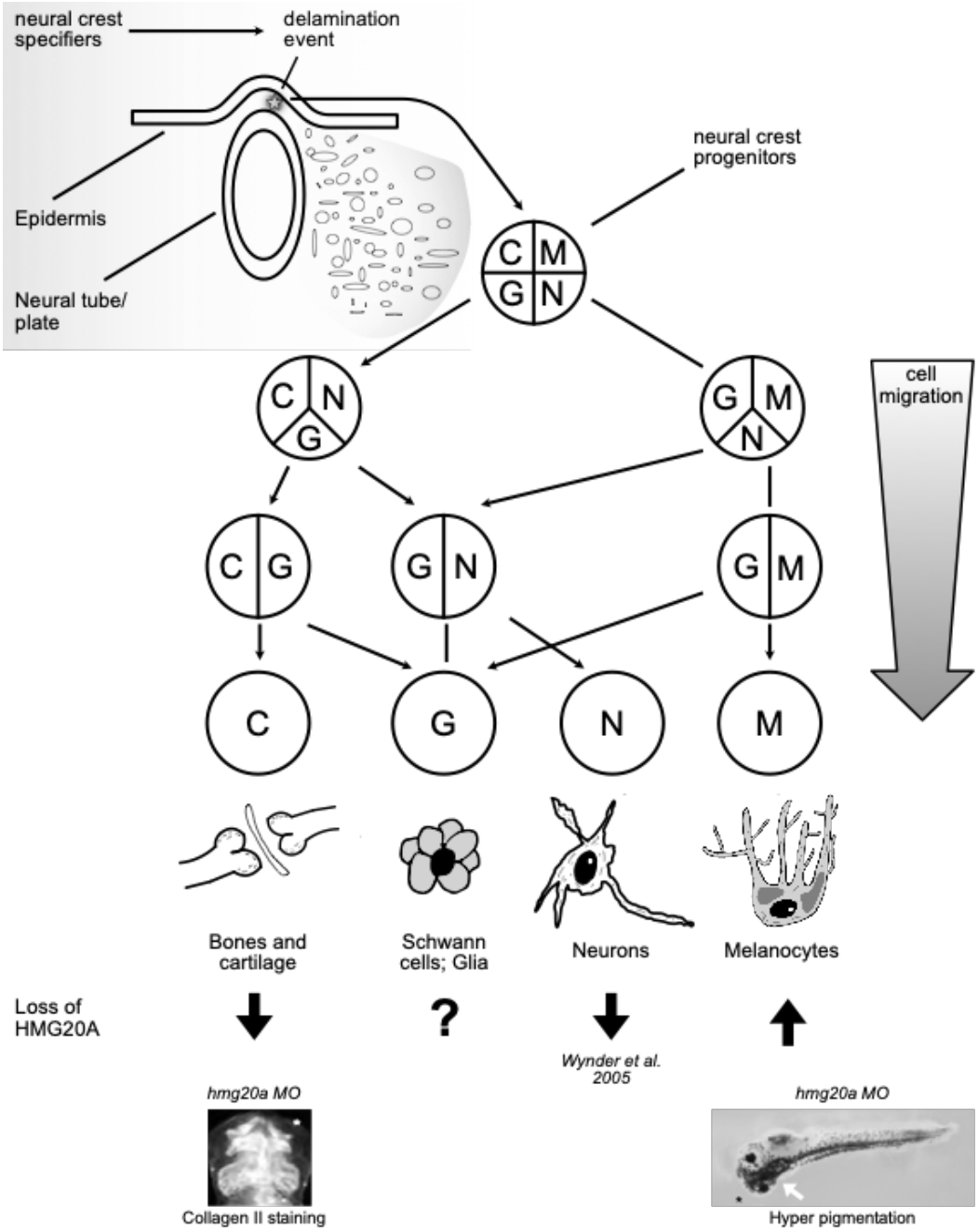


Figure 48: Cell fate of neural crest cells is skewed in HMG20A loss of function scenarios *in vivo* and *in vitro*.
Schematic depiction of neural crest delamination and the neural crest cell fate map. The committed precursors of cartilage/bone (C), glia (G), neurons (N), and melanocytes (M) are derived from intermediate progenitor cells. Direction of arrows implicate increased or decreased formation in HMG20A loss of function scenarios. Adapted from (Martinez-Morales et al., 2007).

Data from murine, pluripotent P19 cells (Wynder et al., 2005) and *Xenopus laevis* embryos (Figure App 3) show reduction in cartilage, and neural structures, but an increase in melanocytes (Figure 48) upon

HMG20A depletion. The HMG20A interacting complexes BHC/CoREST (Ceballos-Chávez et al., 2012; Hirota et al., 2019; Qureshi et al., 2010) and NuRD (Hirota et al., 2019; Laugesen and Helin, 2014) are well-known factors of neural stem cell fate and initiation of neural development, a process that is closely related to neural crest differentiation. At the same, time the BHC/CoREST catalytic subunit KDM1A was reported to be involved in heart development (Nicholson et al., 2011; Nicholson et al., 2013) and the NuRD complex was reported to be essential for cardiac sarcomere formation, the most prominent segment of cardiomyocytes (Wilczewski et al., 2018). Since BHC/CoREST and NuRD are ubiquitous chromatin regulating proteins, it is tempting to speculate, that HMG20A is required to convey they function in neural, neural crest and cardiomyocyte differentiation specifically.

In addition to the interacting repressive complexes, label-free quantitative mass spectrometry reported interaction of HMG20A to TEAD1 a cardiomyocyte marker gene (Akerberg et al., 2019), also shown by others (Gómez-Marín et al., 2022). Yamamoto and colleagues showed, that HMG20A binds to Ca2+/S100A6, a protein that contributes to cellular calcium signaling (Yamamoto et al., 2021). Besides possible developmental defects discussed above, the absence of HMG20A might alter calcium signaling and may abolish or delay beating of cardiomyocytes this way.

6.5 HMG20A and its associated complexes regulate transcription programs during differentiation – a model

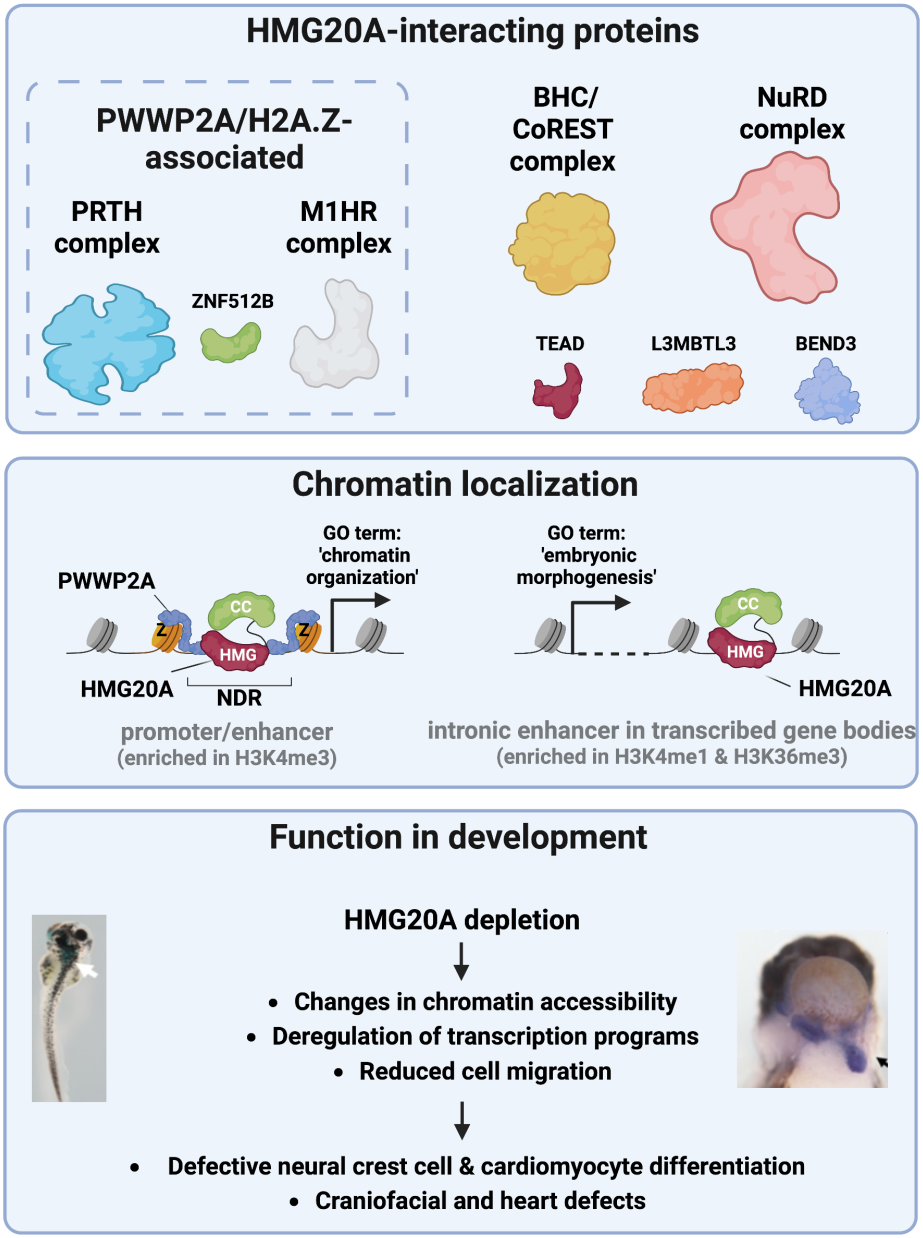


Figure 49: Model of HMG20A's function in chromatin and transcriptional regulation in development
Figure Top: HMG20A associates with H2A.Z- and PWWP2A-associated PRTH and M1HR complexes and ZNF512B as well as BHC/CoREST, NuRD, TEAD, L3MBTL3 and BEND3, that are not part of H2A.Z or PWWP2A interactomes. Middle: HMG20A binds to two distinct chromatin regulatory elements: (1) Nucleosome depleted regions (NDR) at promoter sites that are surrounded by H2A.Z-containing nucleosomes and bound by PWWP2A and that are

associated with genes involved in basic processes, such as ‘chromatin organization‘. (2) H2A.Z-lacking intronic enhancers within transcribed genes belonging to developmental processes, such as ‘embryonic morphology”. Bottom: Depletion of HMG20A in *Xenopus laevis* and mESCs leads to changes in chromatin accessibility, deregulation of transcription programs as well as migration defects. HMG20A depleted cells fail to properly differentiate into neural crest cells or cardiomyocytes in mESCs as well as head and heart in *Xenopus laevis*. Figure was created with BioRender.

In conclusion, the data presented here allow me to postulate a bimodal function for HMG20A (Figure 50). It interacts with the H2A.Z and PWWP2A-associated PRTH complex, ZNF512B and the NuRD subcomplex M1HR, while it also binds to the BHC/CoREST complex, the complete NuRD complex, TEAD proteins, L3MBTL3 and BEND3, interactions that have not been monitored with PWWP2A or H2A.Z. HMG20As chromatin binding is also dual: 1. at H2A.Z and PWWP2A occupied promoters, it binds to genes involved in ‘chromatin organization’ and 2. at intronic enhancers of genes involved in general ‘embryonic morphogenesis’, where it presumably helps to attach repressive complexes. Its depletion causes changes in chromatin accessibility, deregulation of transcriptional programs, and reduced cell migration, resulting in defective NCC and CM differentiation.

6.6 Outlook

There are still major open questions that need to be addressed to better understand HMG20A's function in development. Mechanistically, it is still unclear how exactly HMG20A binds to the NuRD complex, how it influences its activity and whether it competes with other members, as already shown for HMG20B in the BHC complex (Rivero et al., 2015). It is still unclear whether HMG20A achieves specificity for H2A.Z-containing nucleosomes? Since not all HMG20A proteins are associated with H2A.Z, specificity is likely not intrinsic to HMG20A. To assess whether binding in H2A.Z occupied regions is dependent on HMG20A interacting factors, chromatin binding assays in individual knock outs of the interactors are needed to clarify this behavior. At the same time, experiments on the effects of HMG20A depletion on chromatin binding of its interacting protein partners must be conducted. It remains elusive whether PRTH, as a protein complex, actually has a defined function in development. Studies in *Xenopus laevis* showed that RAI1 is also essential for neural crest migration, but data on its influence on gene expression and chromatin dynamics is still lacking. Phenotypically, it should be assessed in more detail how HMG20A depletion affects NCC differentiation. For example: Are their precursor cells, called neural plate border cells, defined differently or is their delamination disturbed by HMG20A depletion.

In summary, further research will to shed light on how exactly HMG20A binds its protein interaction partners and how it affects their chromatin binding. Detailed analysis will reveal, how HMG20A regulates cell fate decision in late gastrula and neurula, and how HMG20A is functionally connected to PHF14, TCF20 and RAI1 in these processes.

7 Acknowledgements

I sincerely thank Prof. Dr. Sandra Hake for giving me the opportunity to work in her laboratory and giving me constant support. Her open mind and passion for biology, together with her compassionate leadership, create a sublime environment to do science. I would like to thank Dr. Jörg Leers for being a tremendous bench mate and mentor, his wit and stunning way to organize bench space is unprecedented. I want to thank Dr. Jie Lan. His expertise and bursting with ideas, kickstarted the stem cell work in the Hake Laboratory and brought this project to the next level. I thank Prof. Dr. Marek Bartkuhn and Tobias Friedrich for their constant exchange with me throughout the years and the amount of great work and time they have put into this project. Their analytical thinking and suggestions on which questions should be asked to reveal the convoluted biology of HMG20A pushed this project forward. Thanks to Prof. Dr. Joel Mackay and Dr. Hakimeh Moghaddas Sani for helping with NuRD interaction studies, to Prof. Dr. Matthias Mann and Dr. Alexander Reim for quantifying the HMG20A interactome, and especially to Prof. Dr. Annette Borchers and Stefanie Gossen for all the work they did to prove that HMG20A is functionally relevant in *Xenopus laevis* development. Their effort determined the direction of this project.

Thanks to Dr. Oliver Rossbach for carefully reading the manuscript of this dissertation, his comments, and for helping me to attain consistency in format and phrasing.

Finally, I would like to thank my family. Their continual support throughout my life is the primary reason I was able to choose this career path.

8 References

Aagaard, L., Laible, G., Selenko, P., Schmid, M., Dorn, R., Schotta, G., Kuhfittig, S., Wolf, A., Lebersorger, A., Singh, P. B., et al. (1999). Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. *Embo J* 18, 1923–1938.

Akerberg, B. N., Gu, F., VanDusen, N. J., Zhang, X., Dong, R., Li, K., Zhang, B., Zhou, B., Sethi, I., Ma, Q., et al. (2019). A reference map of murine cardiac transcription factor chromatin occupancy identifies dynamic and conserved enhancers. *Nat Commun* 10, 4907.

Alatwi, H. E. and Downs, J. A. (2015). Removal of H2A.Z by INO80 promotes homologous recombination. *Embo Rep* 16, 986–994.

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 National Acad Sci* 51, 786–794.

Arai, S. and Miyazaki, T. (2005). Impaired maturation of myeloid progenitors in mice lacking novel Polycomb group protein MBT-1. *Embo J* 24, 1863–1873.

Awad, S. and Hassan, A. H. (2008). The Swi2/Snf2 Bromodomain Is Important for the Full Binding and Remodeling Activity of the SWI/SNF Complex on H3- and H4-acetylated Nucleosomes. *Ann Ny Acad Sci* 1138, 366–375.

Babbs, C., Lloyd, D., Pagnamenta, A. T., Twigg, S. R. F., Green, J., McGowan, S. J., Mirza, G., Naples, R., Sharma, V. P., Volpi, E. V., et al. (2014). De novo and rare inherited mutations implicate the transcriptional coregulator TCF20/SPBP in autism spectrum disorder. *J Med Genet* 51, 737–47.

Bannister, A. J. and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. *Nature* 384, 641–643.

Bao, Y. and Shen, X. (2007). INO80 subfamily of chromatin remodeling complexes. *Mutat Res Fundam Mol Mech Mutagen* 618, 18–29.

Barriga, E. H., Shellard, A. and Mayor, R. (2019). Neural Crest Cells, Methods and Protocols. *Methods Mol Biology Clifton N J* 1976, 135–152.

Beest, M. van, Dooijes, D., Wetering, M. van de, Kjaerulff, S., Bonvin, A., Nielsen, O. and Clevers, H. (2000). Sequence-specific High Mobility Group

Box Factors Recognize 10–12-Base Pair Minor Groove Motifs. *J Biol Chem* 275, 27266–27273.

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 National Acad Sci* 80, 5559–5563.

Bi, W., Ohyama, T., Nakamura, H., Yan, J., Visvanathan, J., Justice, M. J. and Lupski, J. R. (2005). Inactivation of Rai1 in mice recapitulates phenotypes observed in chromosome engineered mouse models for Smith–Magenis syndrome. *Hum Mol Genet* 14, 983–995.

Bird, A., Taggart, M., Frommer, M., Miller, O. J. and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* 40, 91–99.

Bönisch, C., Schneider, K., Pünzeler, S., Wiedemann, S. M., Bielmeier, C., Bocola, M., Eberl, H. C., Kuegel, W., Neumann, J., Kremmer, E., et al. (2012). H2A.Z.2.2 is an alternatively spliced histone H2A.Z variant that causes severe nucleosome destabilization. *Nucleic Acids Res* 40, 5951–5964.

Bornelöv, S., Reynolds, N., Xenophontos, M., Gharbi, S., Johnstone, E., Floyd, R., Ralser, M., Signolet, J., Loos, R., Dietmann, S., et al. (2018). The Nucleosome Remodeling and Deacetylation Complex Modulates Chromatin Structure at Sites of Active Transcription to Fine-Tune Gene Expression. *Mol Cell* 71, 56-72.e4.

Brackertz, M., Gong, Z., Leers, J. and Renkawitz, R. (2006). p66α and p66β of the Mi-2/NuRD complex mediate MBD2 and histone interaction. *Nucleic Acids Res* 34, 397–406.

Branco, M. R. and Pombo, A. (2006). Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations. *Plos Biol* 4, e138.

Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P., Smith, S., McCabe, V. M., Norris, D. P., Penny, G. D., Patel, D. and Rastan, S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* 351, 329–331.

Brownell, J. E. and Allis, C. D. (1995). An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. *Proc National Acad Sci* 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.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10, 1213–1218.

Burgold, T., Barber, M., Kloet, S., Cramard, J., Gharbi, S., Floyd, R., Kinoshita, M., Ralser, M., Vermeulen, M., Reynolds, N., et al. (2019). The Nucleosome Remodelling and Deacetylation complex suppresses transcriptional noise during lineage commitment. *Embo J* e100788–e100788.

Carnesecchi, J., Pinto, P. B. and Lohmann, I. (2018). Hox transcription factors: an overview of multi-step regulators of gene expression. *Int J Dev Biol* 62, 723–732.

Casado-Medrano, V., Barrio-Real, L., Wang, A., Cooke, M., Lopez-Haber, C. and Kazanietz, M. G. (2019). Distinctive requirement of PKCε in the control of Rho GTPases in epithelial and mesenchymally transformed lung cancer cells. *Oncogene* 38, 5396–5412.

Ceballos-Chávez, M., Rivero, S., García-Gutiérrez, P., Rodríguez-Paredes, M., García-Domínguez, M., Bhattacharya, S. and Reyes, J. C. (2012). Control of neuronal differentiation by sumoylation of BRAF35, a subunit of the LSD1-CoREST histone demethylase complex. *Proc National Acad Sci* 109, 8085–8090.

Chambeyron, S. and Bickmore, W. A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Gene Dev* 18, 1119–1130.

Choi, J., Heo, K. and An, W. (2009). Cooperative action of TIP48 and TIP49 in H2A.Z exchange catalyzed by acetylation of nucleosomal H2A. *Nucleic Acids Res* 37, 5993–6007.

Chong, J. X., Yu, J.-H., Lorentzen, P., Park, K. M., Jamal, S. M., Tabor, H. K., Rauch, A., Saenz, M. S., Boltshauser, E., Patterson, K. E., et al. (2016). Gene discovery for Mendelian conditions via social networking: de novo variants in KDM1A cause developmental delay and distinctive facial features. *Genet Med* 18, 788–795.

Christensen, J., Agger, K., Cloos, P. A. C., Pasini, D., Rose, S., Sennels, L., Rappsilber, J., Hansen, K. H., Salcini, A. E. and Helin, K. (2007). RBP2 Belongs to a Family of Demethylases, Specific for Tri-and Dimethylated Lysine 4 on Histone 3. *Cell* 128, 1063–1076.

Ciccone, D. N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Xu, G., Li, E. and Chen, T. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* 461, 415–418.

Clapier, C. R. and Cairns, B. R. (2009). The Biology of Chromatin Remodeling Complexes. *Annu Rev Biochem* 78, 273–304.

Clarke, D. J., O'Neill, L. P. and Turner, B. M. (1993). Selective use of H4 acetylation sites in the yeast Saccharomyces cerevisiae. *Biochem J* 294, 557–561.

Cole, L., Kurscheid, S., Nekrasov, M., Domaschenz, R., Vera, D. L., Dennis, J. H. and Tremethick, D. J. (2021). Multiple roles of H2A.Z in regulating promoter chromatin architecture in human cells. *Nat Commun* 12, 2524.

Colino-Sanguino, Y., Clark, S. J. and Valdes-Mora, F. (2022). The H2A.Z-nucleosome code in mammals: emerging functions. *Trends Genet* 38, 273–289.

Consortium, I. H. G. S., Research:, W. I. for B. R., Center for Genome, Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Conway, S. J., Henderson, D. J. and Copp, A. J. (1997). Pax3 is required for cardiac neural crest migration in the mouse: evidence from the splotch (Sp2H) mutant. *Dev Camb Engl* 124, 505–14.

Coon, J. J., Ueberheide, B., Syka, J. E. P., Dryhurst, D. D., Ausio, J., Shabanowitz, J. and Hunt, D. F. (2005). Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc National Acad Sci* 102, 9463–9468.

Corujo, D. and Buschbeck, M. (2018). Post-Translational Modifications of H2A Histone Variants and Their Role in Cancer. *Cancers* 10, 59.

Côté, J., Quinn, J., Workman, J. L. and Peterson, C. L. (1994). Stimulation of GAL4 Derivative Binding to Nucleosomal DNA by the Yeast SWI/SNF Complex. *Science* 265, 53–60.

Cremer, T., Cremer, C., Schneider, T., Baumann, H., Hens, L. and Kirsch-Volders, M. (1982). Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. *Hum Genet* 62, 201–209.

Cremer, M., Grasser, F., Lanctôt, C., Müller, S., Neusser, M., Zinner, R., Solovei, I. and Cremer, T. (2012). The Nucleus, Volume 1: Nuclei and Subnuclear Components. *Methods Mol Biology* 463, 205–239.

Daal, A. van and Elgin, S. C. (1992). A histone variant, H2AvD, is essential in Drosophila melanogaster. *Mol Biol Cell* 3, 593–602.

Daniele, A. D., Antonucci, Y. and Campello, S. (2022). Migrasomes, new vescicles as Hansel and Gretel white pebbles? *Biol Direct* 17, 8.

Delmas, V., Stokes, D. G. and Perry, R. P. (1993). A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. *Proc National Acad Sci* 90, 2414–2418.

Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K., Zhou, M.-M. and Zhou, M.-M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. *Nature* 399, 491–496.

Draizen, E. J., Shaytan, A. K., Mariño-Ramírez, L., Talbert, P. B., Landsman, D. and Panchenko, A. R. (2016). HistoneDB 2.0: a histone database with variants—an integrated resource to explore histones and their variants. *Database J Biological Databases Curation* 2016, baw014.

Draker, R., Ng, M. K., Sarcinella, E., Ignatchenko, V., Kislinger, T. and Cheung, P. (2012). A combination of H2A.Z and H4 acetylation recruits Brd2 to chromatin during transcriptional activation. *Plos Genet* 8, e1003047–e1003047.

Dürr, H., Körner, C., Müller, M., Hickmann, V. and Hopfner, K.-P. (2005). X-Ray Structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase Core and Its Complex with DNA. *Cell* 121, 363–373.

Ebbert, R., Birkmann, A. and Schüller, H. (1999). The product of the SNF2/SWI2 paralogue INO80 of Saccharomyces cerevisiae required for efficient expression of various yeast structural genes is part of a high-molecular-weight protein complex. *Mol Microbiol* 32, 741–751.

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.

Elfring, L. K., Deuring, R., McCallum, C. M., Peterson, C. L. and Tamkun, J. W. (1994). Identification and characterization of Drosophila relatives of the yeast transcriptional activator SNF2/SWI2. *Mol Cell Biol* 14, 2225–2234.

Elsea, S. H. and Williams, S. R. (2011). Smith–Magenis syndrome: haploinsufficiency of RAI1 results in altered gene regulation in neurological and metabolic pathways. *Expert Rev Mol Med* 13, e14.

Ernst, J. and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 9, 215–216.

Ernst, J. and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* 12, 2478–2492.

Faast, R., Thonglairoam, V., Schulz, T. C., Beall, J., Wells, J. R. E., Taylor, H., Matthaei, K., Rathjen, P. D., Tremethick, D. J. and Lyons, I. (2001). Histone variant H2A.Z is required for early mammalian development. *Curr Biol* 11, 1183–1187.

Fan, J. Y., Rangasamy, D., Luger, K. and Tremethick, D. J. (2004). H2A.Z alters the nucleosome surface to promote HP1alpha-mediated chromatin fiber folding. *Mol Cell* 16, 655–661.

Fan, H., Lv, P., Huo, X., Wu, J., Wang, Q., Cheng, L., Liu, Y., Tang, Q.-Q., Zhang, L., Zhang, F., et al. (2018). The nuclear matrix protein HNRNPU maintains 3D genome architecture globally in mouse hepatocytes. *Genome Res* 28, 192–202.

Fang, R., Chen, F., Dong, Z., Hu, D., Barbera, A. J., Clark, E. A., Fang, J., Yang, Y., Mei, P., Rutenberg, M., et al. (2013). LSD2/KDM1B and Its Cofactor NPAC/GLYR1 Endow a Structural and Molecular Model for Regulation of H3K4 Demethylation. *Mol Cell* 49, 558–570.

Fassio, A., Patry, L., Congia, S., Onofri, F., Piton, A., Gauthier, J., Pozzi, D., Messa, M., Defranchi, E., Fadda, M., et al. (2011). SYN1 loss-of-function mutations in autism and partial epilepsy cause impaired synaptic function. *Hum Mol Genet* 20, 2297–2307.

Feng, C., Zhao, J., Ji, F., Su, L., Chen, Y. and Jiao, J. (2020). TCF20 dysfunction leads to cortical neurogenesis defects and autistic-like behaviors in mice. *Embo Rep* 21, e49239.

Ferrai, C., Xie, S. Q., Luraghi, P., Munari, D., Ramirez, F., Branco, M. R., Pombo, A. and Crippa, M. P. (2010). Poised Transcription Factories Prime Silent uPA Gene Prior to Activation. *Plos Biol* 8, e1000270.

Flanagan, J. F., Mi, L.-Z., Chruszcz, M., Cymborowski, M., Clines, K. L., Kim, Y., Minor, W., Rastinejad, F. and Khorasanizadeh, S. (2005). Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* 438, 1181–1185.

Flemming, W. (1882). *Zellsubstanz, Kern und Zelltheilung*. Vogel.

Fritz, A. J., Sehgal, N., Pliss, A., Xu, J. and Berezney, R. (2019). Chromosome territories and the global regulation of the genome. *Genes Chromosom. Cancer* 58, 407–426.

Fujimoto, S., Seebart, C., Guastafierro, T., Prenni, J., Caiafa, P. and Zlatanova, J. (2012). Proteome analysis of protein partners to nucleosomes containing canonical H2A or the variant histones H2A.Z or H2A.X. *Biol Chem* 393, 47–61.

Garay, P. M., Wallner, M. A. and Iwase, S. (2016). Yin-yang actions of histone methylation regulatory complexes in the brain. *Epigenomics-uk* 8, 1689–1708.

Giaimo, B. D., Ferrante, F., Herchenröther, A., Hake, S. B. and Borggrefe, T. (2019). The histone variant H2A.Z in gene regulation. *Epigenetics & Chromatin* 12, 37–37.

Gocke, C. B. and Yu, H. (2008). ZNF198 Stabilizes the LSD1–CoREST–HDAC1 Complex on Chromatin through Its MYM-Type Zinc Fingers. *Plos One* 3, e3255.

Gómez-Marín, E., Posavec-Marjanović, M., Zarzuela, L., Basurto-Cayuela, L., Guerrero-Martínez, J. A., Arribas, G., Yerbes, R., Ceballos-Chávez, M., Rodríguez-Paredes, M., Tomé, M., et al. (2022). The high mobility group protein HMG20A cooperates with the histone reader PHF14 to modulate TGFβ and Hippo pathways. *Nucleic Acids Res* 50, 9838–9857.

Greaves, I. K., Rangasamy, D., Devoy, M., Graves, J. A. M. and Tremethick, D. J. (2006). The X and Y Chromosomes Assemble into H2A.Z, Containing Facultative Heterochromatin, following Meiosis. *Mol Cell Biol* 26, 5394–5405.

Greaves, I. K., Rangasamy, D., Ridgway, P. and Tremethick, D. J. (2007). H2A.Z contributes to the unique 3D structure of the centromere. *Proc National Acad Sci* 104, 525–530.

Greenberg, R. S., Long, H. K., Swigut, T. and Wysocka, J. (2019). Single Amino Acid Change Underlies Distinct Roles of H2A.Z Subtypes in Human Syndrome. *Cell* 178, 1421-1436.e24.

Gruenbaum, Y., Cedar, H. and Razin, A. (1982). Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature* 295, 620–622.

Grunstein, M. (1997). Histone acetylation in chromatin structure and transcription. *Nature* 389, 349–352.

Gursoy-Yuzugullu, O., Ayrapetov, M. K. and Price, B. D. (2015). Histone chaperone Anp32e removes H2A.Z from DNA double-strand breaks and promotes nucleosome reorganization and DNA repair. *Proc National Acad Sci* 112, 7507–7512.

Gustafson, C. M., Roffers-Agarwal, J. and Gammill, L. S. (2022). Chick cranial neural crest cells release extracellular vesicles critical for their migration. *J Cell Sci* 135, jcs260272.

Hao, Y., Liu, Y., Yang, J., Li, X., Luo, F., Geng, Q., Li, S., Li, P., Wu, W. and Xie, J. (2022). Prenatal and postnatal diagnosis of Phelan–McDermid syndrome: A report of 21 cases from a medical center and review of the literature. *Frontiers Genetics* 13, 961196.

Harland, R. M. (1991). In situ hybridization: an improved whole-mount method for Xenopus embryos. *Methods Cell Biol* 36, 685–95.

Hassan, A. H., Awad, S. and Prochasson, P. (2006). The Swi2/Snf2 Bromodomain Is Required for the Displacement of SAGA and the Octamer Transfer of SAGA-acetylated Nucleosomes*. *J Biol Chem* 281, 18126–18134.

Haynes, S. R., Dollard, C., Winston, F., Beck, S., Trowsdale, J. and Dawid, I. B. (1992). The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. *Nucleic Acids Res* 20, 2603–2603.

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.

Heitz, E. (1928). *Das Heterochromatin der Moose*. Bornträger.

Hendrich, B. and Bird, A. (1998). Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Mol Cell Biol* 18, 6538–6547.

Herchenröther, A., Wunderlich, T. M., Lan, J. and Hake, S. B. (2023). Spotlight on histone H2A variants: From B to X to Z. *Semin Cell Dev Biol* 135, 3–12.

Hirota, A., Nakajima-Koyama, M., Ashida, Y. and Nishida, E. (2019). The nucleosome remodeling and deacetylase complex protein CHD4 regulates neural differentiation of mouse embryonic stem cells by down-regulating p53. *J Biol Chem* 294, 195–209.

Hizume, K., Araki, S., Hata, K., Prieto, E., Kundu, T. K., Yoshikawa, K. and Takeyasu, K. (2011). Nano-scale analyses of the chromatin decompaction induced by histone acetylation. *Arch Histol Cytol* 73, 149–163.

Hong, Y.-K., Ontiveros, S. D. and Strauss, W. M. (2000). A revision of the human XIST gene organization and structural comparison with mouse Xist. *Mamm Genome* 11, 220–224.

Horikoshi, N., Sato, K., Shimada, K., Arimura, Y., Osakabe, A., Tachiwana, H., Hayashi-Takanaka, Y., Iwasaki, W., Kagawa, W., Harata, M., et al. (2013). Structural polymorphism in the L1 loop regions of human H2A.Z.1 and H2A.Z.2. *Acta Crystallogr Sect D* 69, 2431–2439.

Hu, G., Cui, K., Northrup, D., Liu, C., Wang, C., Tang, Q., Ge, K., Levens, D., Crane-Robinson, C. and Zhao, K. (2013). H2A.Z Facilitates Access of Active and Repressive Complexes to Chromatin in Embryonic Stem Cell Self-Renewal and Differentiation. *Cell Stem Cell* 12, 180–192.

Huang, Q., Zhang, L., Wang, Y., Zhang, C., Zhou, S., Yang, G., Li, Z., Gao, X., Chen, Z. and Zhang, Z. (2013). Depletion of PHF14, a novel histone-binding protein gene, causes neonatal lethality in mice due to respiratory failure. *Acta Bioch Bioph Sin* 45, 622–633.

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.

Iouzalen, N., Moreau, J. and Méchali, M. (1996). H2A.Zl, a New Variant Histone Expressed during Xenopus Early Development Exhibits Several Distinct Features from the Core Histone H2A. *Nucleic Acids Res* 24, 3947–3952.

Jenuwein, T. and Allis, C. D. (2001). Translating the histone code. *Science* 293, 1074–80.

Jenuwein, T., Laible, G., Dorn, R. and Reuter, G. (1998). SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell Mol Life Sci Cmls* 54, 80–93.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K. and Felsenfeld, G. (2009). H3.3/H2A.Z double variant–containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions. *Nat Genet* 41, 941–945.

Jones, W. D., Dafou, D., McEntagart, M., Woollard, W. J., Elmslie, F. V., Holder-Espinasse, M., Irving, M., Saggar, A. K., Smithson, S., Trembath, R. C., et al. (2012). De Novo Mutations in MLL Cause Wiedemann-Steiner Syndrome. *Am J Hum Genetics* 91, 358–364.

Joober, R., Benkelfat, C., Toulouse, A., Lafrenière, R. G., Lal, S., Ajroud, S., Turecki, G., Bloom, D., Labelle, A., Lalonde, P., et al. (1999). Analysis of 14 CAG repeat-containing genes in schizophrenia. *Am. J. Med. Genet.* 88, 694–9.

Kaminker, J. S., Bergman, C. M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D. A., Lewis, S. E., Rubin, G. M., et al. (2002). The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. *Genome Biol* 3, research0084.1.

Käsper, E.-L., Hwang, I.-Y., Grötsch, H., Fung, H. K. H., Sérandour, A. A., Arecco, N., Oellers, R., Poliński, P., Gomez-Monedero, A. G., Müller, C. W., et al. (2021). The neurodevelopmental disorder-linked PHF14 complex that forms biomolecular condensates detects DNA damage and promotes repair. *Biorxiv* 2021.10.12.462922.

Kitagawa, M., Takebe, A., Ono, Y., Imai, T., Nakao, K., Nishikawa, S.-I. and Era, T. (2012). Phf14, a Novel Regulator of Mesenchyme Growth via Platelet-derived Growth Factor (PDGF) Receptor-α*. *J Biol Chem* 287, 27983–27996.

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.

Kobor, Michael. S., Venkatasubrahmanyam, S., Meneghini, M. D., Gin, J. W., Jennings, J. L., Link, A. J., Madhani, H. D. and Rine, J. (2004). A Protein Complex Containing the Conserved Swi2/Snf2-Related ATPase Swr1p Deposits Histone Variant H2A.Z into Euchromatin. *Plos Biol* 2, e131.

Kornberg, R. D. (1974). Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* 184, 868–871.

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M. N. and Sergushichev, A. (2021). Fast gene set enrichment analysis. *Biorxiv* 060012.

Krebs, J. E., Kuo, M.-H., Allis, C. D. and Peterson, C. L. (1999). Cell cycle-regulated histone acetylation required for expression of the yeast HO gene. *Gene Dev* 13, 1412–1421.

Kreienbaum, C., Paasche, L. W. and Hake, S. B. (2022). H2A.Z's 'social' network: functional partners of an enigmatic histone variant. *Trends Biochem Sci*.

Krogan, N. J., Dover, J., Khorrami, S., Greenblatt, J. F., Schneider, J., Johnston, M. and Shilatifard, A. (2002). COMPASS, a Histone H3 (Lysine 4) Methyltransferase Required for Telomeric Silencing of Gene Expression*. *J Biol Chem* 277, 10753–10755.

Krogan, N. J., Keogh, M.-C., Datta, N., Sawa, C., Ryan, O. W., Ding, H., Haw, R. A., Pootoolal, J., Tong, A., Canadien, V., et al. (2003). A Snf2 Family ATPase Complex Required for Recruitment of the Histone H2A Variant Htz1. *Mol Cell* 12, 1565–1576.

Kulesa, P. M. and Fraser, S. E. (2000). In ovo time-lapse analysis of chick hindbrain neural crest cell migration shows cell interactions during migration to the branchial arches. *Development* 127, 1161–1172.

Kumar, J. P. (2008). The sine oculis homeobox (SIX) family of transcription factors as regulators of development and disease. *Cell Mol Life Sci* 66, 565.

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. *Gene Dev* 16, 2893–2905.

Kwon, H., Imbalzano, A. N., Khavari, P. A., Kingston, R. E. and Green, M. R. (1994). Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature* 370, 477–481.

Laird, C. D., Pleasant, N. D., Clark, A. D., Sneeden, J. L., Hassan, K. M. A., Manley, N. C., Vary, J. C., Morgan, T., Hansen, R. S. and Stöger, R. (2004). Hairpin-bisulfite PCR: Assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proc National Acad Sci* 101, 204–209.

Lamaa, A., Humbert, J., Aguirrebengoa, M., Cheng, X., Nicolas, E., Côté, J. and Trouche, D. (2020). Integrated analysis of H2A.Z isoforms function reveals a complex interplay in gene regulation. *Elife* 9, e53375.

Lange, M., Kaynak, B., Forster, U. B., Tönjes, M., Fischer, J. J., Grimm, C., Schlesinger, J., Just, S., Dunkel, I., Krueger, T., et al. (2008). Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Gene Dev* 22, 2370–2384.

Laugesen, A. and Helin, K. (2014). Chromatin Repressive Complexes in Stem Cells, Development, and Cancer. *Cell Stem Cell* 14, 735–751.

Lee, M. G., Wynder, C., Cooch, N. and Shiekhattar, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* 437, 432–435.

Leidescher, S., Ribisel, J., Ullrich, S., Feodorova, Y., Hildebrand, E., Galitsyna, A., Bultmann, S., Link, S., Thanisch, K., Mulholland, C., et al. (2022). Spatial organization of transcribed eukaryotic genes. *Nat Cell Biol* 24, 327–339.

Lévy, J., Cogan, G., Maruani, A., Maillard, A., Dupont, C., Drunat, S., Rachid, M., Atzori, P., Delorme, R., Jeyarajah, S., et al. (2022). Rare and de novo duplications containing TCF20 are associated with a neurodevelopmental disorder. *Clin Genet* 101, 364–370.

Lewis, T. S., Sokolova, V., Jung, H., Ng, H. and Tan, D. (2021). Structural basis of chromatin regulation by histone variant H2A.Z. *Nucleic Acids Res* 49, 11379–11391.

Li, Y., Wen, H., Xi, Y., Tanaka, K., Wang, H., Peng, D., Ren, Y., Jin, Q., Dent, S. Y. R., Li, W., et al. (2014). AF9 YEATS Domain Links Histone Acetylation to DOT1L-Mediated H3K79 Methylation. *Cell* 159, 558–571.

Liang, X., Shan, S., Pan, L., Zhao, J., Ranjan, A., Wang, F., Zhang, Z., Huang, Y., Feng, H., Wei, D., et al. (2016). Structural basis of H2A.Z recognition by SRCAP chromatin-remodeling subunit YL1. *Nat Struct Mol Biol* 23, 317–323.

Liao, C. and Seebeck, F. P. (2019). S-adenosylhomocysteine as a methyl transfer catalyst in biocatalytic methylation reactions. *Nat Catal* 2, 696–701.

Lignani, G., Raimondi, A., Ferrea, E., Rocchi, A., Paonessa, F., Cesca, F., Orlando, M., Tkatch, T., Valtorta, F., Cossette, P., et al. (2013). Epileptogenic Q555X SYN1 mutant triggers imbalances in release dynamics and short-term plasticity. *Hum Mol Genet* 22, 2186–2199.

Link, S., Spitzer, R. M. M., Sana, M., Torrado, M., Völker-Albert, M. C., Keilhauer, E. C., Burgold, T., Pünzeler, S., Low, J. K. K., Lindström, I., et al. (2018). PWWP2A binds distinct chromatin moieties and interacts with an MTA1-specific core NuRD complex. *Nat Commun* 9, 4300–4300.

Long, M., Sun, X., Shi, W., Yanru, A., Leung, S. T. C., Ding, D., Cheema, M. S., MacPherson, N., Nelson, C. J., Ausio, J., et al. (2019). A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. *Nucleic Acids Res*.

Lorenzo, P. I., Vazquez, E. M., López-Noriega, L., Fuente-Martín, E., Mellado-Gil, J. M., Franco, J. M., Cobo-Vuilleumier, N., Martínez, J. A. G., Romero-Zerbo, S. Y., Perez-Cabello, J. A., et al. (2021). The metabesity factor HMG20A potentiates astrocyte survival and reactive astrogliosis preserving neuronal integrity. *Theranostics* 11, 6983–7004.

Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550.

Low, J. K. K., Silva, A. P. G., Tabar, M. S., Torrado, M., Webb, S. R., Parker, B. L., Sana, M., Smits, C., Schmidberger, J. W., Brillault, L., et al. (2020). The Nucleosome Remodeling and Deacetylase Complex Has an Asymmetric, Dynamic, and Modular Architecture. *Cell Reports* 33, 108450.

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 Å resolution. *Nature* 389, 251–260.

Luk, E., Ranjan, A., FitzGerald, P. C., Mizuguchi, G., Huang, Y., Wei, D. and Wu, C. (2010). Stepwise Histone Replacement by SWR1 Requires Dual Activation with Histone H2A.Z and Canonical Nucleosome. *Cell* 143, 725–736.

Luo, Z., Gao, X., Lin, C., Smith, E. R., Marshall, S. A., Swanson, S. K., Florens, L., Washburn, M. P. and Shilatifard, A. (2015). Zic2 Is an Enhancer-Binding Factor Required for Embryonic Stem Cell Specification. *Mol Cell* 57, 685–694.

Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A. and Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res* 50, W276–W279.

Manuelidis, L. (1985). Individual interphase chromosome domains revealed by in situ hybridization. *Hum Genet* 71, 288–293.

Mao, Z., Pan, L., Wang, W., Sun, J., Shan, S., Dong, Q., Liang, X., Dai, L., Ding, X., Chen, S., et al. (2014). Anp32e, a higher eukaryotic histone chaperone directs preferential recognition for H2A.Z. *Cell Res* 24, 389–399.

Marabelli, C., Marrocco, B., Pilotto, S., Chittori, S., Picaud, S., Marchese, S., Ciossani, G., Forneris, F., Filippakopoulos, P., Schoehn, G., et al. (2019). A Tail-Based Mechanism Drives Nucleosome Demethylation by the LSD2/NPAC Multimeric Complex. *Cell Reports* 27, 387-399.e7.

Marmorstein, R. and Zhou, M.-M. (2014). Writers and Readers of Histone Acetylation: Structure, Mechanism, and Inhibition. *Csh Perspect Biol* 6, a018762.

Marsano, R. M. and Dimitri, P. (2022). Constitutive Heterochromatin in Eukaryotic Genomes: A Mine of Transposable Elements. *Cells* 11, 761.

Martinez-Morales, J.-R., Henrich, T., Ramialison, M., Wittbrodt, J. and Martinez-Morales, J.-R. (2007). New genes in the evolution of the neural crest differentiation program. *Genome Biol* 8, R36–R36.

Mayran, A. and Drouin, J. (2018). Pioneer transcription factors shape the epigenetic landscape. *J Biol Chem* 293, 13795–13804.

McClellan, D., Casey, M. J., Bareyan, D., Lucente, H., Ours, C., Velinder, M., Singer, J., Lone, M. D., Sun, W., Coria, Y., et al. (2019). Growth Factor Independence 1B-Mediated Transcriptional Repression and Lineage Allocation Require Lysine-Specific Demethylase 1-Dependent Recruitment of the BHC Complex. *Mol Cell Biol* 39,.

McKinney, M. C., Stark, D. A., Teddy, J. and Kulesa, P. M. (2011). Neural crest cell communication involves an exchange of cytoplasmic material through cellular bridges revealed by photoconversion of KikGR. *Dev. Dyn.* 240, 1391–1401.

Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. and Bird, A. P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 58, 499–507.

Megee, P. C., Morgan, B. A., Mittman, B. A. and Smith, M. M. (1990). Genetic Analysis of Histone H4: Essential Role of Lysines Subject to Reversible Acetylation. *Science* 247, 841–845.

Mellado-Gil, J. M., Fuente-Martín, E., Lorenzo, P. I., Cobo-Vuilleumier, N., López-Noriega, L., Martín-Montalvo, A., Gómez, I. de G. H., Ceballos-Chávez, M., Gómez-Jaramillo, L., Campos-Caro, A., et al. (2018). The type 2 diabetes-associated HMG20A gene is mandatory for islet beta cell functional maturity. *Cell Death Dis* 9, 279–279.

Meyers, B. C., Tingey, S. V. and Morgante, M. (2001). Abundance, Distribution, and Transcriptional Activity of Repetitive Elements in the Maize Genome. *Genome Res* 11, 1660–1676.

Milne, T. A., Briggs, S. D., Brock, H. W., Martin, M. E., Gibbs, D., Allis, C. D. and Hess, J. L. (2002). MLL Targets SET Domain Methyltransferase Activity to Hox Gene Promoters. *Mol Cell* 10, 1107–1117.

Mizuguchi, G., Shen, X., Landry, J., Wu, W.-H., Sen, S. and Wu, C. (2004). ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. *Science* 303, 343–348.

Mizzen, C. A., Yang, X.-J., Kokubo, T., Brownell, J. E., Bannister, A. J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S. L., Kouzarides, T., et al. (1996). The TAFII250 Subunit of TFIID Has Histone Acetyltransferase Activity. *Cell* 87, 1261–1270.

Muerdter, F., Boryń, Ł. M., Woodfin, A. R., Neumayr, C., Rath, M., Zabidi, M. A., Pagani, M., Haberle, V., Kazmar, T., Catarino, R. R., et al. (2018). Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat Methods* 15, 141–149.

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.

Musselman, C. A., Khorasanizadeh, S. and Kutateladze, T. G. (2014). Towards understanding methyllysine readout. *Biochimica Et Biophysica Acta Bba - Gene Regul Mech* 1839, 686–693.

Nakagawa, S. and Takeichi, M. (1998). Neural crest emigration from the neural tube depends on regulated cadherin expression. *Development* 125, 2963–2971.

Nan, X., Campoy, F. J. and Bird, A. (1997). MeCP2 Is a Transcriptional Repressor with Abundant Binding Sites in Genomic Chromatin. *Cell* 88, 471–481.

Nasrullah, Hussain, A., Ahmed, S., Rasool, M. and Shah, A. J. (2022). DNA methylation across the tree of life, from micro to macro-organism. *Bioengineered* 13, 1666–1685.

Natsume, T., Kiyomitsu, T., Saga, Y. and Kanemaki, M. T. (2016). Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Reports* 15, 210–218.

Neigeborn, L. and Carlson, M. (1984). Genes affecting the regulation of suc2 gene expression by glucose repression in saccharomyces cerevisiae. *Genetics* 108, 845–858.

Neri, F., Krepelova, A., Incarnato, D., Maldotti, M., Parlato, C., Galvagni, F., Matarese, F., Stunnenberg, H. G. and Oliviero, S. (2013). Dnmt3L Antagonizes DNA Methylation at Bivalent Promoters and Favors DNA Methylation at Gene Bodies in ESCs. *Cell* 155, 121–134.

Nguyen, H. Q., Chattoraj, S., Castillo, D., Nguyen, S. C., Nir, G., Lioutas, A., Hershberg, E. A., Martins, N. M. C., Reginato, P. L., Hannan, M., et al. (2020). 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat Methods* 17, 822–832.

Nicholson, T. B., Su, H., Hevi, S., Wang, J., Bajko, J., Li, M., Valdez, R., Loureiro, J., Cheng, X., Li, E., et al. (2011). Defective heart development in hypomorphic LSD1 mice. *Cell Res* 1–16.

Nicholson, T. B., Singh, A. K., Su, H., Hevi, S., Wang, J., Bajko, J., Li, M., Valdez, R., Goetschkes, M., Capodieci, P., et al. (2013). A Hypomorphic Lsd1 Allele Results in Heart Development Defects in Mice. *Plos One* 8, e60913.

Nieto, M. A., Sargent, M. G., Wilkinson, D. G. and Cooke, J. (1994). Control of Cell Behavior During Vertebrate Development by Slug, a Zinc Finger Gene. *Science* 264, 835–839.

Nodelman, I. M., Bleichert, F., Patel, A., Ren, R., Horvath, K. C., Berger, J. M. and Bowman, G. D. (2017). Interdomain Communication of the Chd1 Chromatin Remodeler across the DNA Gyres of the Nucleosome. *Mol Cell* 65, 447-459.e6.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., Berkum, N. L. van, Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385.

Obri, A., Ouararhni, K., Papin, C., Diebold, M.-L., Padmanabhan, K., Marek, M., Stoll, I., Roy, L., Reilly, P. T., Mak, T. W., et al. (2014). ANP32E is a histone chaperone that removes H2A.Z from chromatin. *Nature* 505, 648–653.

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.

Ohki, I., Shimotake, N., Fujita, N., Jee, J.-G., Ikegami, T., Nakao, M. and Shirakawa, M. (2001). Solution Structure of the Methyl-CpG Binding Domain of Human MBD1 in Complex with Methylated DNA. *Cell* 105, 487–497.

Olins, A. L. and Olins, D. E. (1974). Spheroid Chromatin Units (ν Bodies). *Science* 183, 330–332.

Olins, D. E. and Olins, A. L. (2003). Chromatin history: our view from the bridge. *Nat Rev Mol Cell Bio* 4, 809–814.

O'Neill, L. P. and Turner, B. M. (1995). Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner. *Embo J* 14, 3946–3957.

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.

Ortmann, D., Brown, S., Czechanski, A., Aydin, S., Muraro, D., Huang, Y., Tomaz, R. A., Osnato, A., Canu, G., Wesley, B. T., et al. (2020). Naive Pluripotent Stem Cells Exhibit Phenotypic Variability that Is Driven by Genetic Variation. *Cell Stem Cell* 27, 470-481.e6.

Papamichos-Chronakis, M., Watanabe, S., Rando, O. J. and Peterson, C. L. (2011). Global Regulation of H2A.Z Localization by the INO80 Chromatin-Remodeling Enzyme Is Essential for Genome Integrity. *Cell* 144, 200–213.

Paro, R. and Hogness, D. S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proc National Acad Sci* 88, 263–267.

Payne, A. C., Chiang, Z. D., Reginato, P. L., Mangiameli, S. M., Murray, E. M., Yao, C.-C., Markoulaki, S., Earl, A. S., Labade, A. S., Jaenisch, R., et al. (2021). In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* 371,.

Pazin, M. J. and Kadonaga, J. T. (1997). SWI2/SNF2 and Related Proteins: ATP-Driven Motors That Disrupt-Protein–DNA Interactions? *Cell* 88, 737–740.

Perell, G. T., Mishra, N. K., Sudhamalla, B., Ycas, P. D., Islam, K. and Pomerantz, W. C. K. (2017). Specific Acetylation Patterns of H2A.Z Form Transient Interactions with the BPTF Bromodomain. *Biochemistry-us* 56, 4607–4615.

Potocki, L., Bi, W., Treadwell-Deering, D., Carvalho, C. M. B., Eifert, A., Friedman, E. M., Glaze, D., Krull, K., Lee, J. A., Lewis, R. A., et al. (2007). Characterization of Potocki-Lupski Syndrome (dup(17)(p11.2p11.2)) and Delineation of a Dosage-Sensitive Critical Interval That Can Convey an Autism Phenotype. *Am J Hum Genetics* 80, 633–649.

Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R. and Grant, P. A. (2005). Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* 433, 434–438.

Procida, T., Friedrich, T., Jack, A. P. M., Peritore, M., Bönisch, C., Eberl, H. C., Daus, N., Kletenkov, K., Nist, A., Stiewe, T., et al. (2021). JAZF1, A Novel p400/TIP60/NuA4 Complex Member, Regulates H2A.Z Acetylation at Regulatory Regions. *Int J Mol Sci* 22, 678.

Pünzeler, S., Link, S., Wagner, G., Keilhauer, E. C., Kronbeck, N., Spitzer, R. M., Leidescher, S., Markaki, Y., Mentele, E., Regnard, C., et al. (2017). Multivalent binding of PWWP2A to H2A.Z regulates mitosis and neural crest differentiation. *Embo J* 36, 2263–2279.

Qiu, C., Sawada, K., Zhang, X. and Cheng, X. (2002). The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nat Struct Biol* 9, 217–224.

Qureshi, I. A., Gokhan, S. and Mehler, M. F. (2010). REST and CoREST are transcriptional and epigenetic regulators of seminal neural fate decisions. *Cell Cycle* 9, 4477–4486.

Ramazi, S., Allahverdi, A. and Zahiri, J. (2020). Evaluation of post-translational modifications in histone proteins: A review on histone modification defects in developmental and neurological disorders. *J Biosciences* 45, 135.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A. and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8, 2281–2308.

Rando, O. J., Zhao, K., Janmey, P. and Crabtree, G. R. (2002). Phosphatidylinositol-dependent actin filament binding by the SWI/SNF-like BAF chromatin remodeling complex. *Proc National Acad Sci* 99, 2824–2829.

Razin, A. and Riggs, A. D. (1980). DNA Methylation and Gene Function. *Science* 210, 604–610.

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.

Reid, X. J., Low, J. K. K. and Mackay, J. P. (2023). A NuRD for all seasons. *Trends Biochem Sci* 1, 11–25.

Richmond, T. J. and Davey, C. A. (2003). The structure of DNA in the nucleosome core. *Nature* 423, 145–150.

Rivera, C., Lee, H.-G., Lappala, A., Wang, D., Noches, V., Olivares-Costa, M., Sjöberg-Herrera, M., Lee, J. T. and Andrés, M. E. (2022). Unveiling RCOR1 as a rheostat at transcriptionally permissive chromatin. *Nat Commun* 13, 1550.

Rivero, S., Ceballos-Chávez, M., Bhattacharya, S. S. and Reyes, J. C. (2015). HMG20A is required for SNAI1-mediated epithelial to mesenchymal transition. *Oncogene* 34, 5264–5276.

Roberts, S. M. and Winston, F. (1997). Essential Functional Interactions of SAGA, a Saccharomyces cerevisiae Complex of Spt, Ada, and Gcn5 Proteins, With the Snf/Swi and Srb/Mediator Complexes. *Genetics* 147, 451–465.

Ryan, D. P. and Tremethick, D. J. (2018). The interplay between H2A.Z and H3K9 methylation in regulating HP1 binding to linker histone-containing chromatin. *Nucleic Acids Res* 46, 9353–9366.

Sanmiguel, P. and Bennetzen, J. L. (1998). Evidence that a Recent Increase in Maize Genome Size was Caused by the Massive Amplification of Intergene Retrotransposons. *Ann Bot-london* 82, 37–44.

Santagati, F. and Rijli, F. M. (2003). Cranial neural crest and the building of the vertebrate head. *Nat Rev Neurosci* 4, 806–818.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C. T., Schreiber, S. L., Mellor, J. and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature* 419, 407–411.

Sathyan, K. M., Shen, Z., Tripathi, V., Prasanth, K. V. and Prasanth, S. G. (2011). A BEN-domain-containing protein associates with heterochromatin and represses transcription. *J Cell Sci* 124, 3149–3163.

Scarpa, E. and Mayor, R. (2016). Collective cell migration in development. *J Cell Biol* 212, 143–155.

Schardin, M., Cremer, T., Hager, H. D. and Lang, M. (1985). Specific staining of human chromosomes in Chinese hamster x man hybrid cell lines demonstrates interphase chromosome territories. *Hum Genet* 71, 281–287.

Schindler, U., Beckmann, H. and Cashmore, A. R. (1993). HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. *Plant J* 4, 137–150.

Schöler, H. R., Ruppert, S., Suzuki, N., Chowdhury, K. and Gruss, P. (1990). New type of POU domain in germ line-specific protein Oct-4. *Nature* 344, 435–439.

Shen, X., Mizuguchi, G., Hamiche, A. and Wu, C. (2000). A chromatin remodelling complex involved in transcription and DNA processing. *Nature* 406, 541–544.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A. and Shi, Y. (2004). Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* 119, 941–953.

Shi, X., Hong, T., Walter, K. L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Peña, P., Lan, F., Kaadige, M. R., et al. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* 442, 96–99.

Shukla, M. S., Syed, S. H., Goutte-Gattat, D., Richard, J. L. C., Montel, F., Hamiche, A., Travers, A., Faivre-Moskalenko, C., Bednar, J., Hayes, J. J., et al. (2011). The docking domain of histone H2A is required for H1 binding and RSC-mediated nucleosome remodeling. *Nucleic Acids Res* 39, 2559–2570.

Sims, R. J., Chen, C.-F., Santos-Rosa, H., Kouzarides, T., Patel, S. S. and Reinberg, D. (2005). Human but Not Yeast CHD1 Binds Directly and Selectively to Histone H3 Methylated at Lysine 4 via Its Tandem Chromodomains*. *J Biol Chem* 280, 41789–41792.

Sinha, K. M., Yasuda, H., Coombes, M. M., Dent, S. Y. R. and Crombrugghe, B. de (2010). Regulation of the osteoblast-specific transcription factor Osterix by NO66, a Jumonji family histone demethylase. *Embo J* 29, 68–79.

Song, Y., Dagil, L., Fairall, L., Robertson, N., Wu, M., Ragan, T. J., Savva, C. G., Saleh, A., Morone, N., Kunze, M. B. A., et al. (2020). Mechanism of Crosstalk between the LSD1 Demethylase and HDAC1 Deacetylase in the CoREST Complex. *Cell Reports* 30, 2699-2711.e8.

Soo, K., O'Rourke, M. P., Khoo, P.-L., Steiner, K. A., Wong, N., Behringer, R. R. and Tam, P. P. L. (2002). Twist Function Is Required for the Morphogenesis of the Cephalic Neural Tube and the Differentiation of the Cranial Neural Crest Cells in the Mouse Embryo. *Dev Biol* 247, 251–270.

Stern, M., Jensen, R. and Herskowitz, I. (1984). Five SWI genes are required for expression of the HO gene in yeast. *J Mol Biol* 178, 853–868.

Strahl, B. D. and Allis, C. D. (2000). The language of covalent histone modifications. *Nature* 403, 41–45.

Štros, M., Launholt, D. and Grasser, K. D. (2007). The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell Mol Life Sci* 64, 2590.

Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B. and Zhuang, X. (2020). Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* 182, 1641-1659.e26.

Svorenova, T., Romito, L. M., Colangelo, I., Han, V., Jech, R., Prokisch, H., Winkelmann, J., Skorvanek, M., Garavaglia, B. and Zech, M. (2022). Dystonia as a prominent feature of TCF20-associated neurodevelopmental disorder: Expanding the phenotype. *Parkinsonism Relat D* 102, 89–91.

Swarr, D. T., Bloom, D., Lewis, R. A., Elenberg, E., Friedman, E. M., Glotzbach, C., Wissman, S. D., Shaffer, L. G. and Potocki, L. (2010). Potocki–Shaffer syndrome: Comprehensive clinical assessment, review of the literature, and proposals for medical management. *Am J Med Genet A* 152A, 565–572.

Szerlong, H., Saha, A. and Cairns, B. R. (2003). The nuclear actin-related proteins Arp7 and Arp9: a dimeric module that cooperates with architectural proteins for chromatin remodeling. *Embo J* 22, 3175–3187.

Tachibana, M., Sugimoto, K., Fukushima, T. and Shinkai, Y. (2001). SET Domain-containing Protein, G9a, Is a Novel Lysine-preferring Mammalian Histone Methyltransferase with Hyperactivity and Specific Selectivity to Lysines 9 and 27 of Histone H3*. *J Biol Chem* 276, 25309–25317.

Tahir, R., Kennedy, A., Elsea, S. H. and Dickinson, A. J. (2014). Retinoic acid induced-1 (Rai1) regulates craniofacial and brain development in Xenopus. *Mech Develop* 133, 91–104.

Takei, Y., Yun, J., Zheng, S., Ollikainen, N., Pierson, N., White, J., Shah, S., Thomassie, J., Suo, S., Eng, C.-H. L., et al. (2021). Integrated spatial genomics reveals global architecture of single nuclei. *Nature* 590, 344–350.

Talbert, P. B. and Henikoff, S. (2020). What makes a centromere? *Exp Cell Res* 389, 111895.

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.

Tripathi, K. and Garg, M. (2018). Mechanistic regulation of epithelial-to-mesenchymal transition through RAS signaling pathway and therapeutic implications in human cancer. *J Cell Commun Signal* 12, 513–527.

Trojer, P. and Reinberg, D. (2007). Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Mol Cell* 28, 1–13.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G. and Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *Embo J* 13, 3822–3831.

Tunovic, S., Barkovich, J., Sherr, E. H. and Slavotinek, A. M. (2014). De novo ANKRD11 and KDM1A gene mutations in a male with features of KBG syndrome and Kabuki syndrome. *Am J Med Genet A* 164, 1744–1749.

Tyagi, M., Imam, N., Verma, K. and Patel, A. K. (2016). Chromatin remodelers: We are the drivers!! *Nucleus* 7, 388–404.

Upadia, J., Gonzales, P. R., Atkinson, T. P., Schroeder, H. W., Robin, N. H., Rudy, N. L. and Mikhail, F. M. (2018). A previously unrecognized 22q13.2 microdeletion syndrome that encompasses TCF20 and TNFRSF13C. *Am J Med Genet A* 176, 2791–2797.

Vardabasso, C., Hasson, D., Ratnakumar, K., Chung, C.-Y., Duarte, L. F. and Bernstein, E. (2014). Histone variants: emerging players in cancer biology. *Cell Mol Life Sci* 71, 379–404.

Vardabasso, C., Gaspar-Maia, A., Hasson, D., Pünzeler, S., Valle-Garcia, D., Straub, T., Keilhauer, E. C., Strub, T., Dong, J., Panda, T., et al. (2015). Histone Variant H2A.Z.2 Mediates Proliferation and Drug Sensitivity of Malignant Melanoma. *Mol Cell* 59, 75–88.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., et al. (2001). The Sequence of the Human Genome. *Science* 291, 1304–1351.

Vetrini, F., McKee, S., Rosenfeld, J. A., Suri, M., Lewis, A. M., Nugent, K. M., Roeder, E., Littlejohn, R. O., Holder, S., Zhu, W., et al. (2019). De novo and inherited TCF20 pathogenic variants are associated with intellectual disability, dysmorphic features, hypotonia, and neurological impairments with similarities to Smith-Magenis syndrome. *Genome Med* 11, 12–12.

Vezzoli, A., Bonadies, N., Allen, M. D., Freund, S. M. V., Santiveri, C. M., Kvinlaug, B. T., Huntly, B. J. P., Göttgens, B. and Bycroft, M. (2010). Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nat Struct Mol Biol* 17, 617–619.

Vicient, C. M., Suoniemi, A., Anamthawat-Jónsson, K., Tanskanen, J., Beharav, A., Nevo, E. and Schulman, A. H. (1999). Retrotransposon BARE-1 and Its Role in Genome Evolution in the Genus Hordeum. *Plant Cell* 11, 1769–1784.

Wang, P., Yang, W., Zhao, S. and Nashun, B. (2021a). Regulation of chromatin structure and function: insights into the histone chaperone FACT. *Cell Cycle* 20, 465–479.

Wang, X., Xiang, Y., Yu, Y., Wang, R., Zhang, Y., Xu, Q., Sun, H., Zhao, Z.-A., Jiang, X., Wang, X., et al. (2021b). Formative pluripotent stem cells show features of epiblast cells poised for gastrulation. *Cell Res* 31, 526–541.

Whyte, W. A., Bilodeau, S., Orlando, D. A., Hoke, H. A., Frampton, G. M., Foster, C. T., Cowley, S. M. and Young, R. A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* 482, 221–225.

Wilczewski, C. M., Hepperla, A. J., Shimbo, T., Wasson, L., Robbe, Z. L., Davis, I. J., Wade, P. A. and Conlon, F. L. (2018). CHD4 and the NuRD complex directly control cardiac sarcomere formation. *P Natl Acad Sci Usa* 115, 6727–6732.

Williams, S. R., Aldred, M. A., Kaloustian, V. M. D., Halal, F., Gowans, G., McLeod, D. R., Zondag, S., Toriello, H. V., Magenis, R. E. and Elsea, S. H. (2010). Haploinsufficiency of HDAC4 Causes Brachydactyly Mental Retardation Syndrome, with Brachydactyly Type E, Developmental Delays, and Behavioral Problems. *Am J Hum Genetics* 87, 219–228.

Woodage, T., Basrai, M. A., Baxevanis, A. D., Hieter, P. and Collins, F. S. (1997). Characterization of the CHD family of proteins. *Proc National Acad Sci* 94, 11472–11477.

Wynder, C., Hakimi, M.-A., Epstein, J. A., Shilatifard, A. and Shiekhattar, R. (2005). Recruitment of MLL by HMG-domain protein iBRAF promotes neural differentiation. *Nat Cell Biol* 7, 1113–1117.

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.

Yamamoto, M., Kondo, R., Hozumi, H., Doi, S., Denda, M., Magari, M., Kanayama, N., Hatano, N., Morishita, R. and Tokumitsu, H. (2021). Identification and Biochemical Characterization of High Mobility Group Protein 20A as a Novel Ca2+/S100A6 Target. *Biomol* 11, 510.

Yamanaka, T., Wong, H. K., Tosaki, A., Bauer, P. O., Wada, K., Kurosawa, M., Shimogori, T., Hattori, N. and Nukina, N. (2014). Large-Scale RNA Interference Screening in Mammalian Cells Identifies Novel Regulators of Mutant Huntingtin Aggregation. *Plos One* 9, e93891–e93891.

Yamane, K., Tateishi, K., Klose, R. J., Fang, J., Fabrizio, L. A., Erdjument-Bromage, H., Taylor-Papadimitriou, J., Tempst, P. and Zhang, Y. (2007). PLU-1 Is an H3K4 Demethylase Involved in Transcriptional Repression and Breast Cancer Cell Proliferation. *Mol Cell* 25, 801–812.

Yan, J., Bi, W. and Lupski, J. R. (2007). Penetrance of Craniofacial Anomalies in Mouse Models of Smith-Magenis Syndrome Is Modified by Genomic Sequence Surrounding Rai1: Not All Null Alleles Are Alike. *Am J Hum Genetics* 80, 518–525.

Yanagisawa, Y., Ito, E., Yuasa, Y. and Maruyama, K. (2002). The human DNA methyltransferases DNMT3A and DNMT3B have two types of promoters with different CpG contents. *Biochimica Et Biophysica Acta Bba - Gene Struct Expr* 1577, 457–465.

Yesbolatova, A., Natsume, T., Hayashi, K. and Kanemaki, M. T. (2019). Generation of conditional auxin-inducible degron (AID) cells and tight control of degron-fused proteins using the degradation inhibitor auxinole. *Methods* 164–165, 7–80.

Yin, F., Lan, R., Zhang, X., Zhu, L., Chen, F., Xu, Z., Liu, Y., Ye, T., Sun, H., Lu, F., et al. (2014). LSD1 Regulates Pluripotency of Embryonic Stem/Carcinoma Cells through Histone Deacetylase 1-Mediated Deacetylation of Histone H4 at Lysine 16. *Mol Cell Biol* 34, 158–179.

Yu, G., Wang, L.-G. and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31, 2382–2383.

Yu, M., Zhan, J. and Zhang, H. (2020). HOX family transcription factors: Related signaling pathways and post-translational modifications in cancer. *Cell Signal* 66, 109469.

Zeng, L., Zhang, Q., Li, S., Plotnikov, A. N., Walsh, M. J. and Zhou, M.-M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* 466, 258–262.

Zhang, Y., Ng, H.-H., Erdjument-Bromage, H., Tempst, P., Bird, A. and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Gene Dev* 13, 1924–1935.

Zhang, Y., Ku, W. L., Liu, S., Cui, K., Jin, W., Tang, Q., Lu, W., Ni, B. and Zhao, K. (2017). Genome-wide identification of histone H2A and histone variant H2A.Z-interacting proteins by bPPI-seq. *Cell Res* 27, 1258–1274.

Zhang, T., Wei, G., Millard, C. J., Fischer, R., Konietzny, R., Kessler, B. M., Schwabe, J. W. R. and Brockdorff, N. (2018). A variant NuRD complex containing PWWP2A/B excludes MBD2/3 to regulate transcription at active genes. *Nat Commun* 9, 3798–3798.

Zhang, L., Song, D., Zhu, B. and Wang, X. (2019). The role of nuclear matrix protein HNRNPU in maintaining the architecture of 3D genome. *Semin Cell Dev Biol* 90, 161–167.

Zheng, S., Bi, Y., Chen, H., Gong, B., Jia, S. and Li, H. (2021). Molecular basis for bipartite recognition of histone H3 by the PZP domain of PHF14. *Nucleic Acids Res* 49, gkab670-.

Zhong, Y., Sani, H. M., Paudel, B. P., Low, J. K. K., Silva, A. P. G., Mueller, S., Deshpande, C., Panjikar, S., Reid, X. J., Bedward, M. J., et al. (2022). The role of auxiliary domains in modulating CHD4 activity suggests mechanistic commonality between enzyme families. *Nat Commun* 13, 7524.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C. and Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10, 1523.

Zhou, J., Hamdan, H., Yalamanchili, H. K., Pang, K., Pohodich, A. E., Lopez, J., Shao, Y., Oses-Prieto, J. A., Li, L., Kim, W., et al. (2022). Disruption of MeCP2-TCF20 complex underlies distinct neurodevelopmental disorders. *P Natl Acad Sci Usa* 119, e2119078119.

Zovkic, I. B., Paulukaitis, B. S., Day, J. J., Etikala, D. M. and Sweatt, J. D. (2014). Histone H2A.Z subunit exchange controls consolidation of recent and remote memory. *Nature* 515, 582–586.

Zwaag, B. van der, Franke, L., Poot, M., Hochstenbach, R., Spierenburg, H. A., Vorstman, J. A. S., Daalen, E. van, Jonge, M. V. de, Verbeek, N. E., Brilstra, E. H., et al. (2009). Gene-Network Analysis Identifies Susceptibility Genes Related to Glycobiology in Autism. *Plos One* 4, e5324.

Zweier, C., Sticht, H., Bijlsma, E. K., Clayton-Smith, J., Boonen, S. E., Fryer, A., Greally, M. T., Hoffmann, L., Hollander, N. S. den, Jongmans, M., et al. (2008). Further delineation of Pitt–Hopkins syndrome: phenotypic and genotypic description of 16 novel patients. *J Med Genet* 45, 738.

Żylicz, J. J. and Heard, E. (2020). Molecular Mechanisms of Facultative Heterochromatin Formation: An X-Chromosome Perspective. *Annu Rev Biochem* 89, 1–28.

9 Appendices

9.1 Data on the generation of Hela cells expressing GFP-PRTH

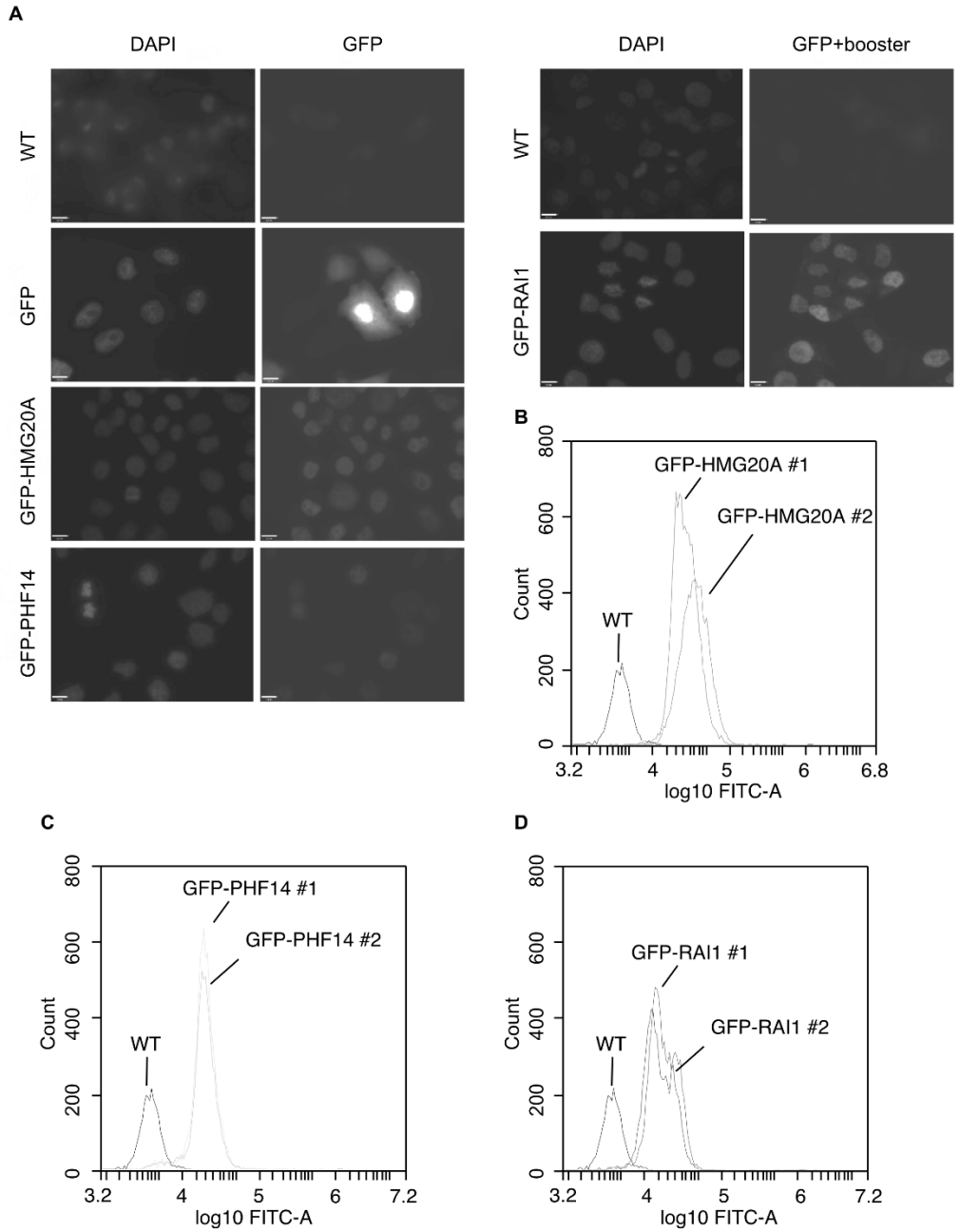


Figure App 1: GFP-PHF14, GFP-RAI1, and GFP-HMG20A localize to the nucleus in Hela cell lines
(A) Fluorescent microscopy images of Hela cells expressing GFP, GFP-PHF14, GFP-HMG20A, to visualize GFP-RAI, the addition of an Alexa 488 coupled nanobody (GFP-booster) was required Scale bar: 5 µm (B-D) Flow cytometry of

GFP-HMG20A (B), GFP-PHF14 (C) and GFP-RAI1 (D) expressing cell clones to assess the purity of cell populations.

9.2 HMG20A regulates neural crest and heart differentiation in *Xenopus laevis*

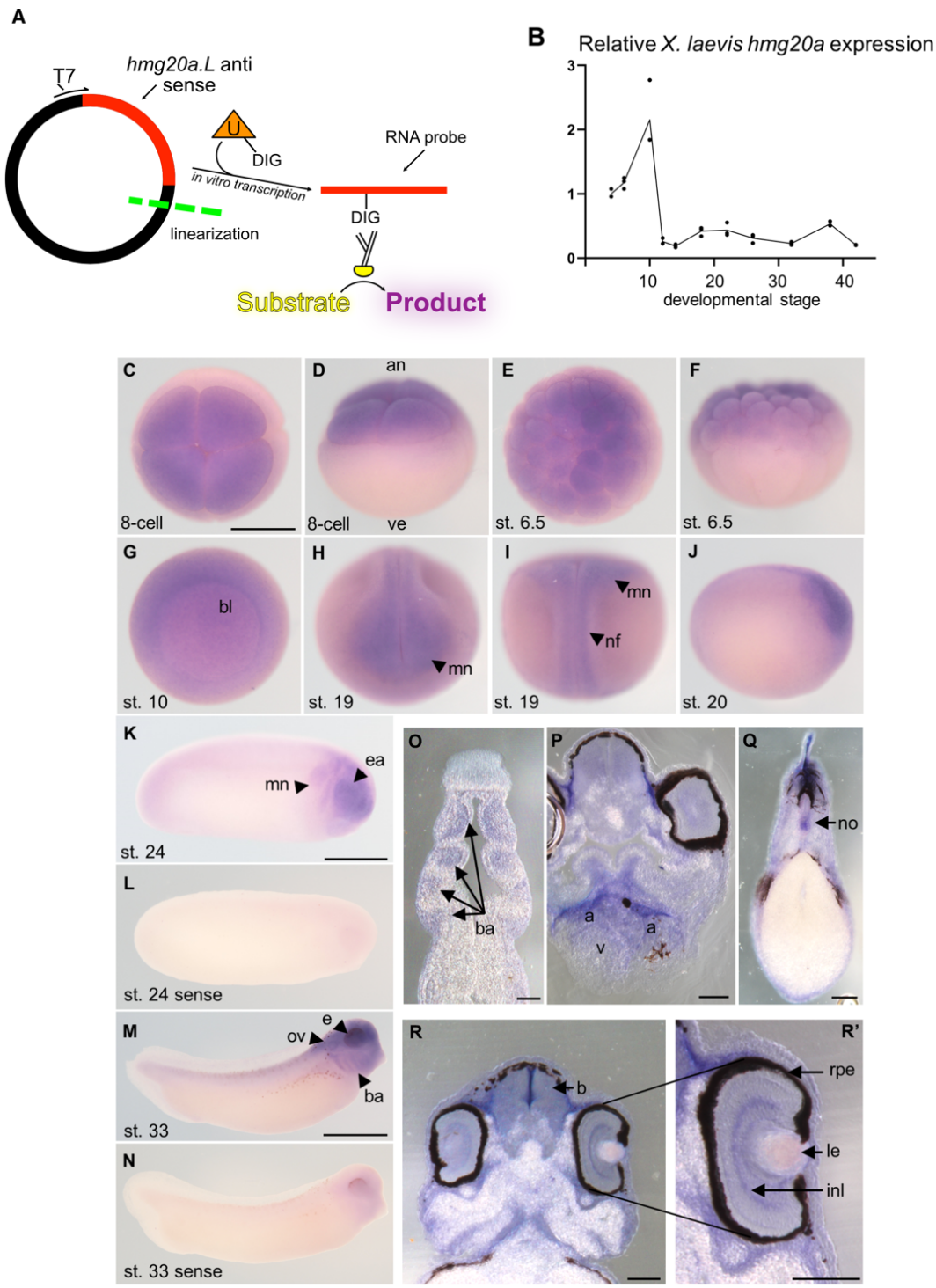


Figure App 2: The spatial expression pattern of hmg20A determined by whole mount hybridization in situ

(A)Hmg20a.L cDNA was cloned into the pDNA3 plasmid. DIG labeled uracil was incorporated into the probes by T7 in vitro transcription. The localization of probes in embryos was detected by an alkaline phosphatase-coupled anti digoxigenin antibody. (B) Temporal expression pattern of Xenopus hmg20a: RT-qPCR of hmg20a mRNA expression covering *Xenopus laevis* developmental stages 4 (8-cell stage) to 42 normalized to odc expression. The error bars indicate the s.e.m. of three technical replicates. hmg20A mRNA is detected in the early stages of *Xenopus laevis* development. (C) 8-cell embryo, anterior view. (D) 8-cell embryo, dorsolateral view, animal and vegetal pole are indicated. (E) Embryo at blastula stage 6.5. anterior view. (F) Same embryo as in E, dorsal view. (G) Embryo at gastrula stage 10. (H) Embryo at neurula stage 19, anterior view. (I) Same embryo as in H, dorsal view. (J) Embryo at stage 20, lateral view. (K) Embryo at stage 24, lateral view. (L) Sense control, embryo at stage 24. (M) Embryo at stage 33, lateral view. (N) Sense control, embryo at stage 33. The scale bar in C-N is 1mm. (O) Transverse section through the branchial arch region of a stage 31 embryo, expression of hmg20A in the branchial arches is indicated by arrows. (P-R') Transverse sections of a stage 42 embryo. (P) hmg20A is partially expressed in the heart region. (Q) Expression of hmg20A within the notochord (no). (R, R') hmg20A is partially expressed in the brain and the eye. The scale bar in O-R is 100 µm. abbreviations: a, atrium, animal; b, brain; ba, branchial arches; bl, blastoporus; ea, eye anlage; e, eye; inl, inner nuclear layer; le, lens; mn, migratory neural crest; nf, neural fold; no, notochord; ov, otic vesicle; rpe, retinal pigment epithelium v, ventricle; ve, vegetal.

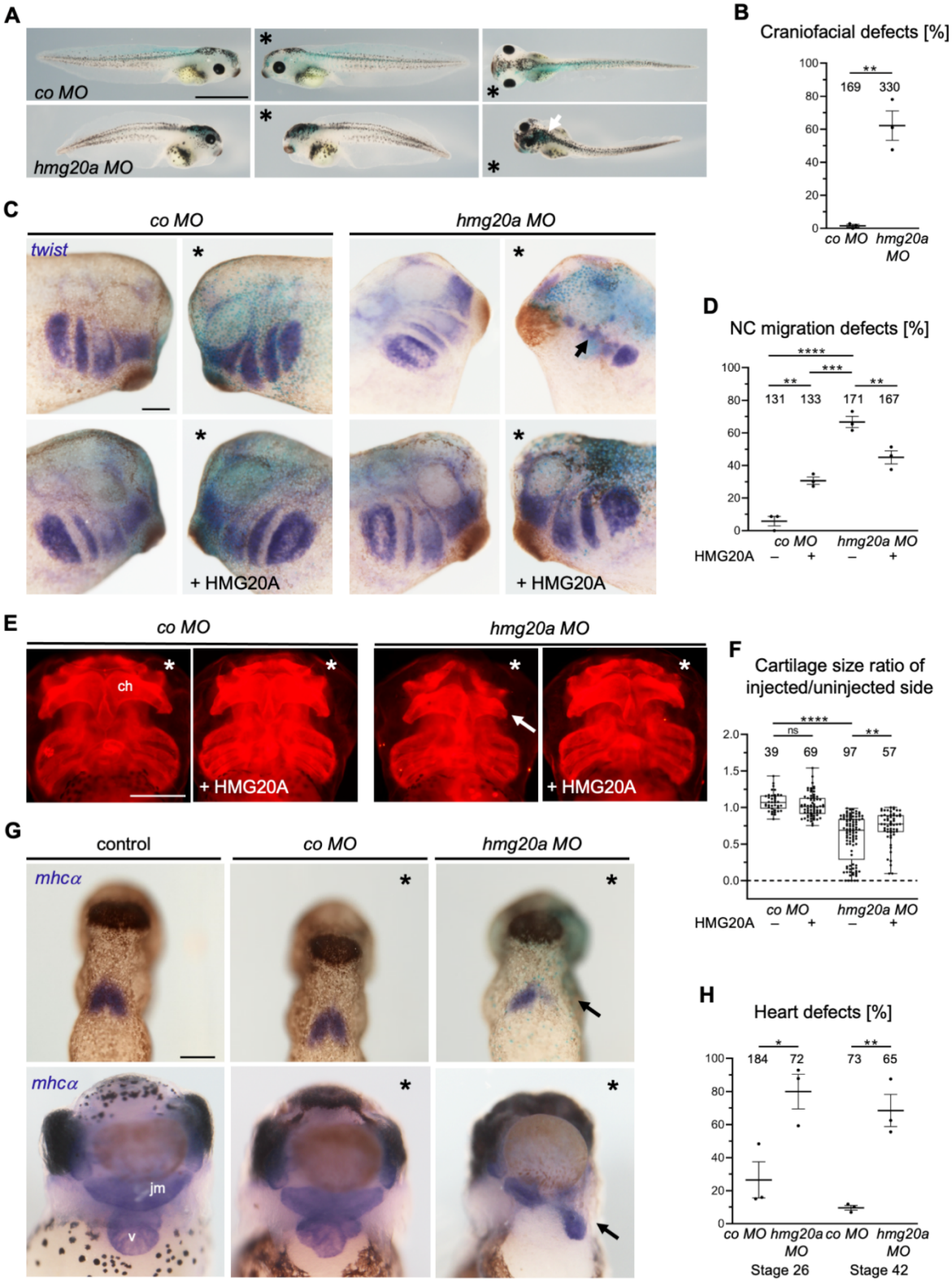


Figure App 3: HMG20A depletion leads to craniofacial and heart malformations in Xenopus laevis

(A) Loss of Hmg20a function leads to craniofacial and pigmentation defects in Xenopus tadpoles. Embryos were injected with 10 ng MO in combination with 80 pg of lacZ RNA in one blastomere at the two-cell stage, * marks the manipulated side, and the white arrow marks pigmentation defects. Scale bar = 1 mm. (B) Graph summarizing the mean percentage of craniofacial defects in

three independent experiments ± s.e.m. The number of embryos is indicated for each column. **p<0.01 (two-tailed unpaired Student's t-test). (C) The loss of function of Hmg20 caused defects in cranial NC migration that can be partially rescued by coinjection of human *HMG20A* cDNA. Embryos were injected with 10 ng MO in combination with 80 pg of lacZ RNA (seen in blue) and analyzed by twist in situ hybridization (seen in purple). For rescue experiments, 130 pg of human *HMG20A* cDNA was coinjected. * marks the manipulated side; the arrow shows the defect in cranial NC migration. Scale bar = 1mm. (D) Quantification of NC migration defects from three independent experiments as shown in (C). Data are presented as mean ± s.e.m., the number of embryos is indicated for each column. **p < 0.01, *** p < 0.001 **** p < 0.0001 (one- way ANOVA, Tukey's multiple comparisons test). (E) Xenopus tadpole embryos depleted with Hmg20a show defects in cartilage formation (arrow). Embryos were injected with 10 ng MO in combination with 80 pg of membraneRFP (mbRFP) RNA and analyzed by collagen II immunostaining. For rescue experiments, 100 pg of *HMG20A* cDNA was coinjected, * marking the manipulated side. Scale bar = 500 µm. (F) Box and whisker plots summarize cartilage defects of at least three independent experiments analysed as in (E) and quantified by measuring the area of the ceratohyale cartilage. The number of embryos (n, above each bar) and the median are indicated. The box extends from the 25th to the 75th percentile, with whiskers maximum at 1.5 IQR. **p < 0.01, **** p < 0.0001, ns.: not significant (one-way ANOVA, Tukey's multiple comparisons test). (G) The loss of function of Hmg20a causes heart defects. The embryos were injected as in (C) and analyzed by mhcα in situ hybridization. Top: The embryos depleted with Hmg20a at stage 26 show defects in the formation of the first heart field (arrow), while the controls are not affected. Bottom: At stage 42, Hmg20a depletion disrupts the three-chambered heart structure consisting of two atria (a) and a ventricle (v); the malformed heart is displaced toward the manipulated side. Jaw muscle (jm), which is also marked by mhcα, is also reduced in Hmg20a-depleted embryos. Scale bar = 1 mm. (H) Graph summarizing three independent experiments as shown in (G), data are presented as mean ± s.e.m., the number of embryos is indicated for each column. *p < 0.05 (two-tailed unpaired Student's t-test).

9.3 Vector maps and cloning strategies

9.3.1 *Cloning strategies applying restriction enzyme followed by DNA ligation*

Used backbone vectors and insert sequences are depicted in the followinging schematics of cloning strategies. These figures were generated with Snapgene *in silico* cloning:

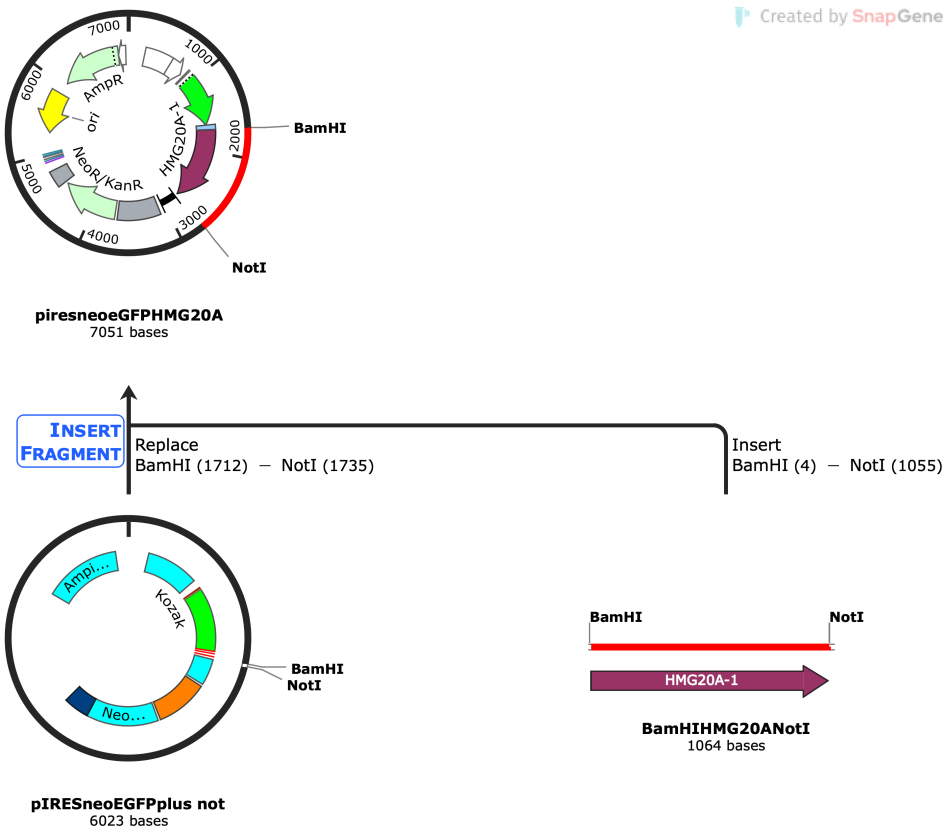


Figure App 4: Cloning strategy to generate GFP-HMG20A plasmids to stably overexpress it in human cells

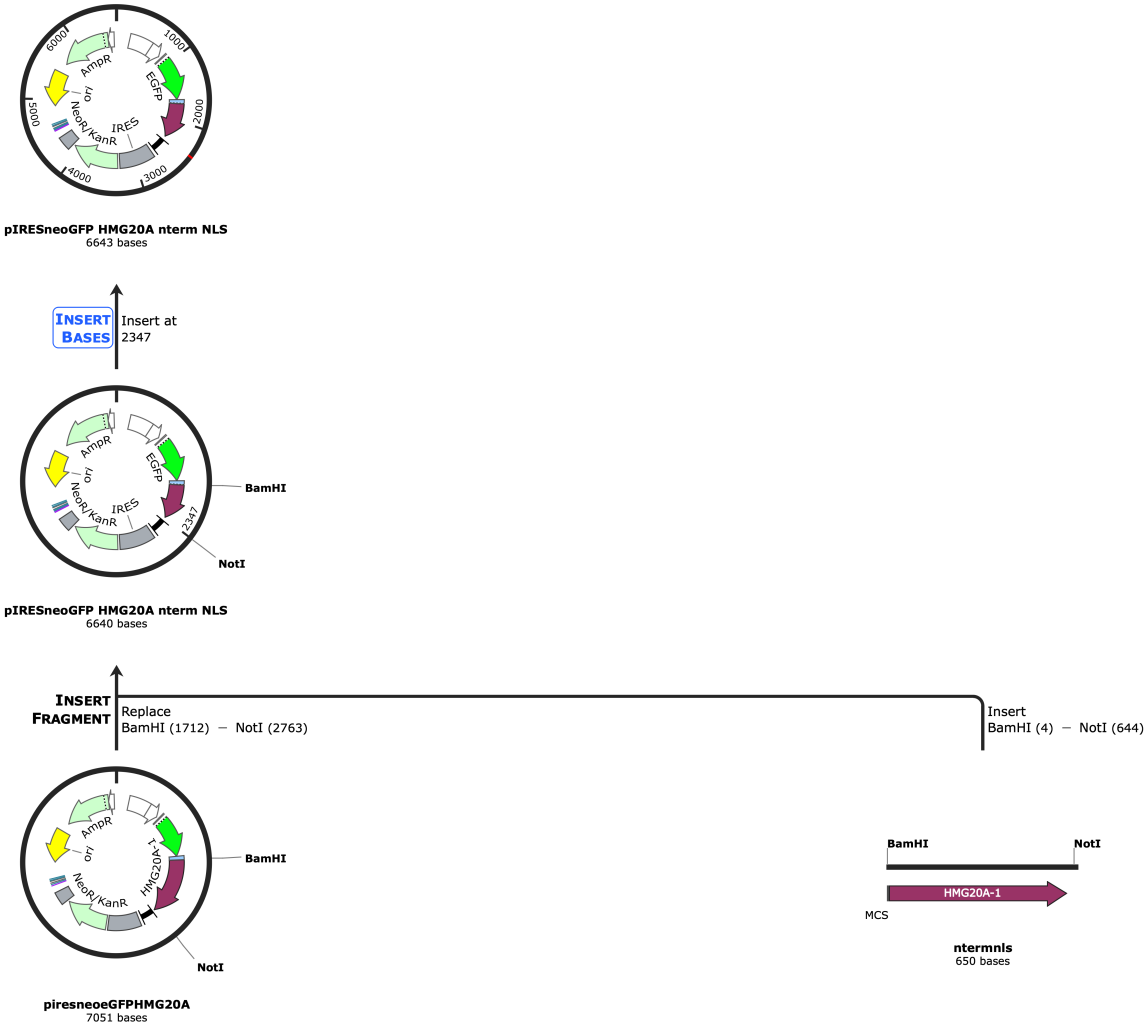


Figure App 5: Cloning strategy to generate GFP-HMG plasmids to stably overexpress it in human cells

To ensure proper nuclear localization a sequence coding for SV40-NLS was included at the 3'-end of the insert.

Figure App 6: Cloning strategy to generate GFP-CC plasmids to stably overexpress it in human cells
To ensure proper nuclear localization a sequence coding for SV40-NLS was included at the 3'-end of the insert.

To generate HMG20A expression vectors for Sf9 cells, pFastbac1 vectors were digested with BamHI and NotI restriction enzymes and ligated with the same insert sequences discussed above (Figure App 5 and Figure App 6):

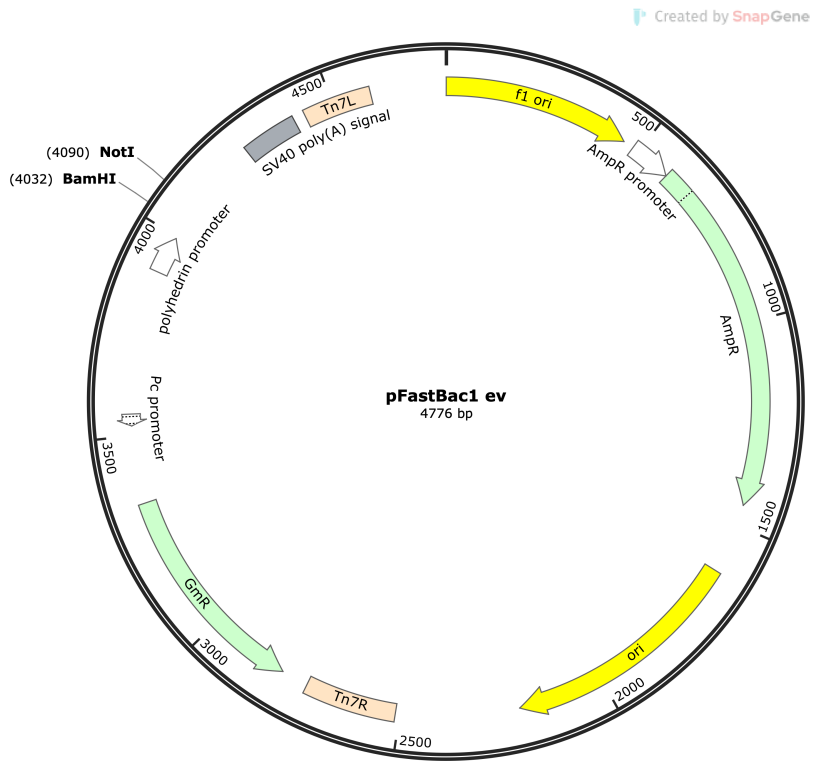


Figure App 7: Vector map of FastBac1

9.3.1.1 *Generation of gRNA and Cas9 expression vectors*

To generate locus specific Cas9 nucleases, px641 plasmids (Ran et al., 2013) were digested with BbsI (NEB). The linearized Vector was fused with hybridized oligonucleotides containing the needed gRNA sequence, flanked by the appropriate overhangs for ligation. For Oligosequence information see Table 7. Backbone of the Vector is depicted here:

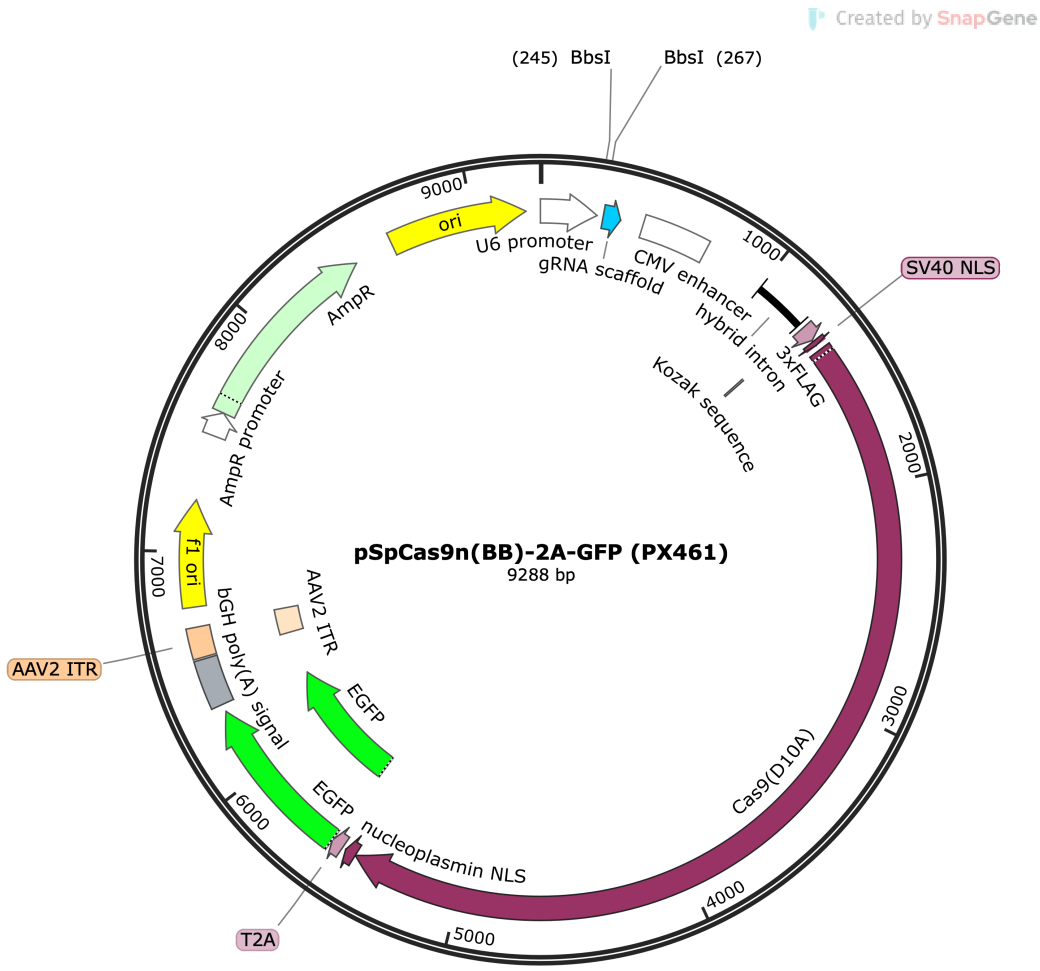


Figure App 8: Vector map of px461
Px461 was digested with BbsI restriction enzyme and re-ligated with appropriate oligonucleotide hybrids

9.3.2 *Recombination based cloning strategies*

To generate homology arms for recombination templates to knock out *Hmg20a* in mouse embryonic stem cells via integration of selection genes followed by transcriptional terminators inside of the *Hmg20a* locus, genomic DNA from mouse cells was integrated into puc19 vectors, in a way, that they flank synthetically generated puromycin or mCherry followed terminator sequences.

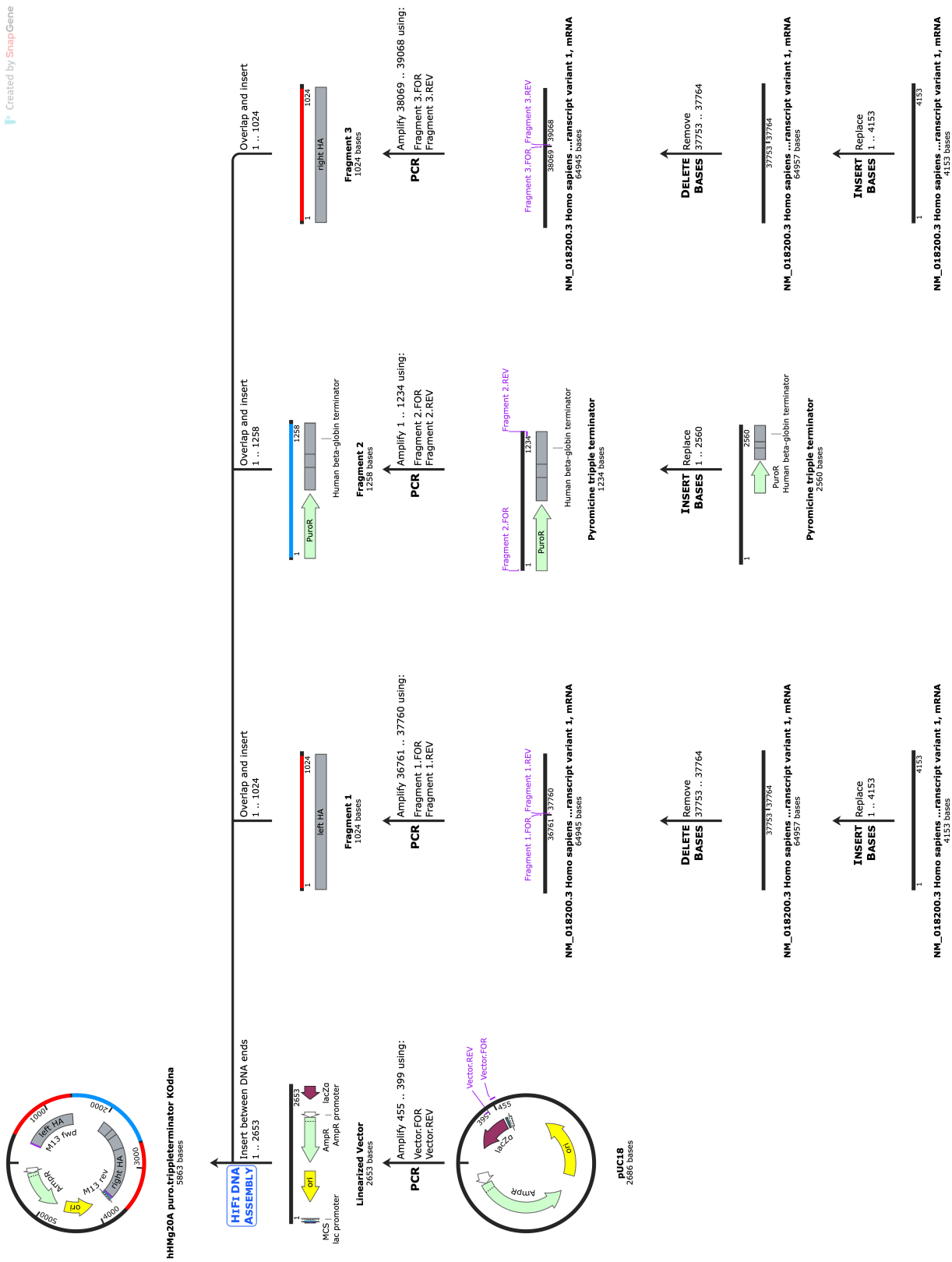


Figure App 9: Cloning strategy to generate Plasmid for puromycin selection of Hmg20a Knock out in murine cells

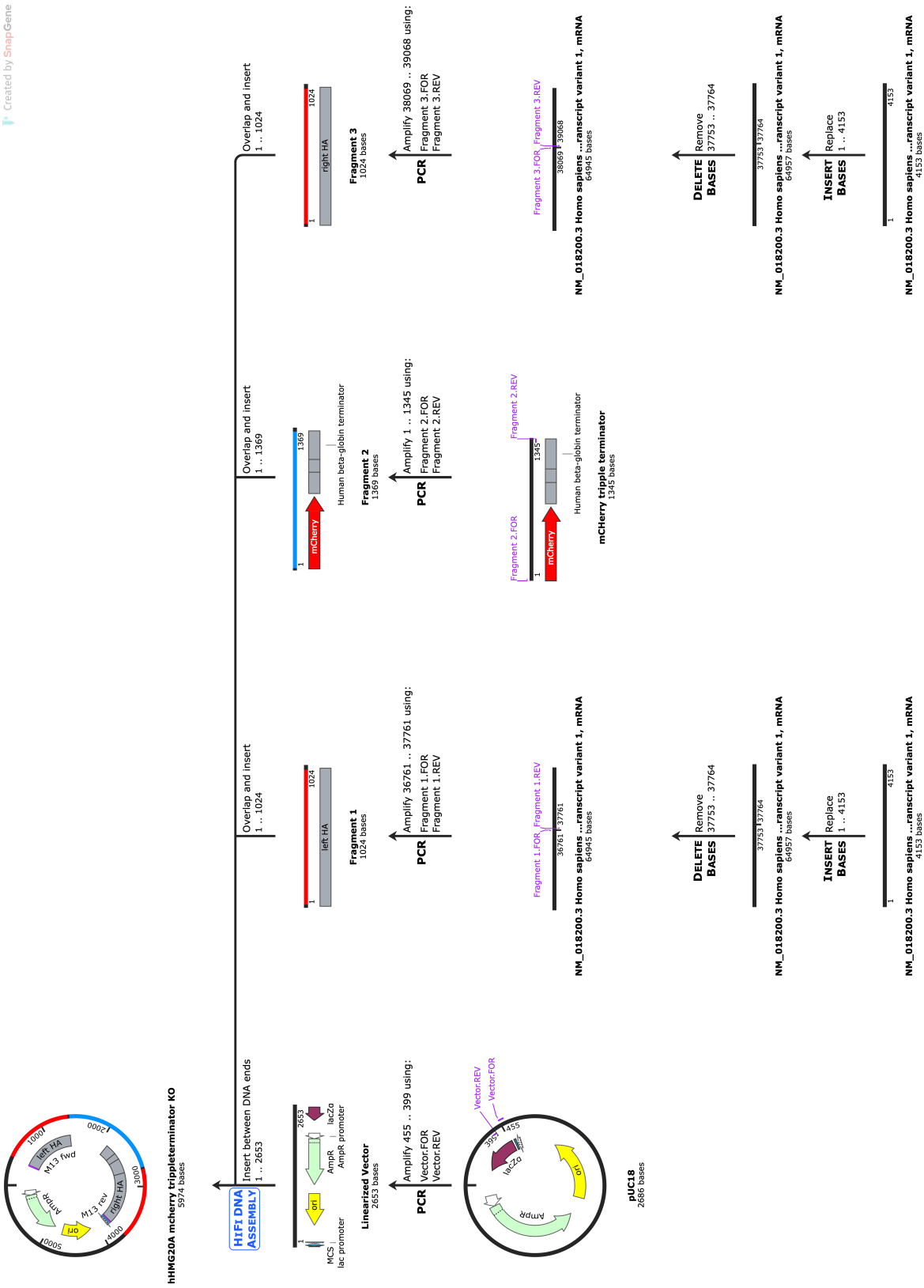


Figure App 10: Cloning strategy to generate Plasmid for mCherry fluorescence selection of Hmg20a Knock out in murine cells