Transcriptional regulation in the neural lineage

Martí Quevedo Calero

ISBN: 978-91-979021-7-5

Cover design and Layout: Martí Quevedo Calero
Printed by Original i Umeå AB, Sweden

The studies presented in this thesis were conducted at the Department of Cell Biology. Erasmus Medical Center (Rotterdam, the Netherlands) and financially supported by the NOW Graduate Programme Erasmus MC – Medical Genetics Grant: 022.004.002.

Transcriptional regulation in the neural lineage

Transcriptionele regulatie in de neurale lijn

Thesis

To obtain the degree of Doctor from the

Erasmus University Rotterdam by command of the rector magnificus

Prof. dr. H.A.P. pols

and in accordance with the decision of the Doctorate Board

The public defense shall be held on

Tuesday 19th June 2018 at 13.30

by

Martí Quevedo Calero

born in Barcelona, Catalunya (Spain)

DOCTORAL COMMITTEE

Promoters:	Prof.dr. F.G. Grosveld
	Prof.dr. D. Huylebroeck
Other members:	Dr. M.P. Creyghton
	Prof.dr. J.N.J. Philipsen
	Prof.dr. J.Gribnau
Co-promoter:	Dr. R.A. Poot

Als meus pares, per cristal·litzar el seu amor en mi

A la Marta, per catalitzar la meva felicitat

i a en Blai, pel dolç augment en entropia vital

To my parents for cristallizing their love in me

To Marta for catalyzing my happiness

and to Blai, for the sweet increase in vital entropy

TABLE OF CONTENTS

ABBREVIATIONS

A	Adenine
Å	Angstrom
ACH	Active chromatinhub
ATP	Adenosinetriphosphate
BDNF	Brain-derivedneurotrophic factor
Brd4	Bromodomain-containingprotein 4
bHLH	Basic helix-loop-helix
BMP	Bone morphogenic protein
C	Cytosine
ChIP-seq	Chromatin immunoprecipitation coupled to sequencing
CNS	Central nervous system
CORE	Clusters of open regulatory elements
CTD	Carboxy-terminal domain
Da	Dalton
DMV	DNA methylation valley
EB	Embryonic body
EM	Electron microscopy
ESC	Embryonic stem cell
eRNA	Enhancer RNA
EtBr	Ethidium bromide
FMR1	Fragile X mental retardation 1
G	Guanine
Gsc	Goosecoid
GSK3	Glycogen synthase kinase 3
HDAC	Histone deacethylase
HNF	Hepatic nuclear factor
HTH	Helix-turn-helix
H3K9me3	Histone 3 lysine 9 trimethylation
H3K27ac	Histone 3 lysine 27 acetylation
ICM	Inner cell mass
IDR	Intrinsic disordered region
iPSC	Induced pluripotent stem cell
DBD	DNA-binding domains
DNA	Deoxyribonucleic acid
Lac	Lactose
LIF	leukemia inhibitory factor
LCR	Locus Control Regions
LUCA	Latest universal common ancestor
MBD	Methyl-CpG-binding domain
MeCP2	Methyl-CpG-binding protein 2
MED	Mediator subunit
MEF	Mouse embryonic fibroblast
NASA	National Aeronautics and Space Administration
NE	Neuropithelium
NFI	Nuclear Factor I
ncRNA	non-coding RNA
NLR	nucleosome length repeat
NSC	Neural stem cell
PcG	Polycomb group

PIC	Preinitiation complex
PRE	Polycombreponse element
PTM	Post-translation modification
RA	Retinoic acid
RARa	Retinoic acid receptor alpha
RGC	Radial glial cell
RNA	Ribonucleic acid
RNApol2	RNA polymerase 2
SE	Super enhancer
SEC	Super elongation complex
SHH	Sonic hedgehog
shRNA	Short hairpin RNA
SNP	Single nucleotide polymorphism
SVZ	Subventricular zone
T	Thymine
TAD	Topologically associated domain
TF	Transcription factor
TRN	Transcription regulatory networks
TSS	Transcription start site
U	Uracil
VZ	Ventricular zone
Znf	Zinc finger
5hmC	5-hydroxymethylcytosine

SCOPE OF THE THESIS

The evolution of life can only be understood as the loom of more efficient ways to replicate genetic material. The development of intricate processes of gene regulation control has allowed the emergence of more elaborated life forms. In the first half of **Chapter 1** of this thesis, I introduce the basis of transcription regulation in the context of evolution. On the second half, I discuss how transcription regulation mechanisms can explain complex processes in animal development such as the formation of the brain.

Chapters 2 to **4** contain the experimental work performed during the course of my PhD studies where the general scope has been the application of state of the art biochemistry technologies to the field of transcription regulation and neurodevelopment. **Chapter 2** involves the study of the core transcription regulatory machinery. The Mediator complex has acted as a bridge for my neuroscience background to cross to the chromatin world, where we have expanded the general understanding of the Mediator interactome and its genomic localization. Moreover, it has provided new paths to explore in further research. One example is **Chapter 3**, where I characterize the role of Cggbp1, a Mediator-interacting transcription factor involved in neural commitment. Exploiting updated protocols of stem cells culture and differentiation, I could follow the role of Cggbp1 in the dynamic model of neural induction. Having seen the early neural induction events described in **Chapter 3** and to the more biochemistry focus studies in neural progenitors in **Chapter 2**; it seems fitting that **Chapter 4** of my thesis focuses in neurons, a terminal point of differentiation in the neural lineage. In this last chapter, we present an epigenetic map of the phenomenon of neuronal maturation, which is very important for neurons but surprisingly understudied. Also related to my experimental work are worth mentioning 2 other publications (reported in my CV at this thesis addendum) where I could contribute with the biochemistry skills learned during my PhD studies.

In the final Chapter of this thesis (**Chapter 5**), I summarize the results of the experimental research described in **Chapters 2** to **4**. In addition, I present preliminary experimental data of new potential projects.

In summary, we have combined biochemistry and molecular cell techniques with the study of neural development, providing notable contributions to the general understanding of how transcription is regulated but also discover new factors involved in transcriptional regulation.

Chapter 1

INTRODUCTION

Prologue

The science of biology is the study of life. Many disciplines, ranging from the morphological description and classification of species in the taxonomy field, to the study of organic chemical reactions seen in biochemistry, are branches of biology that each study living organisms.

But what do we call a *living* organism? What is *life*? These questions have been the subject of discussion between scientists for centuries and continues to arise as our knowledge and understanding expand[1]. The current concept of life may not be far from the one used by the National Aeronautics and Space Administration (NASA) as "a self-sustained chemical system capable of undergoing Darwinian evolution". However, the development of new research fields such as synthetic biology and artificial intelligence has made the boundaries of the definitions of life dimmed[2], and I would suggest an even more minimalistic idea such as "life is an evolving self-sustaining system".

The last short sentence presents two key concepts that will echo through this introduction. The first one, "self-sustention" refers to the autonomy of the replication of the organism, while the second one, "evolution", alludes to the transmission of heritable traits by the aforementioned Darwinian natural selection which could lead to the appearance of new organisms[3]. With these two pillars I will attempt to summarize my molecular neuroscience studies starting from the simplicity of the first life forms on earth towards the development of the brain, adding one layer of complexity at a time.

The origin of life and the central dogma of biology

After the formation of the Earth and the condensation of the oceans around 4 billion years ago[4], the first organic molecules started to appear from inorganic reactions with the energy from the sun and/or volcanic activity filling the oceans with a vast and chaotic spectrum of monomers and polymers in a stage termed "primordial soup"[5]. Several studies have demonstrated the production of both amino acids[6] and nucleotides[7], building blocks of modern life forms, under prebiotic reactions mimicked under laboratory conditions.

In this progressively changing chemical environment, complex molecules were created and destroyed continuously depending on the affinity of their small components and other favorable conditions such as compartmentalization within lipid vesicles and local concentrations at sea shores, crystals, ice sheets or metal precipitates at deep-sea vents[8–10]. There is an ongoing debate about the exact chemical nature of the first pre-living organism, but its main characteristic are clear, i.e. the ability to self-replicate[11]. In other words, the origin of life would have been molecules with the capacity to catalyze the chemical reactions needed to create a new molecule with the same capacity and as result, preserving and propagating its identity in the anarchical mix of reactions ongoing in the primordial soup. To accomplish this, in addition to the catalytic activity per se, which many other molecules perhaps had acquired before, the first life forms would have been the template themselves, carrying a piece of information (what we know today as genes) needed to assemble the right components to carry out the replication reaction[12].

It is important to notice that an exact copy of the molecule would not have been essential to continue the replication chain as long as its "daughter" had retained the same ability. Therefore, several variants of the original structure would have appeared (alleles) and coexisted. Some of these molecules would have changed their template (genotype) so much that they had developed differential traits (phenotype) such as the speed of replication, the accuracy on reading the template, their structural stability, or even starting to carry information for several genes. In short, the evolution arms race would have started, as one type of molecule would have tried to outcompete the others and/or compartmentalize it in each cell of multi-cellular organisms, as a more capable system to transmit its information.

From this ancestor "replication" wars, one particular system rose victorious. The power of this design relayed on a first very stable molecule that could contain several genes, individually encoding into smaller templates which at the end would be processed to produce the functional molecules[13]. This design was so fitting that became the base of all other subsequent living forms until now and constitutes the central dogma of biology, which explains the flow of genetic information within a biological system[14] (Figure 1).

The ultimate molecule then comes as deoxyribonucleic acid (DNA)[15] and usually consists of two antiparallel polymer strands coiled around each other to form a double helix[16,17]. The monomer building blocks are nucleotides, each of them composed by three molecules: a five-carbon sugar (deoxyribose in the case of DNA), at least one phosphate group and one of four nitrogenous bases, guanine (G), adenine (A), cytosine (C), or thymine (T).The order in which these four bases are stacked to form the strand formulates a letter code[18] (Figure 2). Why there is a four letter only alphabet in DNA is intriguing as some laboratories have accomplished to synthesize and coordinate new artificial bases into native DNA and it suggested to be a hint to the first choices in replication evolution[19].

In addition, in the double strand selective pairing (or complementarity) between the letters occurs, where the purine bases G and A form 3 and 2 hydrogen bonds to the pyrimidine bases C and T, respectively. The duplication of the information in two strands provides a very stable and at the same time direct way to replicate the molecule as a single strand can become the template for the synthesis of the complementary counterpart[20].

Through the process called transcription, the letters in the DNA encode for the formation of a very similar yet different molecule, the ribonucleic acid (RNA) by a protein named RNA polymerase[21]. RNA molecules tend to be single-stranded; their nucleotides contain ribose, as well as the same bases except thymine that is replaced by uracil (U). Despite single-stranded RNA being less stable, its properties allow it to gain specific activities[22]. For example, RNA molecules can have catalytic properties representing a paradigm where genotype and phenotype are found in the same molecule. In fact, a strongly supported theory suggests RNA molecules as the true origin of life, before DNA was established as the main genotype carrier[23].

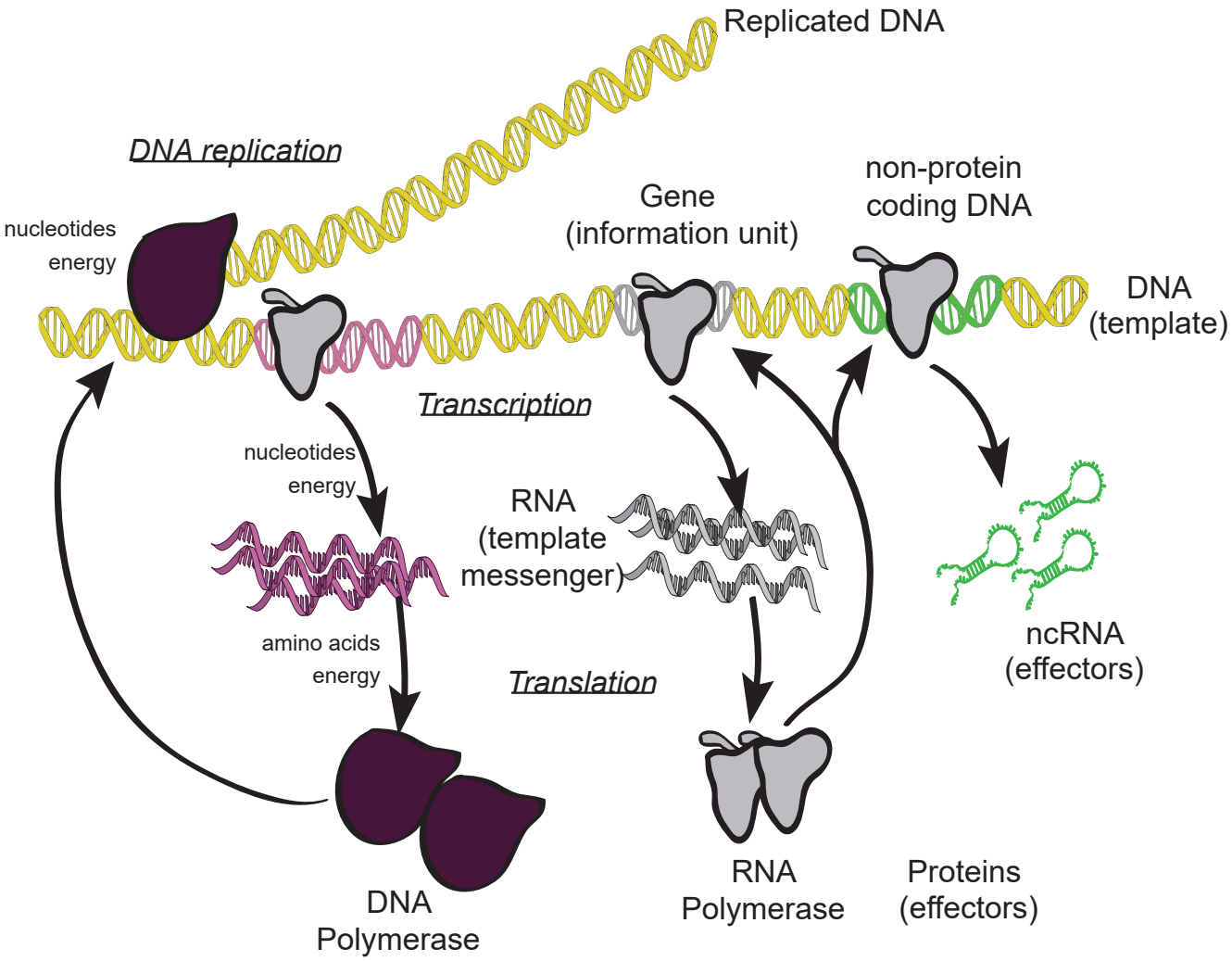


Figure 1 | The central dogma of biology

The normal flow of biological information: DNA can be copied to DNA (DNA replication), DNA information can be transferred into mRNA (transcription), and proteins can be synthesized using the information in mRNA as a template (translation). Proteins are the major effectors in all steps.

Next, RNA is used as an intermediate template to produce the final step on the process, the synthesis of proteins[24]. These polymers are different from DNA and RNA as they are composed of amino acids. It is in this last step, that the 4-letter alphabet is translated, whereby 3 letters are read in combination into 1 out of 22 possible amino acids; different 3-letter combinations can result in the same amino acid (the concept of degeneration of the code), and 1 and 3 of such combinations provide a signal to the system for respectively starting and stopping the incorporation of amino acids during synthesis of proteins.

Amino acids differentiate between one another by their side chains, which contain specific atom groups and grant them different chemical properties. Again, the sequence into which amino acids are arranged in the polypeptide (called primary structure) will be essential as the order of the side chains will determine in turn local interactions between amino acids (secondary structure), the overall

folding of the protein (tertiary structure) and even the formation of multi-protein complexes (quarternary structure). The preeminent characteristic of proteins is the expansion of the complexity of structural conformation possibilities (also called protein domains) by going from the 4 nucleotides to 22 amino acid combinations.

Part I. Evolving genes, evolving transcription

As hinted before, the DNA of the latest universal common ancestor (LUCA) of modern organisms must have carried several genes to code for the proteins needed to carry out the essential process of replication. Probably, either by copying mistakes, by the integration of viral genomes by horizontal gene transfer[25] or by (sometimes incomplete) genome duplication and conversion during evolution, it would have also developed extra genes coding for proteins that had become beneficial in their primitive environment. Examples would be those regulating compartmentalization, the caption of nutrients across lipid-based membranes, the metabolism of some molecules for either energy or limiting intermediate substrates, etc[26].

Promoters and general transcription factors

One of the first mechanisms that appeared to regulate transcription would have been promoters. Even the genes of "simple" bacteria contain combinations of DNA sequence elements (*cis*-regulatory elements) pointing to the transcription start point (TSS). Close-proximity short sequences serve to recruit DNA recognition proteins (*trans*-regulatory elements) such as general transcription factors, which aid the RNA polymerase to initiate transcription. While in bacteria this function is restricted to variants of a single protein (σ factor[28]), the intricacy of transcription increased in more complex organisms (such as eukaryotes). Notably, promoter modularity expanded both in sequences forming the core promoter and in general transcription factors[29].

While storing information into one single-strand of DNA to expand protein functionality would have been helpful, a selective transcription of each of the genes would also have been favored. Hence, the first mechanisms of transcription would have appeared at the beginning especially to maximize resources and separate pieces of information, but later to coordinate responses to signals[27].

Transcription activators and repressors

An early evolutionary innovation in transcription was the ability to switch on or off some of the functions of the organism. Thus, some *trans*-regulatory elements evolved to recognize and bind specific sequences (or DNA motifs) to select which genes to regulate. By definition, transcription factors (TFs) are proteins containing one or more DNA-binding domains (DBD) and they are often classified based on sequence similarity, structural folding of their DBDs and/or the DNA sequence they bind to. Examples of transcription factor families are basic helix-loop-helix (bHLH) factors, characterized by a motif of two α-helices (one of them with basic aminoacids) connected by a loop; zing finger (Znf) factors, which contain multiple finger-like protrusions that make contacts with their

target nucleic acid using zinc or other metals ions to stabilize their folding; or homodomain factors, composed of three alpha helixes, with helices 2 and 3 forming a helix-turn-helix (HTH) structure[30].

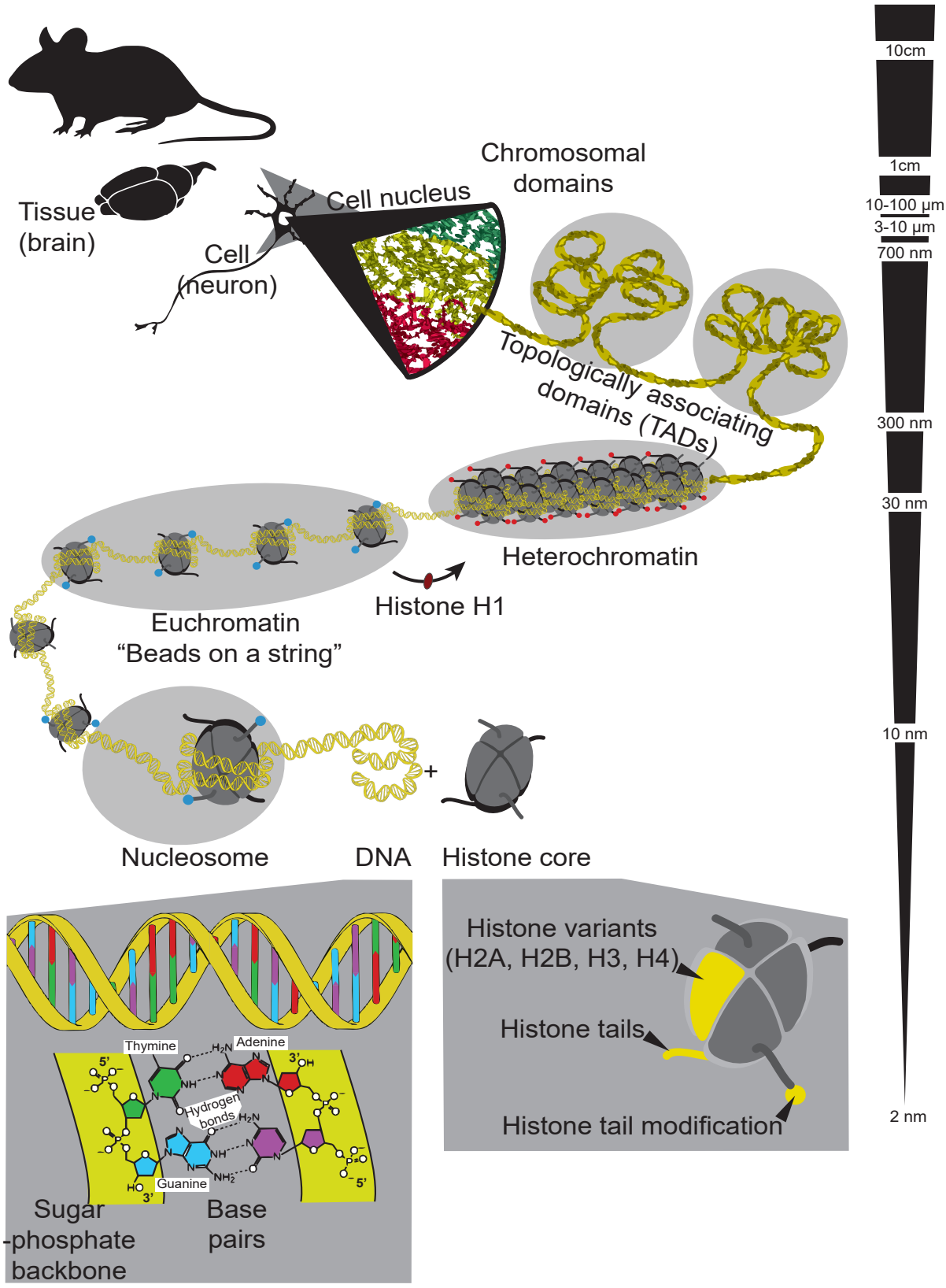


Figure 2 | Range of characteristic sizes of the compaction states of DNA

TFs can either activate or repress transcription, depending on their protein functionality, allowing the modulation of transcriptional output in response to certain extracellular cues linked to cascades of intracellular signals. One of the first described examples is the bacterial lactose (*Lac*) operon, a two-part control mechanism ensuring that the proteins involved in lactose metabolism are only expressed when lactose is available (by constitutive action of a repressor) and there is not a better source of energy such as glucose (activator dependent on a signal molecule)[31,32].

As I will mention below, the increases in the number of regulatory proteins in general and of TFs in particular is connected to phenotypic innovations and the evolution of more complex unicellular and multicellular organisms.

Histones, chromatin and DNA remodeling complexes

An inherent problem in the process of scaling up information storage is the limitation in structure, size and space-time accessibility[33]. As good as the retention of increasing numbers of genes may seem, there is a functional limit where it may become impractical to have an extreme long strand of DNA, roaming inside the cell (even if looping and coiling of naked DNA occurs[34]). In addition, too much available information at the same time makes coordination of genome control difficult[25,35]. Thus, another already early evolutionary innovation in gene regulation was the targeted compaction of DNA. Achaea, a microorganism phylum originally classified as bacteria, differs from the latter by containing innovative genes, especially in the regulation of transcription and translation. Among them, several Achaea have genes encoding for histones, positively charged proteins with the ability to interact with the negatively charged DNA, and fold it around them[34,36].

The term chromatin refers to the DNA in coordination with other molecules such as proteins or RNA. In more modern life forms called eukaryotes, the nucleosome is the basic unit of chromatin packaging[37,38]. It consists of a histone octamer core composed by two times the histone proteins H2A, H2B, H3 and H4 that wraps 146 base pairs plus a short DNA segment as linker[39]. Short-range interactions within an array of nucleosomes, mediated in part by the addition of histone H1, form chromatin fibers. Further compaction (by supposed rosettes) would organize the whole strand of DNA into (each of) the compacted chromosome(s) (Figure 2).

However, chromatin compaction may very well not have emerged to solve the size limitation in the evolutionary process of genomic scaling, but rather – and more importantly to add a layer of gene regulation by dictating which parts of the DNA are accessible to being transcribed. As a consequence, together with the aforementioned nucleosome-based packaging, new regulatory components for efficiently defining the compaction state of specific genomic regions would have emerged. Through evolution, a wide collection of these chromatin remodelers would arise following the different strategies to regulate chromatin. For example, in eukaryotes several groups of proteins act together, i.e. in complexes, which in an ATP-dependent fashion remove, slide or exchange nucleosomes[40].

Although some of these processes and the needed components were already present in Achaea, the real expansion of chromatin remodeling began with the evolution of histones themselves, in

particular by means of providing them with a protruding tail that are accessible on the outside the nucleosome[41]. These histone tails mediate inter-nucleosome interactions, but more importantly serve as more accessible sites to biochemically modify the nucleosome. New chromatin remodelers with enzymatic activity that catalyzes the covalent modification of polypeptides (post-translation modifications, PTMs) provide a means to alter the physico-chemical properties of histone tails depending on the groups conjugated to them and the position of the modification[42]. For example, adding acetyl groups to H3 at its lysine 27 (H3K27ac) would negate positive charges on histones, thereby disrupting the attraction of H3 to DNA[43]. As a consequence, these more relaxed chromatin regions (referred to as euchromatin) are more accessible for transcription to take place. Heterochromatin[44] on the other hand refers to more densely packed regions of DNA which commonly associate with silencing of gene transcription , involving tri-methylation of H3 at its lysine 9 (H3K9me3)[45].

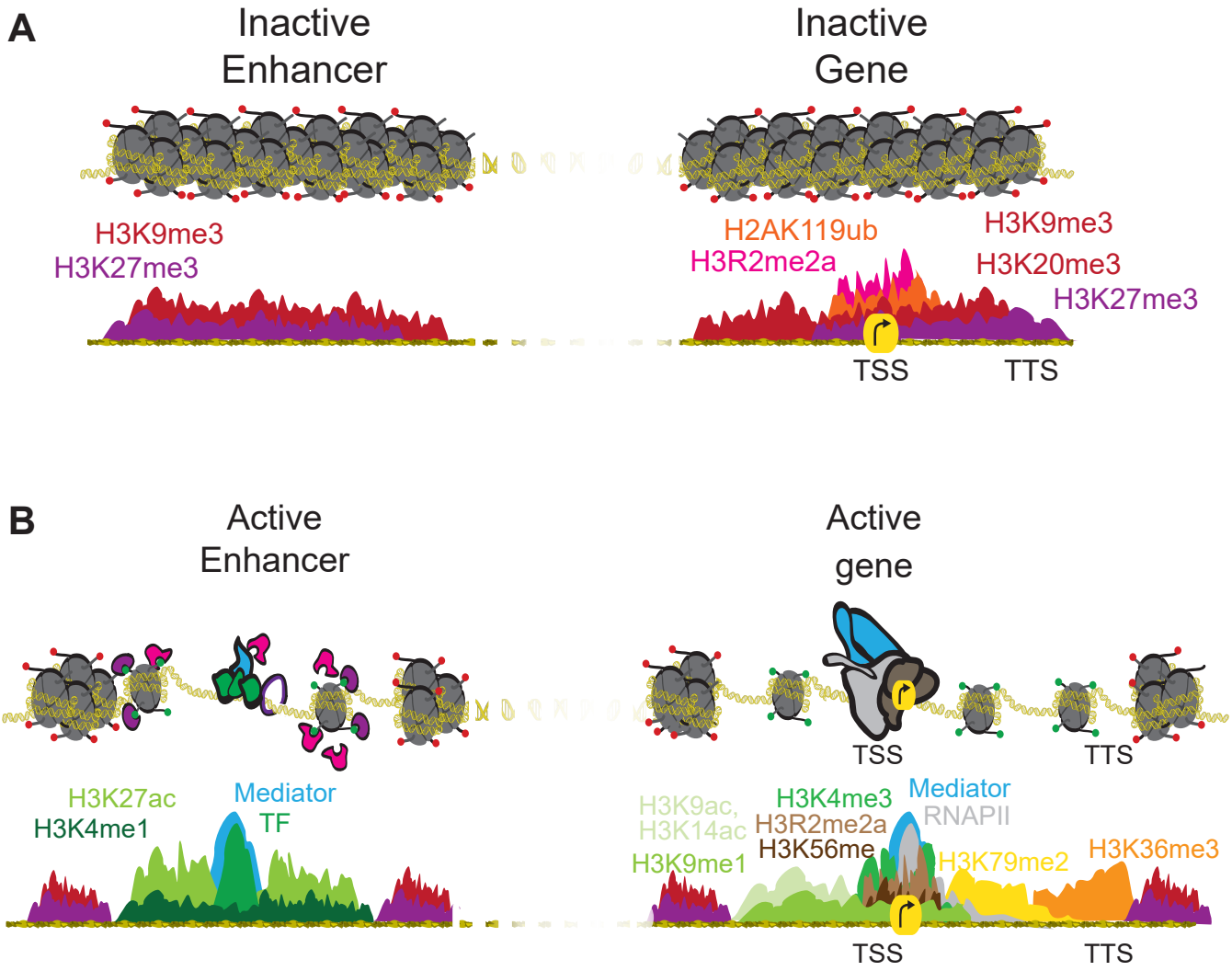


Figure 3 | Epigenetic landscape of (A) heterochromatin and (B) euchromatin

The palette of PTMs at histone tails, or histone code, co-determines the activity of a certain genomic region[46]. Histone modifications, together with nucleosome remodeling, DNA methylation and various non-coding RNAs constitute the major mechanisms of alter gene expression without altering the DNA sequence itself. The study of the heritable transmission of such modulations to daughter cells isknown as epigenetics[47].

In addition to the regulation by chromatin remodelers (acting as writers or erasers), other proteins have evolved to read these histone modifications. Thus, their function became dependent not on the DNA sequence itself, but to previous actionof writers/erasers.Therefore, they outreach to all regions with a certain type of modification[42] (Table 1).

Modification	Histone	Position	Enzyme	Function in transcription
Methylation	H3	K4	Mll1-4, Set1A,b	Activation
		K9	Suv39h, G9a, HMTase I, ESET, SETBD1	Activation (me1), Repression (me3)
		K27	E(Z)	Activation (me1), Repression (me3)
		K36	HYPB, Smyd2, NSD1	Activation and internal gene initiation repression
		K79	Dot1L	Activation (me1, me2, me3)
	H4	K20	PR-Set7, SET8	Activation (me1), Repression (me3)
Acetylation	H3	K27	CBP	Activation
		K56	Asf1+Rtt109	Activation
	H4	K16	hMOF	Activation
	H2A.Z	K14	SAGA (yeast)	Activation
Argenine Methylation	H3	R2 (Asymmetric)	PRMT6	Repression
		R8 (Symmetric)	PRMT5	Repression
		R17 (Asymmetric)	PRMT4 (Carm1)	Activation
		R26 (Asymmetric)	PRMT4 (Carm1)	Activation
	H4	R3 (Symmetric)	PRMT5, PRMT7 (mono me)	Repression
		R3 (Asymmetric)	PRMT1, PRMT6	Activation
	H2A	R3 (Symmetric)	PRMT5, PRMT7 (mono me)	Repression
		R3 (Asymmetric)	PRMT1, PRMT6	Activation
Phosphorylation	H3	S10	Snf1 (yeast)	Activation
Ubiquitination	H2A	K119	hPRC1L	Repression
	H2B	K120	UbcH6, RNF20/40	Activation

Table 1 | Histone post-translation modificationsand enzymes and their main role in gene transcription

Multicellularity and long-range regulatory elements

The increase in response flexibility evolved by transcription regulatory networks(TRNs), which allowed ancestral unicellular organisms to develop the ability to induce different life cycle states tobetter adapt to changes in the environment. In order to orchestrate these radical changes in phenotype via using differential gene expression, their DNA sequence had to incorporate more *cis*-regulatory elements. In other words, one cell would contain the information to transform to a different type as well as the switches to activate this transition.

At the same time, either by selection forces (i.e. predation and the limitation of nutrients, increased global oxygen levels among others[48]) some unicellular organisms, after being replicated, would behave as colonies by starting to aggregate, thereby creating the first multicellular organisms (metazoans)[49,50].

Until this point, the percentage of protein-encoding sequences in the genome of the first life forms was quiet high, fostering "useless" sections of DNA between genes may very well have been disadvantageous[27]. With the innovation of different cellular states and multicellular coordination, more *cis*-regulatory elements would then have to be developed, driving genome expansion. Hence, not only promoters, but pieces further away from the transcription start site (TSS) would then play a major role in gene regulation[50].

Enhancers or silencers are short regions of the genome that contain specific motifs for transcription factors and act as activation/repression switches for a gene (or sometimes a set of genes)that can be located over great distance[51]. As seen before in **Figure 3**, such elements are characterized by certain chromatin features. Although promoters have been recently suggested to act as long-range enhancers for other genes[52], enhancers are the main distant regulatory elements that have expanded the spatio-temporal transcription potential of the genome. This extension further allowed the evolution of more complex organisms, multiplying the number of cell types and developmental steps. As a result, the genome of these organisms started to be filled by non-protein coding regions, many acting as *cis* regulatory elements, to the point that the actual protein coding sequence only accounts for less than the 10% of the total DNA sequence in humans[50].

Chromatin loops, topology domains and insulators

As mentioned above, DNA within the cell is compacted and folded. Even in bacteria, there are proteins involved in the folding and coiling of the DNA, creating a consistent and organized genomic architecture[53].

However, with the introduction of long-range regulatory elements, the conformation of the chromatin ceases to play a structural role only, but also starts to actively function as a further layer of transcriptional regulation. Besides the topologically associated domains (TADs; also seen in bacteria), which are static regions within the DNA where contacts are frequent, a great number of dynamic loops within TADs are observed in eukaryotes[54,55]. In fact, this observation consolidated one of the

proposed models on how enhancers influence transcription at far distance, by looping into the promoter region[56].

Nonetheless, chromatin looping not only facilitates contacts; it can also negate the reach of enhancers or silencers by isolating them in a different domain. This phenomenon is mediated by insulators, with proteins such as CTCF that serve to set boundaries between genomic domains. Insulators can also act as barriers separating and stabilizing different chromatin states[57].

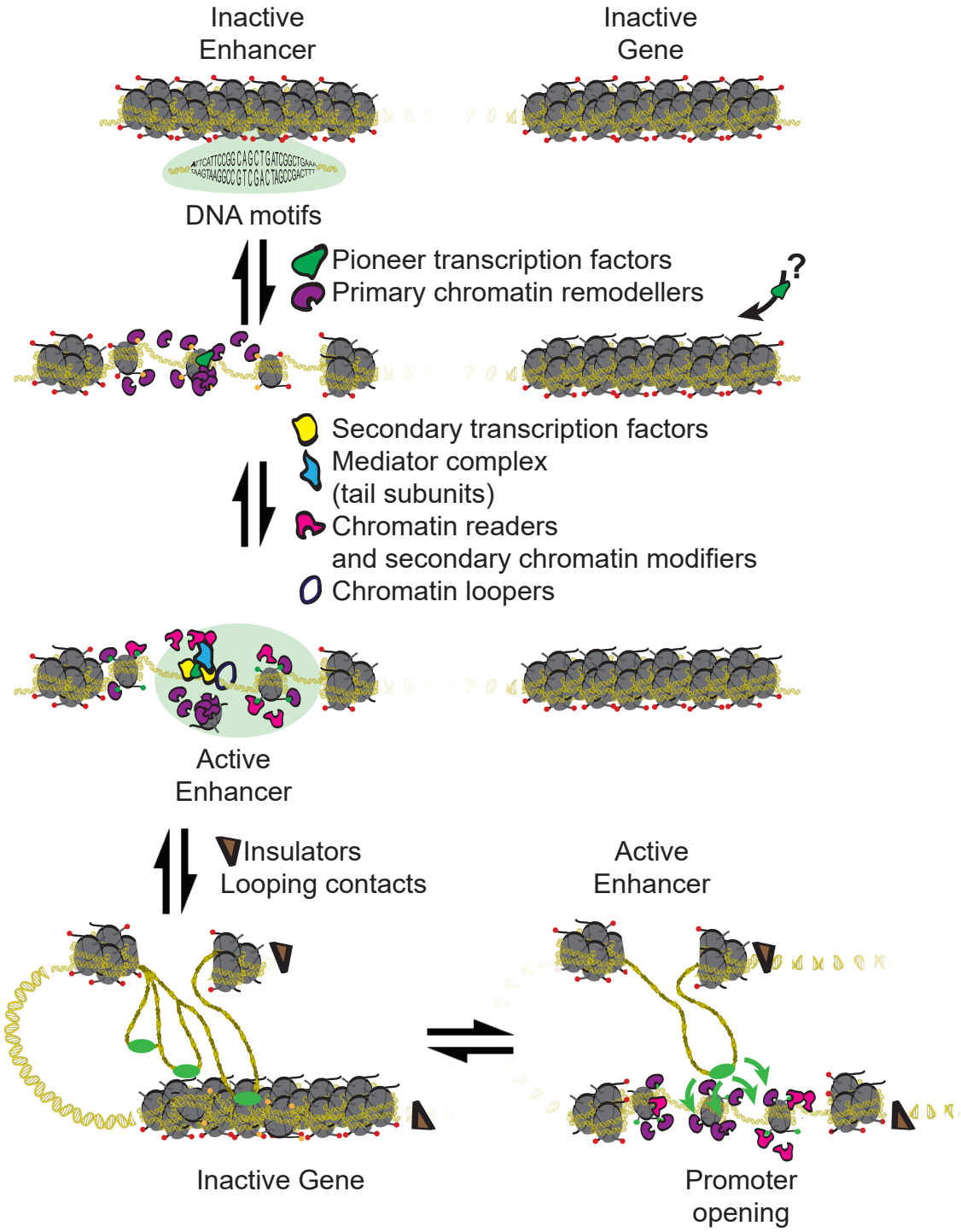
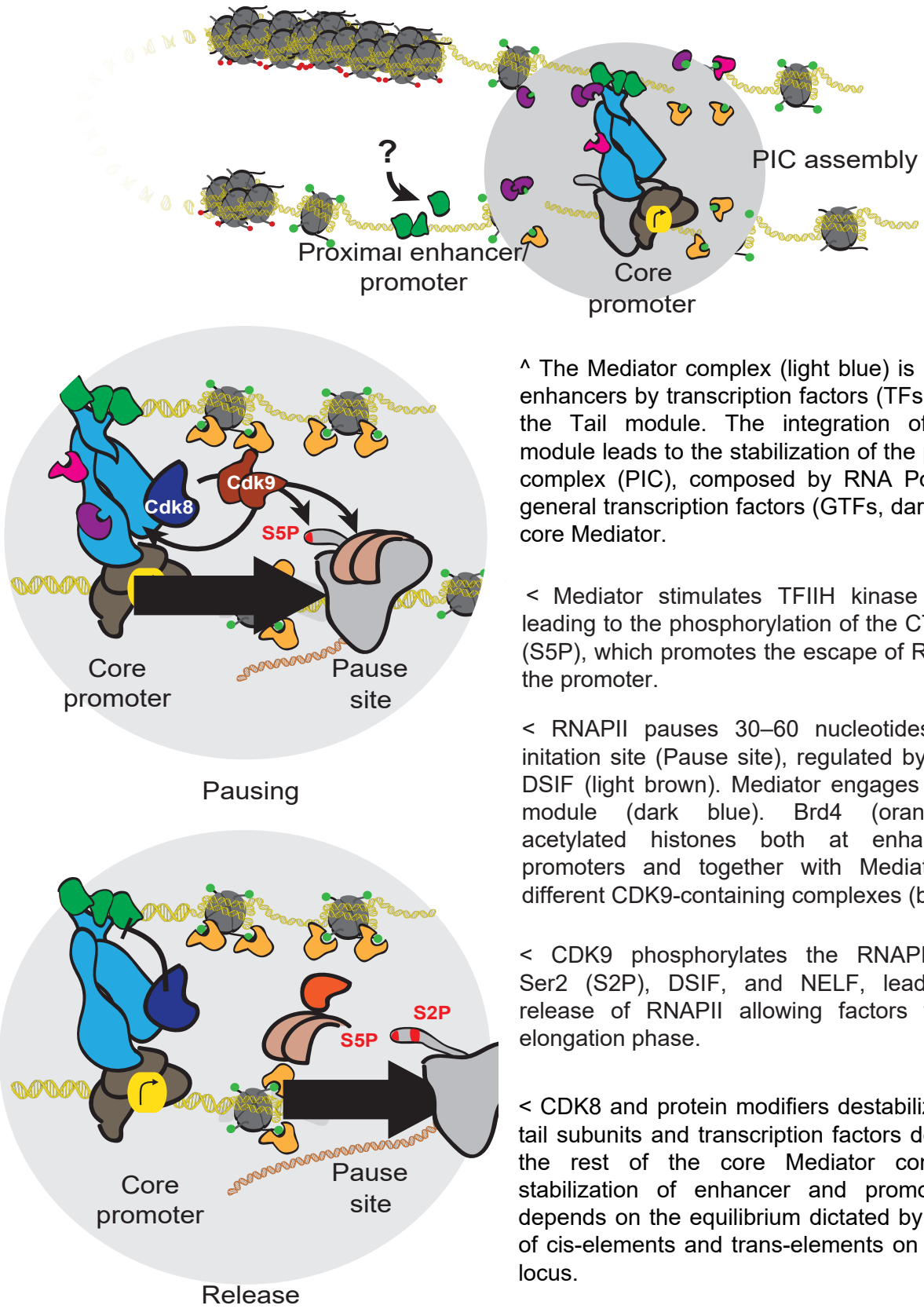


Figure 4 | Enhancer activation and promoter recognition



^ The Mediator complex (light blue) is recruited to enhancers by transcription factors (TFs, green) via the Tail module. The integration of the core module leads to the stabilization of the preinitiation complex (PIC), composed by RNA Pol II (grey), general transcription factors (GTFs, dark grey) and core Mediator.

< Mediator stimulates TFIIH kinase (a GTFs), leading to the phosphorylation of the CTD on Ser5 (S5P), which promotes the escape of RNAPII from the promoter.

< RNAPII pauses 30–60 nucleotides after the initation site (Pause site), regulated by NELF and DSIF (light brown). Mediator engages the Kinase module (dark blue). Brd4 (orange) binds acetylated histones both at enhancers and promoters and together with Mediator recruits different CDK9-containing complexes (brown).

< CDK9 phosphorylates the RNAPII CTD on Ser2 (S2P), DSIF, and NELF, leading to the release of RNAPII allowing factors for RNAPII elongation phase.

< CDK8 and protein modifiers destabilize Mediator tail subunits and transcription factors de-anchoring the rest of the core Mediator complex. Re-stabilization of enhancer and promoter factors depends on the equilibrium dictated by the density of cis-elements and trans-elements on the specific locus.

Figure 5 | Steps of transcription in an active promoter

Box 1. Chromatin regulatory macro-domains

The need to coordinate complex genetic programs and the expansion in chromatin regulation mechanisms lead to the appearance of chromatin macro-domains, large regulatory modules involved in the fine-tuning of complex transcriptional output. Some of these prominent regulatory structures have been known for a long time. However, with the development of new epigenetic techniques and bioinformatic analysis, new regulatory features are being discovered along with their mechanisms of action and their role in global transcriptional regulation. Some of these "special" chromatin features include:

DNA methylation valleys (DMVs): large regions (>3 kb) devoid of methylation that are often located in proximity to promoters of early developmental genes. Part of their mechanism of action is based on a rich GC content and a high association with special chromatin regulatory complexes, such as Polycomb type repressors[58].

H3K4me3 broad domains: regions among the top-5% domains with broadest H3K4me3 span. Mainly associated to promoters, they present high levels of paused RNA polymerase, which correlates with low transcription variability. In addition, these broad domains are relevant to to cell identity genes (factors required to establish and maintain the cell lineage)[59].

Locus Control Regions (LCRs): a combination of regulatory elements, mainly enhancers, that are capable of activating an entire gene locus even when placed in a totally different position in the genome. The first LCR to be identified was in the β-globin locus[60].

Clusters of open regulatory elements (COREs), stretch enhancers and super enhancers: the same as LCRs, these clusters of enhancers were identified by independent groups using genome-wide approaches. Back in 2011, using a combination of DNaseI and FAIRE sequencing approaches, COREs were identified in 7 different cell types[61]. Gene annotation to COREs already revealed that these broad domains were associated to cell-type identity genes. In 2013, a study integrating several histone modification profiles and expression data from 10 cell lines, identified a similar subset of regulatory elements termed stretch enhancers as they display extended lengths in epigenetic marks[62]. In the same year, the concept of super enhancers was proposed[63]. Defined by Mediator complex occupancy (or as seen in other studies by other transcription coactivators or epigenetic marks) and using a pre-defined list of stitched enhancers, the super enhancer label is assigned to the top most-enriched proportion of domains that surpasses an arbitrary defined threshold (i.e. dictated by the slope of a plot). As seen in stretch enhancers, super enhancers are associated to cell identity genes; they are found to be enriched in disease single nucleotide polymorphisms (SNPs).

Part II. The Swiss Army knife of transcription

Discovery of the Mediator complex

The eukaryotic rise of promoter complexity together with the expansion of general transcription factors acting at long distance enhancers was followed by the emergence of the Mediator complex (Table 2)[64].

		Bacteria	Archaea	Protists and fungi	Eukaryotes		
		(*E. coli*)	*Average*	(*S. cerevisae*)	Land plants (*A. Thaliana*)	Drosophila (*D. Melanogaster*)	Human (*H. Sapiens*)
Stimated genome size (bp)		4,6 million	1,5-4 million	12 million	157 million	165 million	3 billion
Protein-coding genes		3200	2000-5000	6000	25000	13000	20000
% of non protein-coding genome		25,5	~20	5-50	70	86,8	98,8
General Transcription Factors		Sigma factor	Ancient TBP, TFII factors	TBP, TFII factors	TBP, TFII factors	TBP, TFII factors	TBP, TFII factors
Core promoter elements		-	TATA, BRE	TATA, INR*	TATA, BRE, INR, MTE, Y-patch	TATA, BRE, INR, MTE,	TATA, BRE, INR, MTE, CpG
Histones		Ancient	Ancient	+	+	+	+
Histone tails		-	-	+	+	+	+
Chromatin looping		Architectural	Architectural	Architectural /Functional	Architectural /Functional	Architectural /Functional	Architectural /Functional
Chromatin remodelling		Minimal	Minimal	+	+	+	+
Mediator complex		-	-	+	+	+	+
HEAD	MED6	-	-	+	+	+	+
	MED8	-	-	+	+	+	+
	MED11	-	-	+	+	+	+
	MED17	-	-	+	+	+	+
	MED19	-	-	+	++	+	+
	MED20	-	-	+	++	+	+
	MED22	-	-	+	++	+	+
MIDDLE	MED1	-	-	+	?	+	+
	MED4	-	-	+	+	+	+
	MED7	-	-	+	+	+	+
	MED9	-	-	+	+	+	+
	MED21	-	-	+	+	+	+
	MED31	-	-	+	+	+	+
TAIL	MED2/29	-	-	+	+	+	+
	MED3/27	-	-	+	+	+	+
	MED5/24	-	-	+	++	+	+
	MED14	-	-	+	+	+	+
	MED15	-	-	+	++++	+	+
	MED16	-	-	+	+	+	+
	MED23	-	-	-	+	+	+
N.A	MED25	-	-	-	+	+	+
	MED26	-	-	-	-	+	+
	MED28	-	-	-	+	+	+
	MED30	-	-	-	+	+	+
KINASE	MED12	-	-	-	+	+	++
	MED13	-	-	+	+	+	++
	CDK8	-	-	+	+	+	++
	CYCC	-	-	+	++	+	++

Table 2 | Genomic features and evolutionary innovations in the kingdoms of life. Adapted from [64]

The first indications of the existence of Mediator came from studies on RNA polymerase II (RNA Pol II) transcription in yeast (reviewed by one of its discovers in [65]). Trying to decipher which components were limiting for the reaction of transcription, Kornberg's group showed in 1990 that adding activators, general transcription factors and polymerase was not sufficient to reach maximum transcription levels. It was only when a different fraction of yeast extract was added that the reaction

was accomplished. This activity was named Mediator as it was hypothesized that would contain the scaffold connecting the rest of the transcription machinery[66].

Parallel studies such as the one from Young's group found a multi-subunit complex associated with the C-terminal domain (CTD) of RNA polymerase II (RNA Pol II) although in that time it was not related to Mediator due to the co-presence of TBP and only 2% of the total yeast polymerase, not taking into account that the association could be transitory[67].

The biggest breakthrough came one year later with the purification of the complex in yeast, where 16 subunits of the Mediator were identified. Besides the function in transcription activation, it was shown that the purified complex stimulated basal transcription by 10-fold and potentiated CTD phosphorylation by at least 30-fold[68].

Further studies highlighted the general role of Mediator in virtually all yeast transcription units[69] and a Mediator cycle model was proposed where it would associate with RNA Pol II holoenzyme in a preinitiation complex (PIC), potentiate CTD phosphorylation that would start transcription and elongation, be released from RNA Pol II and re-start the cycle[70].

Early hints on the evolutionary conservation of Mediator came from the purification of the complex in mammals as a coactivator of nuclear hormone receptors[71] and interestingly, by the electron microscopy observations that, besides differences in sequence, both yeast and mouse Mediator complexes folded in a similar way together with RNA Pol II holoenzyme[72].

Composition and Structure

More than 30 subunits compose the Mediator complex in higher eukaryotes, with a combined mass of more than 1 MDa. From the early electron microscopy studies to chemical protein crosslink and mass spec approaches[73], followed by the most recent cryo-electron microscopy (cryo-EM) experiments[74], many groups have attempted to solve the structure of this macro-complex and to understand the mechanism of its binding to the transcription machinery.

What is known so far is that Mediator subunits constitute four modules; a head domain and middle domain tightly bound with a more flexible tail at the base, plus a kinase module that can reversibly associate with the rest of the complex. Nowadays a unified nomenclature for Mediator subunits is used, established after the discovery of Mediator counterparts across species[75]. The subunits for each module in yeast include MED6, MED8, MED11, MED17, MED18, MED19, MED20, and MED22 in the head module; MED1, MED4, MED7, MED9, MED10, MED21, and MED31 in the middle module; and MED2, MED3, MED5, MED14, MED15, and MED16 in the tail module. Human Mediator subunits MED27, MED24 and MED29 are structural homologs of yeast MED3, MED5, and MED2, respectively. Further work on mammalian Mediator lead to the identification of additional subunits MED28, MED29, MED30, MED23, MED24, MED25, MED26, and MED27. The kinase domain in yeast is composed of MED12, MED13, CDK8 and Cyclin C (in mammals additional paralogs MED12L, MED13L and CDK19 have been found).

Although its presence is widely conserved across the eukaryotic lineage, the protein sequence and the complex subunit composition present high variation[76]. For example, seven Mediator subunits are unique to *Arabidopsis* (named MED32, MED33a, MED33b, MED34, MED35, MED36, and MED37[77]) and some eukaryotic lineages completely lack the kinase module[64] (Table 2).

In the context of transcription evolution, as new chromatin factors emerged it was equally important to coordinate them to the pre-existing transcription apparatus. Indeed, through the course of evolution, the Mediator complex adapted to recognize new partners by the appearance of new subunits, but also through elongation and mutation of existing ones. Most variation in structure resides in intrinsic disordered regions (IDR), which are abundantly found in the middle and tail modules, and proven to be domains of protein interaction and as target for PTMs[78].

Based on the rapid evolution of these IDRs a specific inhibitor with affinity to the fungal MED15 subunit has been developed. This inhibitor disrupts the binding of a transcription factor which is key of the drug resistance pathway in fungi, but has no effects on human MED15 interactions with host transcription factors[79]. Hence, further research on species-specific Mediator differences could provide effective approaches to target eukaryotic pathogens (by disrupting specific IDR-TF interactions) not only focusing on human medicine but also in biotic stresses in plants (such as Mediator-IDRs based pesticides).

Recent structural studies, in particular the two publications of the 3.4-Å crystal structure and 4.4-Å cryo-EM map, have resolved most of the quaternary structure of the head-middle core complex and greatly expand our knowledge on the dynamics of subunit conformation. For example, MED14 acts as a backbone where subunits from the head and middle assemble in addition to its contacts with the tail of the complex. As a consequence, its span over all modules makes MED14 essential for the documented structural shifting of the complex. MED17 serves as the major interface of the head module with MED14. The remaining subunits of the head module assemble in a conformation consisting of a connector neck with a jaw, part of which is movable and connects to RNA Pol II. Several subunits of the middle module interact with MED14 nicely complementary to its shape and forming a more rigid structure termed as hook, hinge, connector, knob, and plank. Both studies coincide that the middle knob and head neck domains of Mediator lock the CTD of RNA Pol II, triggering the further interaction of Mediator plank and the RNA Pol II subunit Rbp1. Due to its high mobility and disorder, only low resolution structures exist of middle subunit MED1 and the tail. Interestingly, the tail has proven not to be completely essential to the core Mediator although its presence is key for binding to DNA-binding transcription factors. Finally, a high resolution structure of the kinase module is currently missing, but it is hypothesized that it docks to the Mediator middle hook domain via MED13 (Figure 6).

Functions of Mediator

In addition to its aforementioned role in PIC assembly and RNA Pol II transcription initiation, the Mediator complex serves a wide range of functions, with most of them in transcriptional activation[76].

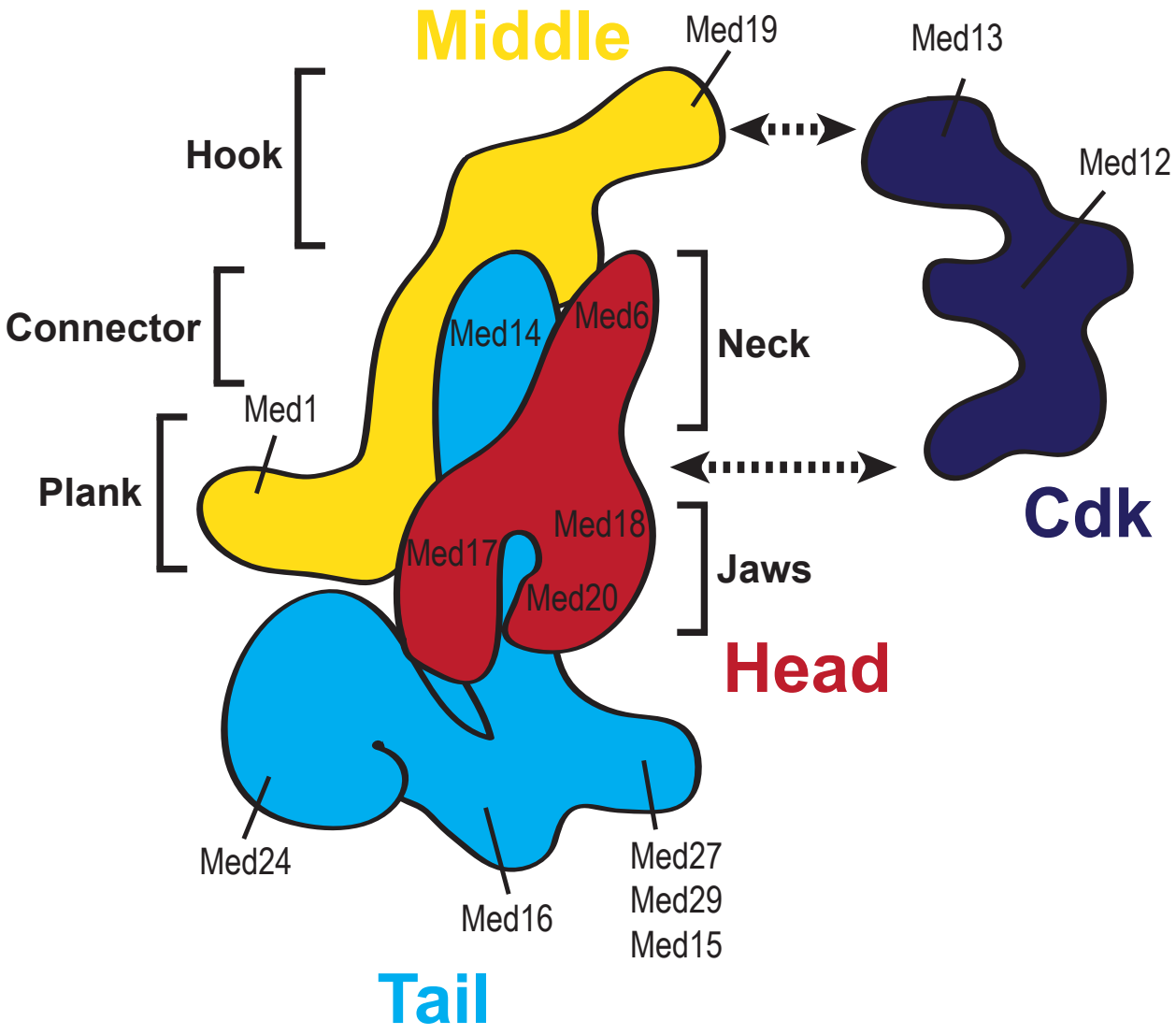


Figure 6 | Subunit localization within the Mediator complex. Adapted from yeast studies[74,80,81]

Often after metazoan transcription initiation, RNA Pol II pauses after 30-60 nucleotides via the action of NELF and DSIF complexes and resumes transcription via a process called pause-release, a rate-limiting step dependent on elongation factors such as CDK9[82]. Until recent studies, the specific localization and function of MED26 subunit was poorly described in part due to its inconsistent appearance in Mediator purifications. Meanwhile, MED26 has been identified as the link between transcription initiation and elongation; it serves as docking for the super elongation complex (SEC), switching Mediators binding from general transcription factors to elongation factors[83,84]. Moreover, CDK8 kinase activity is important for the recruitment of SEC to a different subset of genes, suggesting a parallel mechanism of elongation that depends on the target. Possibly, CDK8-SEC may play a role in early pause-release events when the gene has just been activated, with MED26-SEC ruling steady transcription afterwards[85].

Roles in transcription termination have been also proposed, in particular via MED18. Both in yeast and plants, MED18 binding has been found at gene termination regions showing impairments in expression upon Med18 depletion[86,87].

From affinity purification of MED23 together with mass spectrometry analysis, a link with splicing factors of the hnRNP family was made[88]. Although association with the RNA processing machinery has to be taken with a grain of salt due to Mediator's function in elongation, this new role of Mediator will have to be taken into account in further studies.

Due to its ability to bind RNA Pol II at promoters via its core domains and transcription factors mainly via its tail module (an updated list of them can be found in [89]), the Mediator complex has often been suggested to act as a bridge between enhancers and promoters. However, only recent studies where the genome binding of different Mediator subunits was sequentially studied showing that a single Mediator complex simultaneously contacts enhancers and promoters, finally provided the mechanistic prove to this model[90].

Mediator has meanwhile also been implicated in long-range interactions by helping Cohesin to promote the looping necessary for gene activation[91];in addition, looping is essential for MED18-mediated termination of transcription[92]. More importantly, a recent study in yeast indicates that the chromatin-bound fraction of Mediator occupies chromosomal interacting domain boundaries suggesting a more prominent role of Mediator in high-order genome structure[93].

Another complexity emerged with the inclusion of enhancer RNAs (eRNA) or activating ncRNAs (ncRNA-a), which is related not only to looping, but also to the transcription of non-coding RNAs (ncRNA) and the structure of Mediator. Although there is some controversy as to whether they are the same, it is clear that Mediator is involved in the transcription of ncRNA, which fold in a tridimensional molecular structure, aiding Mediator-mediated looping and potentiating the transcription of its target loci[94].

Due to its strategic location and exceptional size, the Mediator complex also constitutes a platform for coactivator recruitment. To date, more than 550 protein-protein interactions have been accounted for the human Mediator complex (according to Biogrid database). Well known chromatin regulators such as EP300-CBP, CHD1, the TRRAP complex and the SAGA complex interact with Mediator[89,95,96,97]. Recently, CARM1 (coactivator-associated arginine methyltransferase 1), also known as PRMT4 (protein arginine N-methyltransferase 4) has been found in a high-throughput affinity purification based screen using MED9 as bait[98]. Although many of these complexes associate to TFs, the scaffolding effect of Mediator should be also considered for their recruitment. Nevertheless, the interaction of TFs with Mediator is required for the structural shift of the latter, allowing the recruitment of coactivators[96,99].

Along with interactors involved in direct chromatin regulation, Mediator has also been found to be post-translation modified by an increasing range of proteins. As previously mentioned, Mediator IDRs contain abundant sites for PTM and other studies show how signaling cascades converge on

these PTMs, affecting Mediator function in various ways. Global proteomics approaches have uncovered several PTMs on Mediator[100], but very few mechanistic studies have as yet been . Nonetheless, MED1 phosphorylation mediated by MAPK/ERK[101] or PI3K/AKT[102] pathways appears important for MED1 association to the complex, looping and PIC assembly. In addition, work from Grosveld's lab suggests that CDK9 phosphorylates MED1/9 (*unpublished data)*. MED13 and MED13L appear to be phosphorylated and then degraded via the E3-ubiquitin ligase FBW7 mediated ubiquitylation, compromising the recruitment of the kinase module to the complex[103]. CARM1 not only acts as a histone modifier (see above), but has the ability to Arginine-methylate other proteins such as EP300/CBP[104], but also MED12[105] (*see also this PhD thesis*). A new working model on Mediator cycle of transcription implies degradation of not only the recruiting TFs, but also of the tail subunits of Mediator at enhancers[80]. As examples, yeast MED3 tail subunit was found to be degraded after CDK8 phosphorylation[106] and MED15 was found to be destabilized by TRIM11[107].

In contrast to its function in transcription activation, Mediator has also been related to repression and silencing of expression, mainly accredited to the CDK8-kinase module based on its independent actions from the core. First, it was shown that in human cells Mediator containing the kinase-module repressed transcription[108]. In addition, mutations in the kinase-module resulted in gene expression upregulation [109,110,111]. As mentioned, CDK8 kinase activity regulates transcription factor degradation, another example being Notch intracellular domain at enhancers[112]. Finally, the kinase module subunits interact with chromatin repressors such as G9a histone (H3K9) methyltransferase[113], PRMT5 (a histone arginine methyltransferase[114]) and the Polycomb repression complex (PRC)[115]. Along these lines, intriguing studies relate Mediator to pericentromeric heterochromatin, hypothetically via a MED26-HP1 interaction[116], and to telomere maintenance[110,117,118].

Finally, Mediator has been linked to the DNA-damage response (DDR). Indeed, MED17 recruits the DNA repair protein RAD2 to the genome and MED17 mutants result in increased DNA-damage sensitivity to cells[119].

Mediator in development and disease

Subsequent to the recruitment by transcription factors and its interactions with epigenetic regulators, the Mediator complex plays crucial physiological roles. Aberrant function of MED1, MED12, MED21, MED23, MED24, MED31, and CDK8 subunit leads to embryonic lethality[89]. In addition, genetic screens to identify regulators of embryonic stem cell (ESC) state identified a long list of Mediator subunits as essential for *OCT4* mRNA expression, encoding a TF master regulator of embryonic cell pluripotent state[91].

Other subunits, when mutated, display a defined phenotype due to aberrant interactions. such as MED19/26-REST in neurogenesis[120], MED1 in adipogenesis[121],MED14[122] as interactor of PPARγ, GATA1-dependence on MED1[123,124], MED15-Smad2/3/4 in mesoderm development[125], the link of SOX9 and MED12[126] and MED25[127] in chondrogenesis, MED12-SOX10 in oligodendroglia[128] and MED23-RUNX2 in bone development[129].

Extensive studies of Mediator complex have also been carried out in plants. Besides roles in plant development, the idea of Mediator as a hub of transcription really shines in the coordination of signaling cascades in this eukaryotic kingdom. Many studies place Mediator as the nexus of many hormone-mediated responses to both abiotic stress (such as cold and drought), but also in the defense response to plant pathogens[130].

Many human diseases have an origin in Mediator dysfunction[131]. Not surprisingly, many of the Mediator-associated diseases have a developmental component. Remarkably, Mediator subunit gene mutations are a frequent cause of neurodevelopmental disorders, including X-linked intellectual disability (*MED12*[132]), microcephaly (*MED17*[133]), congenital retinal folds and intellectual disability (*CDK19* haplo-insufficiency[134]), Charcot-Marie-Tooth disease (CMTD) and eye-intellectual disability syndrome (*MED25*[135,136]) and intellectual disability (*MED23*[137]). Together with intellectual disability and developmental delay, *MED13L* haplo-insufficiency syndrome features cardiac congenital defects[138]. Also affecting the heart, a chromosome deletion involving *MED15* has been shown to cause cardiac conotruncus defects[139].

The correct fine-tuning of transcription is essential for cell homeostasis, and slight alterations can lead to malignancy. As central operator in transcription, the Mediator complex has the potential to play important roles in oncogenesis[140]. Indeed, many genes encoding for Mediator subunits have been found to be misregulated in cancer[141], but few mechanistic studies have been published. For example, the very well described MED1 interaction with nuclear hormone receptors[142] explains its implication to androgen[143] and estrogen[144] dependent tumorigenesis. In addition to that, the role in modulation of Wnt/beta-catenin[145] signaling could explain in many cases Mediator´s implication in tumorigenesis[146,147]. Finally, the oncogenic role of the CDK8-kinase module[148] could be targeted with the recent development of CDK8/19 inhibitors[149].

Part IV. Let's get neural

In addition to the described increase in transcription complexity, the expansion of genes involved in cell-cell communication and cell adhesion allowed the diverse evolution of metazoans and their wide radiation[150]. The innovation in signaling systems (biochemical pathways and their nuclear interpretation resulting in genomic transcriptional responses) granted the ability to generate more sophisticated body structures. This way, in early metazoans endodermal cells give rise to an internal digestive epithelium; the ectoderm, originally forming a protective epithelium towards the environment; and as a result of endoderm-ectoderm interaction, the induction from ectoderm of mesoderm, a mesenchymal layer between the other two[151], giving rise to many cell types of may later tissues and organs.

Neurons are ancient

Even prior to the presence of mesoderm in the animal kingdom, a specialized cell type of the ectoderm (and in some cases endoderm) made its appearance, the neuron. Until that point, the chase of other organisms as a source of energy may have happened by sensing nutritional, chemical, light or

temperature gradients, basic processes that could be achieved by sensory cilia[152]. However, together with the formation of multicellular organisms, predation may have pushed the development of new fast and highly coordinated sensing-response strategies[153]. Neurons are specialized and high-energy demanding cells with the role of transmitting signals via chemical and/or electrical reactions to other neurons or other cells. Their shape can vary but they share common features such as the soma, the main body of the cell containing the nucleus; dendrites, cellular extensions acting in signal inputs, and axons, the principal projections acting as connection fibers and commonly acting in output signaling. The synapse is the contact structure between neurons (or between neurons and non-neuronal cells) where chemical neurotransmitters are exchanged[154]. The establishment of synapses (synaptogenesis) requires a complex machinery of proteins acting as synthesizers, releasers, transporters, receptors and modulators. Interestingly, a basic neural genetic toolkit is already present in more ancient organisms such as choanoflagellates, unicellular organisms closely related to the first metazoans[155] and it has been proposed that multicellularity and gene duplications unlocked their potential to form the first synaptic structures in evolution.

From hundreds of neurons to millions

Soon in metazoan evolution the appearance of an embryonic region capable of generating a nervous system was selected in order to integrate and coordinate neuronal networks across the body. Particularly in symmetric bilaterians, the nervous system became internalized, anteriorized and concentrated in a mass termed brain and a connecting web of nerve cords[156]. Early evolutionary examples of the first bilaterians with brain are nematodes such *Caenorhabditis elegans* which contain 302 neurons in the whole body, and its study has helped the general understanding in eukaryotic development and neurophysiology[157].

Gene duplication is a major evolutionary mechanism as it provides new copies of genes that can diverge to acquire new functions[158]. Vertebrate genomes contain multiple paralogs of many genes of the fruit fly (*Drosophila melanogaster*). Such is the case of the *Hox* genes which invertebrates have a single *Hox* cluster corresponding to four human and mouse equivalent A-D *HOX* clusters, although the duplications are not perfect[159].Notably, the number of coding sequences in vertebrate genomes does not scale proportionally to their increased length, indicating that – as illustrated above – many if not most of the duplicated genes were lost. However, and quite interestingly, there is a disproportional retention of genes involved in developmental processes and neural activity. This increase in the genetic toolkit in addition to the refinement of *cis*-regulatory regions coincides with the appearance of the first vertebrates (chordates) almost 500 million years ago[160].During the course of evolution this combination allowed the expansion of the nervous system both in size and complexity[161].

From an egg to a brain, study of neural development

As brains became larger, the number of neurons and their connectivity also increased, allowing also animals to adapt to more diverse environments and facilitating their radiation. This phenomenon of

evolutionary *encephalization* is more patent since the emergence of placental mammals 100-150 million years ago. The forebrain began to expand rapidly, producing additional cortical subdivisions and more complex neural networks[167].

BOX 2. The mouse as a model organism

Nowadays many different eukaryotic species are used in research ranging from the unicellular yeast, a wide range of plants, small worms and flies to bigger vertebrates such as fish, frogs, mice and rats, guinea pigs or even monkeys and apes. All of them are powerful model organisms to study *in vivo* biological process that can be, always with certain bias, extrapolated to the human physiology. The use of model organisms has been fundamental for the advance of not only our general understanding of biology but to great improvement in medicine of the past centuries[162].

Mice have been formally studied since the beginnings of the 20th century. Their resemblance to the human physiopathology and development, their small size and easy handling and relative short life cycle have fomented its use as a model organism. Currently mice account for more than 60% of all vertebrate models used in research with more than 7 million exemplars used each year only in the European union (stats from 2011[163]). In 2002, its genome became the first mammalian one completely sequenced, and with the sequencing of the human genome a year later; it was shown to share around 80% of the same protein coding genes[164]. Due to their high similarity to humans, mice often provide good models to study and understand human physiology and complex genetic diseases. Furthermore, the development of genetic engineering has allowed the creation of mice carrying specific mutations to mimic different phenotypes and up to this date there are more than 41000 different mice strains[165]. Nonetheless, mice are used not only as research models but also as producers of therapeutic agents such as antibodies, which with recent technologies have reach the milestone of humanized monoclonal peptides[166].

But how is this intricate structure that we called brain formed? As hinted before, the answer relies on the tight spatio-temporal combination of genes and regulatory signals that shapes the development of the organism from its starting point, the fertilized egg or zygote.

Embryonic stem cells, mothers of all cells

Indeed, at the moment of the fertilization of an oocyte by a spermatocyte, yieldingthe 1-cell zygote, all the genetic information to generate, maintain and reproduce the new organism is already contained within the zygote. In mammals, this developmental plan starts already while the zygote and the arising cleavage-stage embryos travels to the uterus (for implantation). In the mouse, it takes about 2.5 embryonic days (E2.5) to generate a mass of 8-16 cells named morula. Between 16-32 cells the first developmental decision is taken as cells of the morula after compaction have to provide the embryo with cells that will become the proper embryo on the one hand and on the other hand, cells needed for implantation of the early (E3.5) and then late (E4.5) blastocyst. The net result is the formation in

the (cavitated) blastocyst of asymmetrically distributed inner cell mass (ICM) cells (at the embryonic pole of the blastocyst) and the trophectoderm cells surrounding the entire blastocyst, respectively. Interestingly, not only transcription factors play a role as chromatin modifiers such as CARM1 may also be essential for this process[168].

Until this point ICM cells have the potential to give rise to all of the cell types of the future embryonic and adult body, just like this is achieved in the mouse embryo by gastrulation, which starts at E6.5. Hence the term embryonic stem cells (ESCs), the cell culture derived counterparts of ICM cells of the pre-implantation blastocyst, for their pluripotency allows them to generate all cells for the development of the organism.

Indeed, ESCs can be isolated from pre-implantation blastocyst stage mouse embryos and their pluripotent state can be maintained in well-defined cell culture conditions[169]. This enables their expansion and, using different cell culture conditions, their differentiation along the three germ layers and cells derived thereof[170]. Undifferentiated ESCs can be modified by genetic engineering and then transplanted back to a non-compacted morula or injected into a forming blastocyst from an acceptor embryo giving rise to chimeric mice, which after appropriate crossing can generate full genetically modified organisms[171]. Moreover, the ability to expand ESCs in high numbers and differentiate them to particular cell types with high or sufficient efficiency has been fundamental for the development of new cell-based therapeutic strategies in regenerative medicine[172]. Thus, the study of ESCs, based on initial crucial work with mouse ESCs, has attracted a lot of attention not only due to its human clinical potential, but also – and important for this PhD thesis - as an excellent cell model to study transcriptional regulation during development.

One of the major fields in ESCs research is the study of the extrinsic and intrinsic signaling systems and resulting pathways that govern the self-renewal and (the meanwhile various) pluripotency states of these cells. For example, the inhibition of glycogen synthase kinase 3 (GSK3) by Wnt signaling supports ESC self-renewal and, together with the block of the FGF pathway inhibition of ERK, constitutes a 2-inhibitor (2i) cocktail widely used in cell cultures. On top of that, LIF, a product of the trophoectoderm, signals to ESCs via the LIFR and the downstream STATs, supporting self-renewal, hence many protocols opt to culture ESCs in serum/LIF conditions. However, ESCs cultures with serum/LIF seem more heterogeneous, resembling more the ICM cells of the late blastocyst, and are not identical to the 2i-mediated ground state[170,173].

The integration of the aforementioned LIF, FGF, Wnt and likely BMP (present in serum) extrinsic signals converges to the nucleus where the action is taken by downstream transcription regulators. Among them, Oct4, Sox2 and Nanog constitute a well-described transcription factor core system that is key to pluripotency acquisition and maintenance, and acts via auto-regulatory feedback loops. The study of these factors has led to the discovery of many others acting with them[174] and the genomic characterization of the epigenetic landscape of ESCs have expanded the core TFs to include others such as Klf4, Esrrb and Prmd14[175]. One of the most notable accomplishments in the study of ESCs transcription has been the use of Oct4/Sox2/cMyc/Klf4 TFs in order to reprogram somatic cells to i.e.

induced pluripotent stem cells (iPSCs)[176]. Although the process is not very efficient, it provides circumvention to the ethical problems of obtaining human embryonic tissues. Such iPSCs represent the opportunity to develop therapeutic strategies using cell systems derived from patient-own cells[177].

Another particularity of ESCs is their epigenetic landscape. Due to their ground-state in development and their potential to differentiate to the three main lineages, ESCs chromatin seems to be more permissive than more mature cells. Instead of strong defined heterochromatin silenced regions, many developmental genes in ESCs appear to be repressed in a less sturdy manner showing a poised state with activation marks. These bivalent domains are characterized by the histone mark H3K27me3 and they are regulated by the repression of Polycomb group (PcG) of protein complexes, PRC1 and PRC2[178,179]. Hence, the specific de-repression of some of these genes casts the path that the cell will take into its final lineage. Most bivalent genome domains shift to a single state upon differentiation, although bivalent domains can also rise in several steps of development when the cell is at a crossroad of determination[180].

Neural ectoderm and neural stem cells

Even before implantation to the uterus, the second lineage specification begins to take place within the ICM cells to separate them in epiblast, which will compose the mesoderm and ectoderm; and hypoblast (or primitive endoderm) which will give rise to the visceral and parietal endoderm. A round of division later, around E4,5 a cavity starts to form in a process called gastrulation and the embryo starts to reorganize into a multilayered structure[181] (Figure 7A).

The nervous system originates from the induced neuroectoderm, which around E7.5 as a thickened, but flat neural plate wherein all cells have the potential to become neural cell types, but they will not all do so, as Delta-Notch signaling will provoke lateral inhibition in these cells; furthermore, the neural plate matures via patterning in anterior-posterior direction. FGF secreted from the anterior neural ridge (ANR) plays an important role in this, while the neural plate is also flanked by neural crest cells and the ectodermal placode cells, which can only arise at intermediate concentrations of BMP, whereas BMP activity has to be avoided in the neural plate itself. In response to signals between this neuroepithelium (NE) of the neural plate and surrounding tissues, also a longitudinal groove forms along the neural plate (referred to as the process of neurulation) and, at different points, the neural plate will display hinges around which it curves on itself to give rise to the neural tube.

This developing neuroepithelium will generate most of the neurons and the non-neuronal cells (glial cells) of the CNS[182]. At the start of gastrulation, cells from any part of the ectoderm can still develop as either epidermis or neural tissue. Here is where morphogenetic positional signals produced both from within and outside the ectoderm play a crucial role in the process of neural induction (Figure 7A).

One of the most prominent signals in this stage of development is BMP, the production of which in *Xenopus* progressively concentrates in the ventral and lateral mesoderm and acts as a ventralizer of mesoderm, dose-controlled mesoderm patterning factor, epidermis promoting factor and neural

induction inhibitor. In the mouse its antagonism (by molecules such as Noggin, Chordin and Follistatin, and the Wnt/Nodal/BMP antagonist Cerberus) is not sufficient to induce neural fate[183].

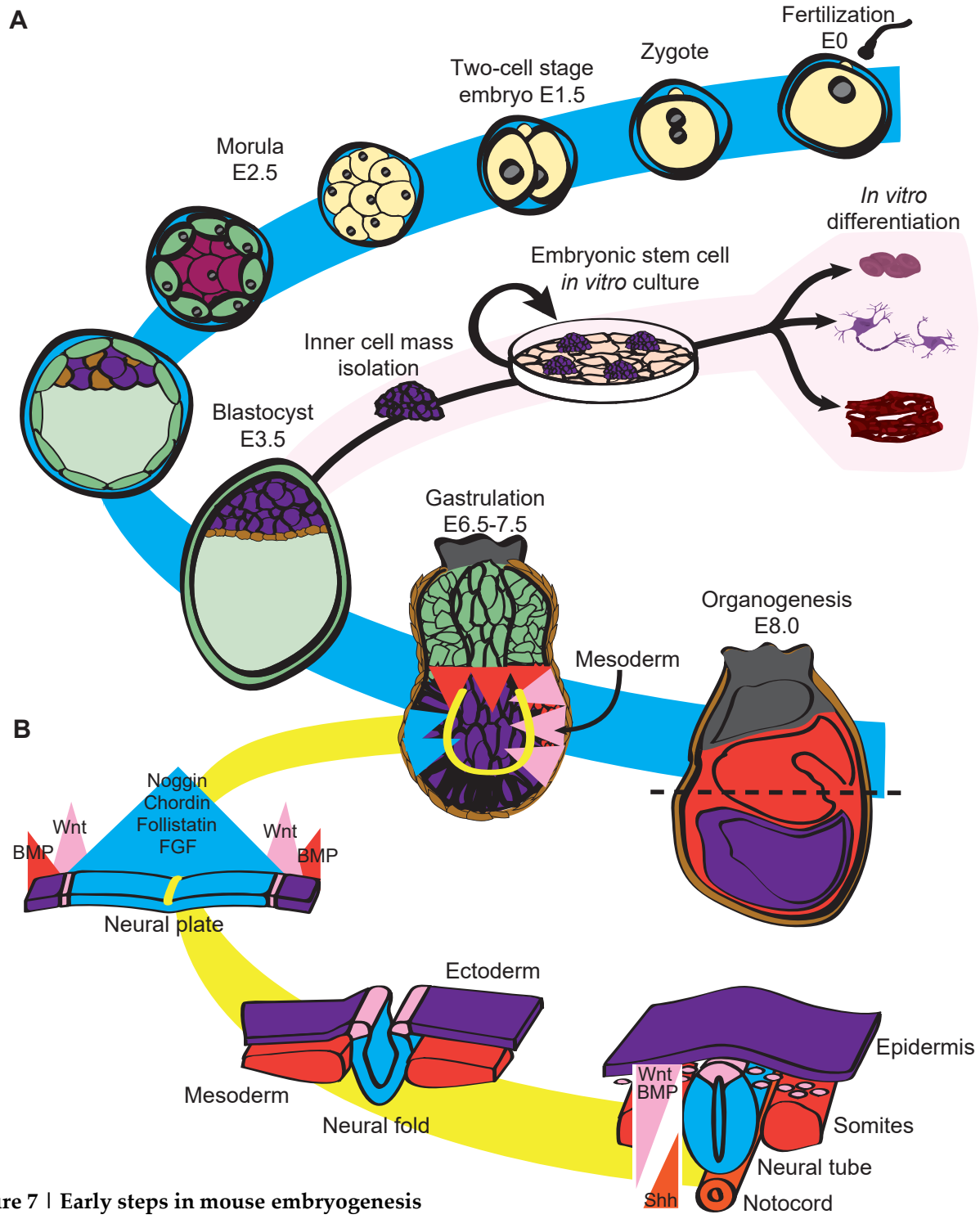


Figure 7 | Early steps in mouse embryogenesis

A. From fertilization to embryonic day 8 (E8); Note that embryonic stem cells can be isolated from the inner cell mass of the blastocyst to establish cell culture models. Around E6.5 the process of gastrulation starts (triangles indicate morphogen signaling, yellow line indicates the primitive line). B. Process of neural tube formation (triangles indicate morphogen signaling).

Indeed, FGFs produced by the prospective neural cells, acting as neural inducer proteins possibly by a direct effect in preventing BMP transcription in the mouse anterior visceral endoderm (AVE) and embryonic ectoderm[184]. However, lateral epiblast cells in close vicinity have been claimed to be insensitive to the FGF gradient. Consequently, it has been shown that Wnt signaling is also a critical determinant of non-neural fate as its continued signaling blocks the FGF response[185] (Figure 7B).

While the neural tube is formed, it is at the same time subdivided into discrete territories according to the levels of morphogens (and neural-promoting antagonists) generated by several signaling centers. In addition to the anterior(rostral)/posterior(caudal) patterning established by BMP, FGF and Wnt gradients, a dorsal/ventral axis settles during neurulation and in the closed neural tube with the action once more of BMPs, this time secreted by the roof plate of the neural tube; and Sonic hedgehog (SHH), secreted in large amounts from the notochord and in the most ventral part (non-neural cells; floor plate) of the neural tube. Furthermore, along the neural tube the paraxial mesoderm also influences the neural tube anterior-posterior patterning in the trunk region of the embryo by the generation of retinoic acid (RA), which is essential for the development of the medial segments of the neural tube that give rise to the hindbrain and the spinal cord. The interplay between these gradients is not only fundamental for the proper patterning of the CNS, but also for the specification of the neuronal sub-types along the dorsal-ventral axis of the neural tube that will give rise to interneurons and motor neurons, respectively, from cell progenitors. Only with the study of the proper signaling within and to the CNS it has been possible to generate specific neuronal populations in cell culture[186].

The combination of dosed signals and their interplay triggers the activation of cell-intrinsic factors that regulate the competence of the neural ectoderm. For example, SHH signaling is integrated by the vertebrate Gli family (GLI1-3) of TFs acting early in neural plate formation[187]. Very similar to Gli protein domains, the Zic family is involved in neuroectoderm differentiation after BMP inhibition[188,189]. In addition to Zic family proteins, the BMP antagonism is mediated in part by (likely) anti-BMP-Smad action of Zeb2, a multifunctional DNA-binding transcription factor that is needed intact for activation of neurogenesis genes such as *Sox1-2*and also acts in BMP-sensitive neural crest cells[190,191].

Derepression of neural genes by the degradation of pluripotency factors and differentiation blockers is also a well described mechanism of neural induction. For example, REST neuronal repressor already starts being degraded in the transition from ESCs to the neural commited progenitors allowing the expression of neural induction transcription factors like *Ascl1* while repressing terminal neuronal differentiation[192]. Another case is the down regulation of neural tissue relevant miRNAs that e.g. target *Zeb2* and *Pax6* TFs during neural induction and the loss of ESCs pluripotency[193]. Finally, many poised genes resolve into an active state by the regulation of Polycomb-mediated repression and the action of chromatin modifiers such as demethylase JMJD3 (also known as KDM6B), which removes H3K27me3 marks of neural genes[194].

Following the exit from pluripotency to neural commitment Sox2, which is already present in ESCs, has been shown to be bound to poised neural enhancers[195]. In fact, the whole SoxB1 family of

transcriptions factors act early in neural induction. *Sox1* expression starts as soon as the appearance of the neural plate and is downregulated after neural tube formation, hence it may be a marker of early neural induction[196]. *Sox3*, another early marker, is expressed promptly in the ectoderm and it has been shown to co-occupy *Sox2*-bound genomic regions in neural-induced cells[195]. Together with other TFs such as Geminin and Zic2, Sox2 and Sox3 are essential for the maintenance of an immature proliferating state of neuroepithelial cells (NE cells).

Thus, NE cells are neural stem cells (NSCs) and start to appear in the ventricular zone (VZ) of the neural tube as early as E8.5 in the mouse, dividing mostly symmetrically to rapidly expand the stem cell pool. The correct regulation of their expansion and survival is vital to achieve a suitable nervous system. In fact, disorders affecting the NSCs pool lead to major neurodevelopmental disorders such as microencephaly, where the exhaustion of self-renewing NSCs leads to smaller brains often associated with intellectual disability in various human syndromes[197].

NSCs can be isolated from mouse embryos and, by applying the right culture conditions (adding growth factors like FGF and EGF), can be expanded while retaining their multipotent capacity[198]. The *in vitro* culture of NSCs has widened the possibilities for their study, notably at the biochemistry level, given insights in the chromatin landscape dictating their physiology[199–201]. The transcriptional networks regulating NSCs are enriched for basic helix-loop-helix (bHLH) factors such as factors of the Hes, Id, Olig and NeuroD families as well as other E-box binding proteins such as E2-2 (Tcf4). Their mRNA expression vary according to the signaling they receive, depending on their localization across the developing CNS[202]. This dynamic regulation of the level of expression of these neural genes also determines the future fate of NSC progeny. The antagonistic action between Hes and Id self-renewal factors against the proneural differentiation effects of NeuroD1, Ascl1 and Ngn2 appears to oscillate[203], keeping the progenitor state in a fragile equilibrium until single-fate determination factors take control.

The most anterior part of the neural tube termed as telencephalon will give rise to a significant part of the human brain, the cerebral cortex, believed to endow mammals with the ability to perform complex tasks. Around E10 NE cells in this part of the embryo transit into a new NSC type, the radial glial cells (RGCs). RGCs define the architecture of the entire future neural tube as their soma and basal body lie in the VZ and a long radial process extends from their cell body to the pial surface of the (rapidly broadening) neural tube. While both NE cells and RGCs express the neural progenitor marker *Nestin*, RGCs also contain the specific markers Glast, Blbp and Rc2, and later Pax6[204]. Different populations of RGCs have been described, but in general terms it has been shown that RGCs undergo symmetric and asymmetric cell divisions, not only expanding the numbers of NSCs, but also generating more differentiated neuronal cells[205].

Neurogenesis

The onset of neurogenesis in the mouse cortex starts around E10-11 with the asymmetric divisions of RGCs, which will give rise to post-mitotic neurons or intermediate progenitors, which move to the subventricular zone (SVZ) (Figure 8).

The neurogenesis switch seems to be triggered *in vivo* by many cues that break the exclusive expanding progenitor state of the neuroepithelium and favors the transition to neuron- producing RGCs. Besides Notch signaling (through Hes family of transcription factors[206]), Pax6 and FoxG1 seem to be directly involved in the NE to RGCs transition together with the action of FGF (*Fgf10* expression starting around E10 in the VZ)[207]. The BMP role is complex as after its initial neural-inhibitory activity it seems to be involved subsequently in blocking neurogenesis and promote astrogliogenesis[208]. RA produced by the meninges (mesenchymal layers that surround the CNS and act as protection and exchange layers) is also fundamental to cortical neurogenesis[209].

The intrinsic effectors of all these extrinsic signals are generally termed proneural genes as they activate neuronal terminal differentiation or (in the mouse, subsequent gliogenesis). As already mentioned, bHLH factors such as the Neurogenin family and Ascl1, can induce rapid and full neuronal differentiation. Interestingly, both Ngn2 and Ascl1 have the capacity to directly reprogram, including *in vivo*, other differentiated cells into neurons[210,211], denoting a pioneer master regulatory effect. Downstream this wave of early proneural genes, a wide range of secondary factors such as those of the Dlx, NeuroD, SoxC (Sox4 and Sox11) families, together with Tbr2 settles the neuronal fate[212]. Besides the downregulation of progenitor functions, several of these proneural transcription factors later target genes involved in cell-adhesion, axon guidance, guided cell migration, cell shape changes and synapse formation, preparing the newborn neurons first to reach their correct destination and later develop an arborized axon projecting structure[213].

In addition, cell cycle control is decisive during CNS development[214]. First, NSCs undergo rapid divisions with the need to tightly control the mitotic apparatus. On the other hand, the birth of post-mitotic neurons is linked with the exit from the cell cycle, establishing them in an undefined G_0 phase. Hence, many transcription factors also target cell cycle regulators like cyclinD1 (*ccnd1*) and CDK inhibitor *p21*[215–217].

As neurons are born from RGCs or intermediate progenitors in the SVZ, an essential process of lamination into layers starts. RGCs provide the pillars where newly generate neurons migrate from the VZ/SVZ into the cortical plate forming eventually, as time proceeds and the brain cortex grows in thickness, an "inside-out" gradient of maturation. The correct migration of the different neuronal populations into layers is regulated not only by the contact with RGCs but a wide range of signaling molecules. For example, the earliest-formed cortical neurons that generate the outer layer I, termed preplate, express Reelin, an extracellular matrix molecule essential for the correct positioning of newborn neurons[218]. Meanwhile, it has been shown that TFs such as Zeb2, also an E-box binding TF, in the upper layers signal in a non-cell autonomous fashion back to the stem cells in the VZ by

regulating (as a transcriptional repressor) the levels of neurotrophins and FGF ligands that time neurogenesis and gliogenesis, respectively[219].

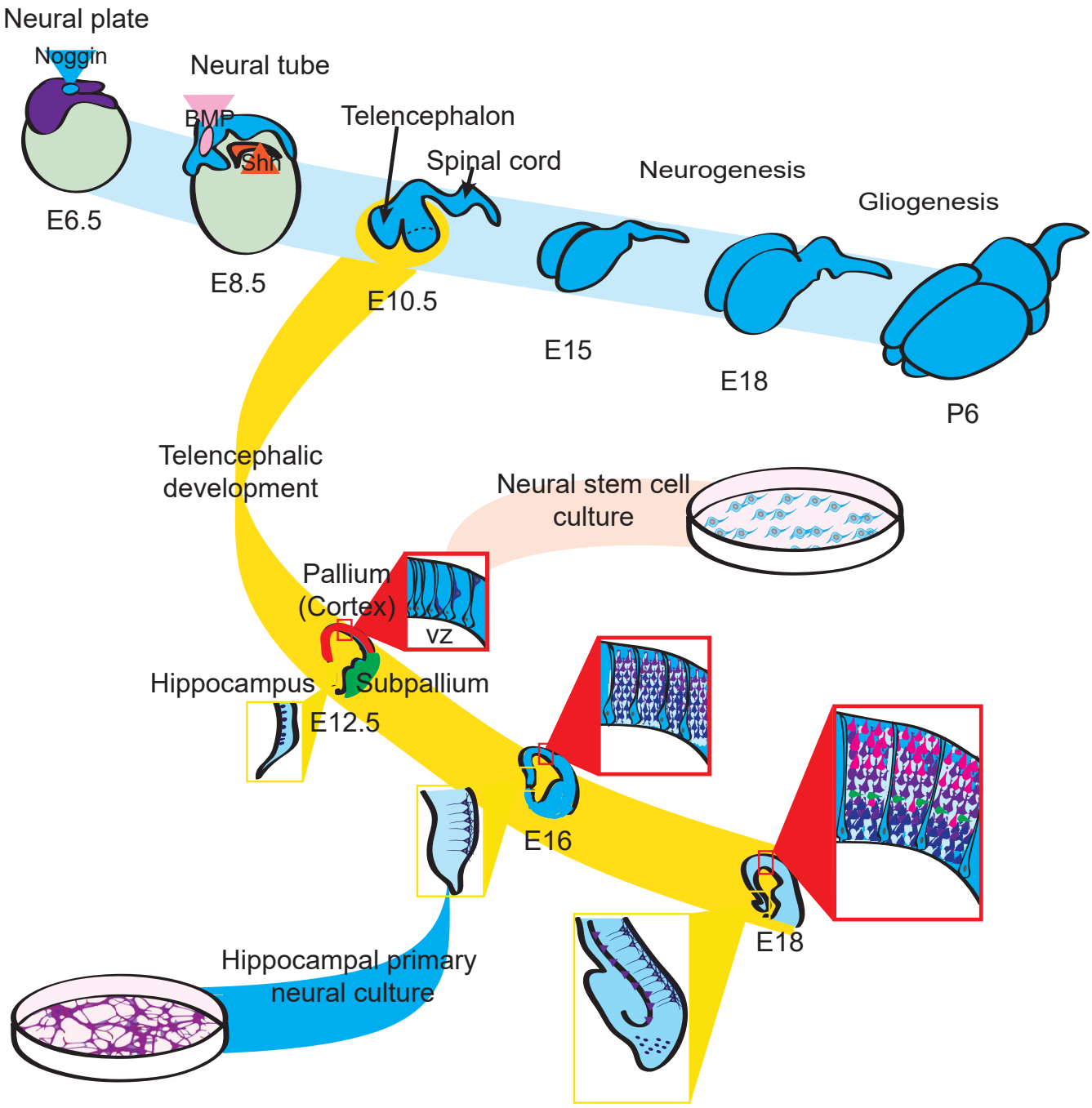


Figure 8 | Mouse central nervous system development
Telencephalic development is highlighted apart. Note that several cell culture models can be established by isolating specific parts of the brain in different windows of time.

While neurons born in the dorsal part of the telencephalon (pallium) occupy their designated layers by radial migration and the neurons formed are pyramidal, excitatory neurons, the inhibitory GABAergic interneurons, which compose around 20% of total neurons in the cortex, migrate tangentially from subpallial structures (termed ganglionic eminences) along well-defined paths to invade and later integrate radially into the cortical neuron network[220].

As seen, such migration is fundamental to many phases and locations of the nervous system development and its impairment is accordingly the cause of several brain diseases[221]. Many molecules that direct migration can act as attractants or repulsers first for the whole cell but also later for its projections. Among them, Netrin, Ephrin and Semaphorin families have a central role[222]. Fascinatingly, mutations in the Integrator complex, which is involved in metazoan RNA Pol II pause-release among other functions[223], also produce cortical migration defects, insinuating specific roles in development for the general transcription machinery[224]. In addition to that, large numbers of neurons die by programmed cell death during the first steps of brain maturation in a competition process for establishing the proper connections. Axon guidance molecules and the final connection with synapses, accompanied by the action of neurotriphins such as BDNF, dictate the survival of post-mitotic neurons[225].

According to the time of generation and their positioning, neurons from each layer and region of the cortex are influenced by specific signal inputs determining the specific cocktail of proneural genes that define their neuronal sub-type identity. First, RGCs at the VZ/SVZ are identified with *Pax6*, *FoxG1* plus *Emx2* and *Lhx2*. In addition, intermediate progenitors in the SVZ expressed *Cux1-2* and *Tbr2*. Later, upper layer markers are *Cux1-2*, *Mef2c*, *Brn1-2* (*Pou3f3*/*Pou3f2*, respectively) while *Sox5*, *Foxp2*, *Fezf2* and *Tbr1* represent deep layer markers[226,227].

After the process of embryonic neurogenesis that mostly takes place between E11-16 in the mouse, a major switch in RGCs occurs as they start producing macroglia, which are different subtypes of non-neural cells. Macroglia are involved in the homeostatsis of neurons in processes such as metabolic support, myelinization of the axons, signaling in neuronal migration and maturation, brain repair after damage, immune reaction response (in combination with the microglia of mesodermal origin) and even on synaptic communication[228].

The switch to gliogenesis, time wise, is also highly regulated by proneural genes as many of them act as glial lineage blockers. Hence, the downregulation of genes such *Ngn2* is the first step of gliogenesis. It has been shown that neural intermediate progenitors and early born neurons signal the surrounding cells via the Notch pathway[229], which in coordination with JAK/STAT signaling, lead Nuclear Factor IA (*Nfia*), *Olig2* and SoxE family (*Sox8*, *Sox9* and *Sox10*) to promote glial specification. Zeb2 is needed for the correct timing and extent of gliogenesis in the developing brain and for counteract inhibitory effects of BMP (and Wnt) action in myelinogenesis[219,230].

BOX 3. Seahorse of memory, the hippocampus

Buried deep within the medial temporal lobe lays the hippocampus. Although residing in the telencephalon, its network is organized quite differently from that found in the cortex. With all principal cells in a single layer and the synaptic inputs to well defined dendritic lamina, the hippocampus has become the major model to study neurophysiology. In addition, its capacity to survive sliced *in vitro* for long periods of time have revolutionized the field of electrophysiology.

Furthermore, its abundance of pyramidal neurons has selected the hippocampus as a good source of homogenous neurons frequently used to study plasticity, pharmacological effects and intracellular features. Hence, the hippocampus is one of the most widely studied regions of the brain not only for its central role in memory and its dysfunction in Alzheimer's disease but also as a model of synaptic plasticity.

On top of this, the extraordinary discovery of the hippocampus as one of the few sources of new neurons in the adulthood (together with the SVZ in humans, plus the olfactory bulb in mice), has pushed the study of the hippocampus to new frontiers[231].

Neural maturation

The generation and migration of neurons to their predetermined locations are only the first steps in the developing brain. Right after birth, post-mitotic neurons are far from being functional. Instead, neurons will endure a long process of neural maturation that will keep ongoing even postnatally. Interestingly, the same post-mitotic cell that is round-shaped at birth will first have to mold into a migrating format in order to reach its destination but later will suffer major morphological changes to develop the synaptic structure. Thus, neurons switch from the expression of cell movement and guidance receptor proteins to synapse machinery and axon guidance, accompanied also by changes in genes involved in metabolic control[232].

Once neurons integrate to their destinations and establish connections via their axon and dendrites, specific transcriptional pathways regulate the processing of information. The connectivity of the CNS is flexible in order to be able to achieve processes such as memory and learning[233]. This synaptic plasticity can be achieved due to the transcriptional control of genes that regulate synaptic formation via cascades triggered by synaptic activity.

Calcium is the major intracellular molecule of synaptic activity as its influx to the neuron is regulated by channels that depend on membrane depolarization. Calcium is recognized by sensors such as kinases CamKI-II which activate by phosphorylation the transcription factor CREB, which in turn quickly induces the immediate-early genes *c-Jun* and *c-Fos*. These three factors together with MEF2 and NPAS4 control the activity-dependent transcription of a wide number of activity-regulated genes[234]. Nevertheless many other transcription factors have been suggested to be involved as well, i.e. SRF, ELK, NFAT, NFkB, DREAM, NeuroD, SP4 and CREST[235].

Perhaps the gene *Bdnf* (encoding Brain-derived neurotrophic factor) is the best reference of activated targets upon synaptic activity. Secreted BDNF has a central role in neuronal survival, dendritic growth and synaptic development. At the molecular level, the regulation of its multiple isoforms, the peptide maturation and secretion have been a focus of incessant interest in neuroscience[236]. Besides Bdnf, other targets such *Arc, Homer-1a* and *Zif268* genes and encoded actors consolidate synapses and help the long-term potentiating of connections[237].

Interestingly many of these activity responses start to take effect intermittently before neurons do fully mature. In fact, synaptic activity is crucial for the survival and maturation of the neural cell, and neural maturation can only be understood in the context of synaptic activity[238].

As seen, the TRN governing differentiation and functionality of matured neurons have been molecularly explored in depth and continuously receive attention. However, the transition period representing the maturation from early post-mitotic phase to a fully arborized neuron remains still poorly understood. While differentiation and synapse formation seem to be highly extrinsic regulated processes, maturation seems to be more cell-intrinsic[239–241].

Indeed, there is abundant evidence pointing to a major autonomous switch in chromatin regulation happening in a short window of time during neuron development. One of the earliest indications that neurons transform their chromatin landscape was the observation of changes in the location, size, and number of nuclear structures such as chromocenters and nucleoli during their maturation[242]. More at the biochemical level, it was shown that neurons present shorter nucleosomal repeat length and lower levels of histone H1 than glia, suggesting a genome-wide lineage-specific regulation of chromatin[243].

More recent epigenetic studies have revealed a shift from hypo to hyper DNA methylation during neural maturation. DNA methylation at cytosines mediated by methyltransferases such as Dnmt family is a very dynamic phenomenon across development. Interestingly, while most methylation in the embryo occurs at CpG, during maturation neurons present significant methylation at non-CpG sites[244]. Methylation of the DNA is a powerful mechanism of epigenetic repression due to its recognition by methyl-CpG-binding domain (MBD) proteins that in turn recruit corepressors such histone deacethylases (HDACs)[245]. In CNS development, one of most studied MBDs is methyl-CpG-binding protein 2 (MeCP2) as its mutation causes Rett syndrome, a progressive neurodevelopmental disorder and one of the most common causes of mental retardation in females[246]. Its role in repression has been widely studied[247]. For example, its phosphorylation after synaptic activity allows the derepression of the *Bdnf* gene[248].

One of the latest breakthroughs in neuronal epigenetics has been the discovery of an enrichment of 5-hydroxymethylcytosine (5hmC) in the developing CNS[249,250]. Previously thought to be just an intermediate in the demethylation processes mediated by TET family proteins, it has been shown to play a physiological role in neurons. More specifically, 5hmC increases during neuronal maturation accumulating at active neural-specific gene bodies[250]. Fascinatingly, MeCP2 Rett mutations impair the binding to 5hmC rather than to 5mC thus revealing a new role for MeCP2 in transcription

activation[251]. With the previous observation that MeCP2 competes with linker histone H1*in vitro*[252], and that MeCP2 is nearly as abundant as the histone octamer only in neurons[253], a model has been postulated where MeCP2 largely substitutes for histone H1 and is distributed throughout the genome; at heterochromatin, acting as a suppressor but also at euchromatin, where its binding to 5hmC would favor transcription[254]. Hence, MeCP2 seem to be a central epigenetic regulator during neuron maturation.

Nonetheless, a transcription factor response during neuronal maturation is missing. While other epigenetic mechanisms like micro RNAs could explain MeCP2 post-transcriptional regulation[255–257], a sequential regulation of transcription factors could be expected, starting from early post-mitotic phases and switching to synaptic active states.

REFERENCES

1. Bruylants, G., Bartik, K. & Reisse, J. Is it Useful to Have a Clear-cut Definition of Life? On the Use of Fuzzy Logic in Prebiotic Chemistry. *Orig. Life Evol. Biosph.* **40,** 137–143 (2010).

2. Popa, R. Necessity, futility and the possibility of defining life are all embedded in its origin as a punctuated-gradualism. *Orig. Life Evol. Biosph.* **40,** 183–190 (2010).

3. Darwin, C. Sexual Selection. *On the Origin of Species* 83–85 (1859).

4. Cloud, P. Major Features of Crustal Evolution. *Trans. Geol. Soc. S. Afr.* (1976).

5. Haldane, J. B. S. The origin of life. *Ration. Annu.* 3–10 (1929).

6. Miller, S. L. A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science (80-.).* **117,** 528–529 (1953).

7. Oró, J. and Kamat, S. S. Amino-acid Synthesis from Hydrogen Cyanide under Possible Primitive Earth Conditions. *Nature* **190,** 442–443 (1961).

8. Russell, M. J. & Hall, A. J. The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J. Geol. Soc. London.* **154,** 377–402 (1997).

9. Feller, G. Cryosphere and Psychrophiles: Insights into a Cold Origin of Life? *Life* **7,** 25 (2017).

10. Jheeta, S. The Landscape of the Emergence of Life. *Life* **20,** (2016).

11. Lanier, K. A. & Williams, L. D. The Origin of Life: Models and Data. *J. Mol. Evol.* **84,** 85–92 (2017).

12. Dawkins, R. *Selfish Gene*. (Oxford University Press, 2006).

13. Lazcano, A., Guerrero, R., Margulis, L. & Oró, J. The evolutionary transition from RNA to DNA in early cells. *J. Mol. Evol.* **27,** 283–290 (1988).

14. Crick, F. On Protein Synthesis. *Symp. Soc. Exp. Biol.* 138–166 (1958).

15. Miescher-Rüsch, F. *Ueber die chemische Zusammensetzung der Eiterzellen.* (1871).

16. Franklin, R. E. and Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **171,** 740–741 (1953).

17. Watson, J. D. and Crick, F. H. C. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* 4356 (1953).

18. Crick, F. H. C., Barnett, Leslie Brenner, S. and Watts-Tobin, T. J. General Nature of the Genetic Code for Proteins. *Nature* **192,** 1227–1232 (1961).

19. Szathmáry, E. Opinion: Why are there four letters in the genetic alphabet? *Nat. Rev. Genet.* **4,** 995–1001 (2003).

20. Alberts, B. et al. Molecular Biology of the Cell. in *Molecular Biology of the Cell* (Garland Science, 1994).

21. Hurwitz, J. The discovery of RNA polymerase. *J. Biol. Chem.* **280,** 42477–42485 (2005).

22. Jandura, A. & Krause, H. M. The New RNA World: Growing Evidence for Long Noncoding RNA Functionality. *Trends Genet.* **33,** 665–676 (2017).

23. Higgs, P. G. & Lehman, N. The RNA World: molecular cooperation at the origins of life. *Nat. Rev. Genet.* **16,** 7–17 (2014).

24. Mulder, G. 'Ueber die Zusammensetzung einiger thierischen Substanzen'. *J. für Prakt. Chemie* **16,** 129–152. (1839).

25. Koonin, E. V. & Wolf, Y. I. Genomics of bacteria and archaea: The emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **36,** 6688–6719 (2008).

26. Zuckerkandl, E. The appearance of new structures and functions in proteins during evolution. *J Mol Evol* **7,** 1–57 (1975).

27. Lynch, M. & Marinov, G. K. The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci.* **112,** 201514974 (2015).

28. Burgess, R. R., Travers, A. A., Dunn, J. J. & Bautz, E. K. F. Factor Stimulating Transcription by RNA Polymerase. *Nature* **221,** 43–46 (1969).

29. Koster, M. J. E., Snel, B. & Timmers, H. T. M. Genesis of chromatin and transcription dynamics in the origin of species. *Cell* **161,** 724–736 (2015).

30. Charoensawan, V., Wilson, D. & Teichmann, S. A. Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucleic Acids Res.* **38,** 7364–77 (2010).

31. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 318–356 (1961).

32. Zubay, G., Schwartz, D. & Beckwith, J. Mechanism of activation of catabolite-sensitive genes: a positive control system. *Proc. Natl. Acad. Sci. U. S. A.* **66,** 104–110 (1970).

33. Bondi, A. B. Characteristics of scalability and their impact on performance. *Proc. Second Int. Work. Softw. Perform. - WOSP '00* 195–203 (2000). doi:10.1145/350391.350432

34. Luijsterburg, M. S., White, M. F., van Driel, R. & Dame, R. T. The Major Architects of Chromatin: Architectural Proteins in Bacteria, Archaea and Eukaryotes. *Crit. Rev. Biochem. Mol. Biol.* **43,** 393–418 (2008).

35. Van Nimwegen, E. Scaling laws in the functional content of genomes. *Trends Genet.* **19,** 479–484 (2003).

36. Malik, H. S. & Henikoff, S. Phylogenomics of the nucleosome. *Nat. Struct. Biol.* **10,** 882–891 (2003).

37. Hewish., D. R. & Burgoyn, L. A. Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem. Biophys. Res. Commun.* **52,** 504–510 (1973).

38. Olins, A. L. & Olins, D. E. Spheroid Chromatin Units (ngr Bodies). *Science (80-.).* **183,** 330–332 (1974).

39. Kornberg, R. Chromatin Structure◉: A Repeating Unit of Histones and DNA Chromatin structure is based on a repeating unit of eight. *Science (80-.).* **184,** 868–871 (1974).

40. Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* **16,** 178–189 (2015).

41. Reeve, J. N. Archaeal chromatin and transcription. *Mol. Microbiol.* **48,** 587–598 (2003).

42. Barth, T. K. & Imhof, A. Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem. Sci.* **35,** 618–626 (2010).

43. Brownell, J. E. *et al.* Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84,** 843–851 (1996).

44. Emil, H. Das Heterochromatin der Moose. *Jahrbücher für Wissenschaftliche Bot.* **69,** 762–818 (1928).

45. Saksouk, N., Simboeck, E. & Déjardin, J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin* **8,** 3 (2015).

46. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403,** 41–5 (2000).

47. Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: A Landscape Takes Shape. *Cell* **128,** 635–638 (2007).

48. Delsemme, A. The origin of the atmosphere and of the oceans. *Adv. Astrobiol. Biogeophysics* 29–68 (2006). doi:10.1007/3-540-33088-7_2

49. King, N. The unicellular ancestry of animal development. *Dev. Cell* **7,** 313–325 (2004).

50. Sebé-Pedrós, A., Degnan, B. M. & Ruiz-Trillo, I. The origin of Metazoa, a unicellular perspective. *Nat. Rev. Genet.* **in press,** (2017).

51. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144,** 327–339 (2011).

52. Dao, L. T. M. *et al.* Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat. Genet.* **49,** 1073–1081 (2017).

53. Le, T. B. K., Imakaev, M. V, Mirny, L. a & Laub, M. T. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* **342,** 731–4 (2013).

54. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).

55. Hofmann, A. & Heermann, D. W. The role of loops on the order of eukaryotes and prokaryotes. *FEBS Lett.* **589,** 2958–2965 (2015).

56. Kolovos, P., Knoch, T. A., Grosveld, F. G., Cook, P. R. & Papantonis, A. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin* **5,** 1 (2012).

57. Gaszner, M. & Felsenfeld, G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* **7,** 703–713 (2006).

58. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153,** 1134–1148 (2013).

59. Benayoun, B. A. *et al.* H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell* **158,** 673–688 (2014).

60. Kioussis, D., Vanin, E., DeLange, T., Flavell, R. A. & Grosveld, F. G. Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature* **306,** 662–6 (1983).

61. Song, L. *et al.* Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* **21,** 1757–67 (2011).

62. Parker, S. C. J. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci.* **110,** 17921–17926 (2013).

63. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–319 (2013).

64. Bourbon, H. M. Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic Acids Res.* **36,** 3993–4008 (2008).

65. Myers, L. C. & Kornberg, R. D. Mediator of Transcription Regulation. *Annu. Rev. Biochem* **69,** 729–749 (2000).

66. Roger D.Kornberg, Raymond J.Kelleher III, P. M. F. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* **61,** 1209–1215 (1990).

67. Thompson CM, Koleske AJ, Chao DM, Young RA. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73,** 1361–1375 (1993).

68. Kim YJ, Björklund S, Li Y, Sayre MH, Kornberg RD. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77,** 599–608 (1994).

69. Thompson, C. M. & Young, R. a. General requirement for RNA polymerase II holoenzymes in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 4587–90 (1995).

70. Svejstrup, J. Q. *et al.* Evidence for a mediator cycle at the initiation of transcription. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 6075–8 (1997).

71. Fondell, J. D., Ge, H. & Roeder, R. G. Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 8329–33 (1996).

72. Asturias FJ, Jiang YW, Myers LC, Gustafsson CM, K. R. Conserved structures of mediator and RNA polymerase II holoenzyme. *Science (80-.).* **283,** 985–87 (1999).

73. Robinson, P. J. *et al.* Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* **166,** 1411–1422.e16 (2016).

74. Tsai, K.-L. *et al.* Mediator structure and rearrangements required for holoenzyme formation. *Nature* **544,** 196–201 (2017).

75. Bourbon, H. M. *et al.* A unified nomenclature for protein subunits of mediator complexes linking transcriptional regulators to RNA polymerase II. *Mol. Cell* **14,** 553–557 (2004).

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

77. Bäckström, S., Elfving, N., Nilsson, R., Wingsle, G. & Björklund, S. Purification of a Plant Mediator from Arabidopsis thaliana Identifies PFT1 as the Med25 Subunit. *Mol. Cell* **26,** 717–729 (2007).

78. Nagulapalli, M., Maji, S., Dwivedi, N., Dahiya, P. & Thakur, J. K. Evolution of disorder in Mediator complex and its functional relevance. *Nucleic Acids Res.* **44,** 1591–1612 (2015).

79. Nishikawa, J. L. *et al.* Inhibiting fungal multidrug resistance by disrupting an activator–Mediator interaction. *Nature* **530,** 485–489 (2016).

80. Jeronimo, C. & Robert, F. The Mediator Complex: At the Nexus of RNA Polymerase II Transcription. *Trends Cell Biol.* **xx,** 1–19 (2017).

81. Nozawa, K., Schneider, T. R. & Cramer, P. Core Mediator structure at 3.4 Å extends model of transcription initiation complex. *Nature* **545,** 248–251 (2017).

82. Paparidis, N. F. dos S., Durvale, M. C. & Canduri, F. The emerging picture of CDK9/P-TEFb: more than 20 years of advances since PITALRE. *Mol. BioSyst.* **13,** 246–276 (2017).

83. Takahashi, H. *et al.* Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell* **146,** 92–104 (2011).

84. Lens, Z. *et al.* Solution structure of the N-terminal domain of Mediator subunit MED26 and molecular characterization of its interaction with EAF1 and TAF7. *J. Mol. Biol.* **429,** 3043–3055 (2017).

85. Conaway, R. C. & Conaway, J. W. The Mediator complex and transcription elongation. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829,** 69–75 (2013).

86. Mukundan, B. & Ansari, A. Novel role for mediator complex subunit Srb5/Med18 in termination of transcription. *J. Biol. Chem.* **286,** 37053–37057 (2011).

87. Lai, Z. *et al.* MED18 interaction with distinct transcription factors regulates multiple plant functions. *Nat. Commun.* **5,** 1–14 (2014).

88. Huang, Y., Yao, X. & Wang, G. 'Mediator-ing' messenger RNA processing. *Wiley Interdiscip. Rev. RNA* **6,** 257–269 (2015).

89. Yin, J. -w. & Wang, G. The Mediator complex: a master coordinator of transcription and cell lineage development. *Development* **141,** 977–987 (2014).

90. Petrenko, N., Jin, Y., Wong, K. H. & Struhl, K. Mediator Undergoes a Compositional Change during Transcriptional Activation. *Mol. Cell* **64,** 443–454 (2016).

91. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Y. R. Mediator and Cohesin Connect Gene Expression and Chromatin Architecture. *Nature* **467,** 430–435 (2010).

92. Mukundan, B. & Ansari, A. Srb5/Med18-mediated termination of transcription is dependent on gene looping. *J. Biol. Chem.* **288,** 11384–11394 (2013).

93. Chereji, R. V. *et al.* Mediator binds to boundaries of chromosomal interaction domains and to proteins involved in DNA looping, RNA metabolism, chromatin remodeling, and actin assembly. *Nucleic Acids Res.* **45,** 8806–8821 (2017).

94. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494,** 497–501 (2013).

95. Lin, J. J. *et al.* Mediator coordinates PIC assembly with recruitment of CHD1. *Genes Dev.* **25,** 2198–2209 (2011).

96. Ebmeier, C. C. & Taatjes, D. J. Activator-Mediator binding regulates Mediator-cofactor interactions. *Proc. Natl. Acad. Sci.* **107,** 11283–11288 (2010).

97. Lemieux, K. & Gaudreau, L. Targeting of Swi/Snf to the yeast GAL1 UAS G requires the Mediator, TAF IIs, and RNA polymerase II. *EMBO J.* **23,** 4040–50 (2004).

98. Huttlin, E. L. *et al.* Architecture of the human interactome defines protein communities and disease networks. *Nature* **545,** 505–509 (2017).

99. Meyer, K. D., Lin, S., Bernecky, C., Gao, Y. & Taatjes, D. J. p53 activates transcription by directing structural shifts in Mediator. *Nat. Struct. Mol. Biol.* **17,** 753–760 (2010).

100. Olsen, J. V. *et al.* Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* **127,** 635–648 (2006).

101. Belakavadi, M., Pandey, P. K., Vijayvargia, R. & Fondell, J. D. MED1 phosphorylation promotes its association with mediator: implications for nuclear receptor signaling. *Mol. Cell. Biol.* **28,** 3932–3942 (2008).

102. Chen, Z. *et al.* Phospho-MED1-enhanced UBE2C locus looping drives castration-resistant prostate cancer growth. *EMBO J.* **30,** 2405–2419 (2011).

103. Davis, M. A. *et al.* The SCF-Fbw7 ubiquitin ligase degrades MED13 and MED13L and regulates CDK8 module association with Mediator. *Genes Dev.* **27,** 151–156 (2013).

104. Xu, W. *et al.* R ESEARCH A RTICLES A Transcriptional Switch Mediated by Cofactor Methylation. *Science (80-.).* **294,** 2507–2511 (2001).

105. Wang, L. *et al.* MED12 methylation by CARM1 sensitizes human breast cancer cells to chemotherapy drugs. *Sci. Adv.* **1,** e1500463 (2015).

106. Gonzalez, D. *et al.* Suppression of Mediator is regulated by Cdk8-dependent Grr1 turnover of the Med3 coactivator. *Proc. Natl. Acad. Sci.* **111,** 2500–2505 (2014).

107. Ishikawa, H., Tachikawa, H., Miura, Y. & Takahashi, N. TRIM11 binds to and destabilizes a key component of the activator-mediated cofactor complex (ARC105) through the ubiquitin-proteasome system. *FEBS Lett.* **580,** 4784–4792 (2006).

108. Taatjes, D. J., Na, A. M., Nogales, E. & Tjian, R. Structure , Function , and Conformations of the CRSP Coactivator. *Science (80-.).* **295,** 1058–1062 (2002).

109. Holstege, F. C. . *et al.* Dissecting the Regulatory Circuitry of a Eukaryotic Genome. *Cell* **95,** 717–728 (1998).

110. Lenstra, T. L. *et al.* The Specificity and Topology of Chromatin Interaction Pathways in Yeast. *Mol. Cell* **42,** 536–549 (2011).

111. Van De Peppel, J. *et al.* Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. *Mol. Cell* **19,** 511–522 (2005).

112. Fryer, C. J., White, J. B. & Jones, K. A. Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover. *Mol. Cell* **16,** 509–520 (2004).

113. Ding, N. *et al.* Mediator Links Epigenetic Silencing of Neuronal Gene Expression with X-Linked Mental Retardation. *Mol. Cell* **31,** 347–359 (2008).

114. Tsutsui, T. *et al.* Mediator complex recruits epigenetic regulators via its two cyclin-dependent kinase subunits to repress transcription of immune response genes. *J. Biol. Chem.* **288,** 20955–20965 (2013).

115. Papadopoulou, T., Kaymak, A., Sayols, S. & Richly, H. Dual role of Med12 in PRC1-dependent gene repression and ncRNA-mediated transcriptional activation. *Cell Cycle* **15,** 1479–1493 (2016).

116. Marr, S. K., Lis, J. T., Treisman, J. E. & Marr, M. T. The Metazoan-Specific Mediator Subunit 26 (Med26) Is Essential for Viability and Is Found at both Active Genes and Pericentric Heterochromatin in Drosophila melanogaster. *Mol. Cell. Biol.* **34,** 2710–2720 (2014).

117. Carlsten, J. O. *et al.* Mediator Promotes CENP-A Incorporation at Fission Yeast Centromeres. *Mol. Cell. Biol.* **32,** 4035–4043 (2012).

118. Peng, J. & Zhou, J. Q. The tail-module of yeast Mediator complex is required for telomere heterochromatin maintenance. *Nucleic Acids Res.* **40,** 581–593 (2012).

119. Eyboulet, F. *et al.* Mediator links transcription and DNA repair by facilitating Rad2 / XPG recruitment. 2549–2562 (2013). doi:10.1101/gad.225813.113.associated

120. Ding, N. *et al.* Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation. *Mol. Cell* **31,** 347–59 (2008).

121. Ge, K. *et al.* Alternative mechanisms by which mediator subunit MED1/TRAP220 regulates peroxisome proliferator-activated receptor gamma-stimulated adipogenesis and target gene expression. *Mol. Cell. Biol.* **28,** 1081–91 (2008).

122. Grøntved, L., Madsen, M. S., Boergesen, M., Roeder, R. G. & Mandrup, S. MED14 tethers mediator to the N-terminal domain of peroxisome proliferator-activated receptor gamma and is required for full transcriptional activity and adipogenesis. *Mol. Cell. Biol.* **30,** 2155–2169 (2010).

123. Crawford, S. E. *et al.* Defects of the heart, eye, and megakaryocytes in peroxisome proliferator activator receptor-binding protein (PBP) null embryos implicate GATA family of transcription factors. *J. Biol. Chem.* **277,** 3585–3592 (2002).

124. Stumpf, M. *et al.* The mediator complex functions as a coactivator for GATA-1 in erythropoiesis via subunit Med1/TRAP220. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 18504–9 (2006).

125. Kato, Y., Habas, R., Katsuyama, Y., Näär, A. M. & He, X. A component of the ARC/Mediator complex required for TGFβ/Nodal signalling. *Nature* **418,** 641–646 (2002).

126. Zhou, H. *et al.* MED12 mutations link intellectual disability syndromes with dysregulated GLI3-dependent Sonic Hedgehog signaling. *Proc. Natl. Acad. Sci.* **109,** 19763–19768 (2012).

127. Nakamura, Y. *et al.* Wwp2 is essential for palatogenesis mediated by the interaction between Sox9 and mediator subunit 25. *Nat. Commun.* **2,** 251 (2011).

128. Vogl, M. R. *et al.* Sox10 Cooperates with the Mediator Subunit 12 during Terminal Differentiation of Myelinating Glia. *J. Neurosci.* **33,** 6679–6690 (2013).

129. Liu, Z. *et al.* Mediator MED23 cooperates with RUNX2 to drive osteoblast differentiation and bone development. *Nat. Commun.* **7,** 11149 (2016).

130. Malik, N., Agarwal, P. & Tyagi, A. Emerging functions of multi-protein complex Mediator with special emphasis on plants. *Crit. Rev. Biochem. Mol. Biol.* **52,** 475–502 (2017).

131. Spaeth, J. M., Kim, N. H. & Boyer, T. G. Mediator and human disease. *Semin. Cell Dev. Biol.* **22,** 776–787 (2011).

132. MJ. Lyons. *MED12-Related Disorders*. *GeneReviews [Internet]* (2008).

133. Kaufmann, R. *et al.* Infantile cerebral and cerebellar atrophy is associated with a mutation in the MED17 subunit of the transcription preinitiation mediator complex. *Am. J. Hum. Genet.* **87,** 667–670 (2010).

134. Mukhopadhyay, A. *et al.* CDK19 is disrupted in a female patient with bilateral congenital retinal folds, microcephaly and mild mental retardation. *Hum. Genet.* **128,** 281–291 (2010).

135. Leal, A. *et al.* Identification of the variant Ala335Val of MED25 as responsible for CMT2B2: Molecular data, functional studies of the SH3 recognition motif and correlation between wild-type MED25 and PMP22 RNA levels in CMT1A animal models. *Neurogenetics* **10,** 275–287 (2009).

136. Basel-Vanagaite, L. *et al.* Homozygous MED25 mutation implicated in eye–intellectual disability syndrome. *Hum. Genet.* **134,** 577–587 (2015).

137. Hashimoto, S. *et al.* MED23 mutation links intellectual disability to dysregulation of immediate early gene expression. *Science* **333,** 1161–3 (2011).

138. Adegbola, A. *et al.* Redefining the MED13L syndrome. *Eur. J. Hum. Genet.* **23,** 1308–1317 (2015).

139. Chen, C. P. *et al.* Chromosome 22q11.2 deletion syndrome: Prenatal diagnosis, array comparative genomic hybridization characterization using uncultured amniocytes and literature review. *Gene* **527,** 405–409 (2013).

140. Schiano, C. *et al.* Involvement of Mediator complex in malignancy. *Biochim. Biophys. Acta - Rev. Cancer* **1845,** 66–83 (2014).

141. Syring, I. *et al.* Comprehensive analysis of the transcriptional profile of the Mediator complex across human cancer types. *Oncotarget* **7,** 23043–23055 (2016).

142. Chen, W. & Roeder, R. G. Mediator-dependent nuclear receptor function. *Semin. Cell Dev. Biol.* **22,** 749–758 (2011).

143. Vijayvargia, R., May, M. S. & Fondell, J. D. A coregulatory role for the mediator complex in prostate cancer cell proliferation and gene expression. *Cancer Res.* **67,** 4034–4041 (2007).

144. Hasegawa, N. *et al.* Mediator subunits MED1 and MED24 cooperatively contribute to pubertal mammary gland development and growth of breast carcinoma cells. *Mol. Cell. Biol.* **32,** 1483–95 (2012).

145. Rocha, P. P., Scholze, M., Bleiss, W. & Schrewe, H. Med12 is essential for early mouse development and for canonical Wnt and Wnt/PCP signaling. *Development* **137,** 2723–2731 (2010).

146. Al-Hendy, A. *et al.* Silencing Med12 gene reduces proliferation of human leiomyoma cells mediated via Wnt/β-catenin signaling pathway. *Endocrinology* **158,** 592–603 (2017).

147. Yuan, H. *et al.* Knockdown of mediator subunit Med19 suppresses bladder cancer cell proliferation and migration by downregulating Wnt/β-catenin signalling pathway. *J. Cell. Mol. Med.* **XX,** 1–10 (2017).

148. Firestein, R. *et al.* CDK8 is a colorectal cancer oncogene that regulates β-catenin activity. *Nature* **455,** 547–551 (2008).

149. Mallinger, A. *et al.* Discovery of Potent, Selective, and Orally Bioavailable Small-Molecule Modulators of the Mediator Complex-Associated Kinases CDK8 and CDK19. *J. Med. Chem.* **59,** 1078–1101 (2016).

150. Rokas, A. The Origins of Multicellularity and the Early History of the Genetic Toolkit For Animal Development. *Annu. Rev. Genet.* **42,** 235–251 (2008).

151. Nielsen, C. Six major steps in animal evolution: Are we derived sponge larvae? *Evol. Dev.* **10,** 241–257 (2008).

152. Singla, V. The Primary Cilium as the Cell's Antenna: Signaling at a Sensory Organelle. *Science (80-.).* **313,** 629–633 (2006).

153. Monk, T. & Paulin, M. G. Predation and the origin of neurones. *Brain. Behav. Evol.* **84,** 246–261 (2014).

154. Dale Purves, George J Augustine, David Fitzpatrick, Lawrence C Katz, Anthony-Samuel LaMantia, James O McNamara, and S. M. W. *Neuroscience*. (Sinauer Associates, 2001).

155. Burkhardt, P. The origin and evolution of synaptic proteins - choanoflagellates lead the way. *J. Exp. Biol.* **218,** 506–514 (2015).

156. Holland, N. D. Early central nervous system evolution: an era of skin brains? *Nat. Rev. Neurosci.* **4,** 617–627 (2003).

157. Hobert, O. A map of terminal regulators of neuronal identity in Caenorhabditis elegans. *Wiley Interdiscip. Rev. Dev. Biol.* **5,** 474–498 (2016).

158. Zhang, J. Evolution by gene duplication: An update. *Trends Ecol. Evol.* **18,** 292–298 (2003).

159. Hrycaj, S. M. & Wellik, D. M. Hox genes and evolution. *F1000Research* **5,** 859 (2016).

160. McLysaght, A., Hokamp, K. & Wolfe, K. H. Extensive genomic duplication during early chordate evolution. *Nat. Genet.* **31,** 200–204 (2002).

161. Northcutt, R. G. Understanding vertebrate brain evolution. *Integr Comp Biol* **42,** 743–756 (2002).

162. Henrique Franco, N. Animal experiments in biomedical research: A historical perspective. *Animals* **3,** 238–273 (2013).

163. EU. Report From the Commission To the Council and the European Parliament. (2014).

164. Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).

165. Eppig, J. T., Motenko, H., Richardson, J. E., Richards-Smith, B. & Smith, C. L. The International Mouse Strain Resource (IMSR): cataloging worldwide mouse and ES cell line resources. *Mamm. Genome* **26,** 448–455 (2015).

166. Nelson, A. L., Dhimolea, E. & Reichert, J. M. Development trends for human monoclonal antibody therapeutics. *Nat. Rev. Drug Discov.* **9,** 767–774 (2010).

167. Northcutt, R. G. & Kaas, J. H. The emergence and evolution of mammalian neocortex. *Trends Neurosci.* **18,** 373–9 (1995).

168. Goolam, M. *et al.* Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos. *Cell* **165,** 61–74 (2016).

169. M. Evans and M. Kaufman. Establishment in culture of pluripotential cells from mouse embryos. *Nature* 154 (1981).

170. Ying, Q. L. & Smith, A. The Art of Capturing Pluripotency: Creating the Right Culture. *Stem Cell Reports* **8,** 1457–1464 (2017).

171. Bradley A, Evans M, Kaufman MH, R. E. Formation of germ-line chimaeras from embryoderived teratocarcinoma cell lines. *Nature* 255–256 (1984).

172. Mahla, R. S. Stem cells applications in regenerative medicine and disease therapeutics. *Int. J. Cell Biol.* **2016,** (2016).

173. Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149,** 590–604 (2012).

174. van den Berg, D. L. C. *et al.* An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells. *Cell Stem Cell* **6,** 369–381 (2010).

175. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155,** (2013).

176. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126,** 663–676 (2006).

177. Medvedev, S. P., Shevchenko, A. I. & Zakian, S. M. Induced Pluripotent Stem Cells: Problems and Advantages when Applying them in Regenerative Medicine. *Acta Naturae* **2,** 18–28 (2010).

178. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125,** 315–326 (2006).

179. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441,** 349–353 (2006).

180. Oguro, H. *et al.* Poised Lineage Specification in Multipotential Hematopoietic Stem and Progenitor Cells by the Polycomb Protein Bmi1. *Cell Stem Cell* **6,** 279–286 (2010).

181. Kojima, Y., Tam, O. H. & Tam, P. P. L. Timing of developmental events in the early mouse embryo. *Semin. Cell Dev. Biol.* **34,** 65–75 (2014).

182. Greene, N. D. E. & Copp, A. J. Development of the vertebrate central nervous system: formation of the neural tube. *Prenat. Diagn.* **29,** 303–311 (2009).

183. Streit, A. *et al.* Chordin regulates primitive streak development and the stability of induced neural cells, but is not sufficient for neural induction in the chick embryo. *Development* **125,** 507–519 (1998).

184. Streit, A., Berliner, A. J., Papanayotou, C., Sirulnik, A. & Stern, C. D. Initiation of neural induction by FGF signalling before gastrulation. *Nature* **406,** 74–78 (2000).

185. Wilson, S. I. *et al.* The status of Wnt signalling regulates neural and epidermal fates in the chick embryo. *Nature* **411,** 325–330 (2001).

186. Suzuki, I. K. & Vanderhaeghen, P. Is this a brain which I see before me? Modeling human neural development with pluripotent stem cells. *Development* **142,** 3138–3150 (2015).

187. Ruiz I Altaba, A. Combinatorial Gli gene function in floor plate and neuronal inductions by Sonic hedgehog. *Development* **125,** 2203–2212 (1998).

188. Aruga, J. The role of Zic genes in neural development. *Mol. Cell. Neurosci.* **26,** 205–221 (2004).

189. Houtmeyers, R., Souopgui, J., Tejpar, S. & Arkell, R. The ZIC gene family encodes multi-functional proteins essential for patterning and morphogenesis. *Cell. Mol. Life Sci.* **70,** 3791–811 (2013).

190. Conidi, A. *et al.* Few Smad proteins and many Smad-interacting proteins yield multiple functions and action modes in TGFβ/BMP signaling in vivo. *Cytokine Growth Factor Rev.* **22,** 287–300 (2011).

191. Hegarty, S. V, Sullivan, A. M. & O'Keeffe, G. W. Zeb2: A multifunctional regulator of nervous system development. *Prog. Neurobiol.* **132,** 81–95 (2015).

192. Ballas, N., Grunseich, C., Lu, D. D., Speh, J. C. & Mandel, G. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell* **121,** 645–657 (2005).

193. Du, Z.-W., Ma, L.-X., Phillips, C. & Zhang, S.-C. miR-200 and miR-96 families repress neural induction from human embryonic stem cells. *Development* **140,** 2611–2618 (2013).

194. Burgold, T. *et al.* The histone H3 lysine 27-specific demethylase Jmjd3 is required for neural commitment. *PLoS One* **3,** (2008).

195. Wegner, M. SOX after SOX: SOXession regulates neurogenesis. *Genes Dev.* **25,** 2423–2428 (2011).

196. Pevny, L. H., Sockanathan, S., Placzek, M. & Lovell-Badge, R. A role for SOX1 in neural determination. *Development* **125,** 1967–78 (1998).

197. Homem, C. C. F., Repic, M. & Knoblich, J. A. Proliferation control in neural stem and progenitor cells. *Nat Rev. Neurosicence* **16,** 647–659 (2015).

198. Conti, L. *et al.* Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.* **3,** 1594–1606 (2005).

199. Engelen, E. *et al.* Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat. Genet.* **43,** 607–611 (2011).

200. Moen, M. J. *et al.* An interaction network of mental disorder proteins in neural stem cells. *Transl. Psychiatry* **7,** e1082 (2017).

201. Mateo, J. L. *et al.* Characterisation of the neural stem cell gene regulatory network identifies OLIG2 as a multi-functional regulator of self-renewal. *Genome Res.* **in press,** 41–56 (2014).

202. Bertrand, N., Castro, D. S. & Guillemot, F. Proneural genes and the specification of neural cell types. *Nat. Rev. Neurosci.* **3,** 517–530 (2002).

203. Imayoshi, I. *et al.* Oscillatory Control of Factors Determining Multipotency and Fate in Mouse Neural Progenitors. *Science (80-.).* **342,** 1203–1208 (2013).

204. Pinto, L. & Götz, M. Radial glial cell heterogeneity-The source of diverse progeny in the CNS. *Prog. Neurobiol.* **83,** 2–23 (2007).

205. Götz, M. & Huttner, W. B. The cell biology of neurogenesis. *Nat. Rev. Mol. Cell Biol.* **6,** 777–788 (2005).

206. Hatakeyama, J. Hes genes regulate size, shape and histogenesis of the nervous system by control of the timing of neural stem cell differentiation. *Development* **131,** 5539–5550 (2004).

207. Sahara, S. & O'Leary, D. D. M. Fgf10 Regulates Transition Period of Cortical Stem Cell Differentiation to Radial Glia Controlling Generation of Neurons and Basal Progenitors. *Neuron* **63,** 48–62 (2009).

208. Bond, A. The dynamic role of bone morphogenetic proteins in neural stem cell fate and maturation. *Dev. Neurobiol.* **72,** 1068–1084 (2012).

209. Siegenthaler, J. A. *et al.* Retinoic Acid from the Meninges Regulates Cortical Neuron Generation. *Cell* **139,** 597–609 (2009).

210. Heinrich, C. *et al.* Directing astroglia from the cerebral cortex into subtype specific functional neurons. *PLoS Biol.* **8,** (2010).

211. Malik, A. N. *et al.* Genome-wide identification and characterization of functional neuronal activity–dependent enhancers. *Nat. Neurosci.* **17,** 1330–1339 (2014).

212. Guillemot, F. Spatial and temporal specification of neural fates by transcription factor codes. *Development* **134,** 3771–3780 (2007).

213. Castro, D. S. *et al.* A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. *Genes Dev.* **25,** 930–945 (2011).

214. Yoshikawa, K. Cell cycle regulators in neural stem cells and postmitotic neurons. *Neurosci. Res.* **37,** 1–14 (2000).

215. Martín-Ibáñez, R. *et al.* Ikaros-1 couples cell cycle arrest of late striatal precursors with neurogenesis of enkephalinergic neurons. *J. Comp. Neurol.* **518,** 329–351 (2010).

216. Lacomme, M., Liaubet, L., Pituello, F. & Bel-Vialar, S. NEUROG2 drives cell cycle exit of neuronal precursors by specifically repressing a subset of cyclins acting at the G1 and S phases of the cell cycle. *Mol. Cell. Biol.* **32,** 2596–607 (2012).

217. Chen, J. Y., Lin, J. R., Tsai, F. C. & Meyer, T. Dosage of Dyrk1a shifts cells within a p21-Cyclin D1 signaling map to control the decision to enter the cell cycle. *Mol. Cell* **52,** 87–100 (2013).

218. Frotscher, M. Cajal-Retzius cells, Reelin, and the formation of layers. *Curr. Opin. Neurobiol.* **8,** 570–575 (1998).

219. Seuntjens, E. *et al.* Sip1 regulates sequential fate decisions by feedback signaling from postmitotic neurons to progenitors. *Nat. Neurosci.* **12,** 1373–80 (2009).

220. Hu, J. S., Vogt, D., Sandberg, M. & Rubenstein, J. L. Cortical interneuron development: a tale of time and space. *Development* **144,** 3867–3878 (2017).

221. Sun, X. Z. *et al.* Normal and abnormal neuronal migration in the developing cerebral cortex. *J Med.Invest* **49,** 97–110 (2002).

222. Peyre, E., Silva, C. G. & Nguyen, L. Crosstalk between intracellular and extracellular signals regulating interneuron production, migration and integration into the cortex. *Front. Cell. Neurosci.* **9,** 1–18 (2015).

223. Baillat, D. & Wagner, E. J. Integrator: Surprisingly diverse functions in gene expression. *Trends Biochem. Sci.* **40,** 257–264 (2015).

224. van den Berg, D. L. C. *et al.* Nipbl Interacts with Zfp609 and the Integrator Complex to Regulate Cortical Neuron Migration. *Neuron* **93,** 348–361 (2017).

225. Yamaguchi, Y. & Miura, M. Programmed cell death in neurodevelopment. *Dev. Cell* **32,** 478–490 (2015).

226. Molyneaux, B. J., Arlotta, P., Menezes, J. R. L. & Macklis, J. D. Neuronal subtype specification in the cerebral cortex. *Nat. Rev. Neurosci.* **8,** 427–437 (2007).

227. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19,** 335–346 (2016).

228. Neuroglia: Definition, Classification, Evolution, Numbers, Development. *Glial Physiol. Pathophysiol.* 73–104 (2013). doi:10.1002/9781118402061.ch3

229. Namihira, M. *et al.* Committed Neuronal Precursors Confer Astrocytic Potential on Residual Neural Precursor Cells. *Dev. Cell* **16,** 245–255 (2009).

230. Weng, Q. *et al.* Dual-mode modulation of Smad signaling by Smad-interacting protein Sip1 is required for myelination in the central nervous system. *Neuron* **73,** 713–28 (2012).

231. *The Hippocampus Book*. (Oxford University Press, 2006).

232. Molyneaux, B. J. *et al.* DeCoN: Genome-wide analysis of invivo transcriptional dynamics during pyramidal neuron fate selection in neocortex. *Neuron* **85,** 275–288 (2015).

233. Takeuchi, T., Duszkiewicz, A. J. & Morris, R. G. M. The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Philos. Trans. R. Soc. B Biol. Sci.* **369,** 20130288–20130288 (2013).

234. West, A. E. & Greenberg, M. E. Neuronal Activity – Regulated Gene Transcription in Synapse Development and Cognitive Function. 1–21 (2011). doi:10.1101/cshperspect.a005744

235. Greer, P. L. & Greenberg, M. E. From Synapse to Nucleus: Calcium-Dependent Gene Transcription in the Control of Synapse Development and Function. *Neuron* **59,** 846–860 (2008).

236. Kowiański, P. *et al.* BDNF: A Key Factor with Multipotent Impact on Brain Signaling and Synaptic Plasticity. *Cell. Mol. Neurobiol.* 1–15 (2017). doi:10.1007/s10571-017-0510-4

237. Cortés-Mendoza, J., Díaz de León-Guerrero, S., Pedraza-Alva, G. & Pérez-Martínez, L. Shaping synaptic plasticity: The role of activity-mediated epigenetic regulation on gene transcription. *Int. J. Dev. Neurosci.* **31,** 359–369 (2013).

238. Bischofberger, J. & Schinder, A. F. Maturation and Functional Integration of New Granule Cells into the Adult Hippocampus. *Cold Spring Harb. Monogr. Arch.* **52,** 299–319 (2016).

239. Montague, P. R. & Friedlander, M. J. Expression of an intrinsic growth strategy by mammalian retinal neurons. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 7223–7 (1989).

240. Metzger, F. Molecular and cellular control of dendrite maturation during brain development. *Curr. Mol. Pharmacol.* **3,** 1–11 (2010).

241. Sances, S. *et al.* Modeling ALS with motor neurons derived from human induced pluripotent stem cells. *Nat. Neurosci.* **16,** 542–553 (2016).

242. Manuelidis, L. Active Nucleolus Organizers Are Precisely Positioned in Adult Central Nervous System Cells but Not in Neuroectodermal Tumor Cells. *J. Neuropathol. Exp. Neurol.* **43,** 225–241 (1984).

243. Pearson, E. C., Bates, D. L., Prospero, T. D. & Thomas, J. O. Neuronal nuclei and glial nuclei from mammalian cerebral cortex: Nucleosome repeat lengths, DNA contents and H1 contents. *Eur. J. Biochem.* **144,** 353–360 (1984).

244. Sharma, A., Klein, S. S., Barboza, L., Lohdi, N. & Toth, M. Principles Governing DNA Methylation during Neuronal Lineage and Subtype Specification. *J. Neurosci.* **36,** 1711–1722 (2016).

245. Moore, L. D., Le, T. & Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **38,** 23–38 (2013).

246. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23,** 185–188 (1999).

247. Chahrour, M. *et al.* MeCP2, a Key Contributor to Neurological Disease, Activates and Represses Transcription. *Science (80-.).* **320,** 1224–1229 (2008).

248. Chen, W. G. Derepression of BDNF Transcription Involves Calcium-Dependent Phosphorylation of MeCP2. *Science (80-.).* **302,** 885–889 (2003).

249. Kriaucionis, S. & Heintz, N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science (80-.).* **324,** 929–930 (2009).

250. Hahn, M. A. *et al.* Dynamics of 5-Hydroxymethylcytosine and Chromatin Marks in Mammalian Neurogenesis. *Cell Rep.* **3,** 291–300 (2013).

251. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151,** 1417–1430 (2012).

252. Ghosh, R. P., Horowitz-Scherer, R. A., Nikitina, T., Shlyakhtenko, L. S. & Woodcock, C. L. MeCP2 Binds Cooperatively to Its Substrate and Competes with Histone H1 for Chromatin Binding Sites. *Mol. Cell. Biol.* **30,** 4656–4670 (2010).

253. Skene, P. J. *et al.* Neuronal MeCP2 Is Expressed at Near Histone-Octamer Levels and Globally Alters the Chromatin State. *Mol. Cell* **37,** 457–468 (2010).

254. Shin, J., Ming, G. L. & Song, H. By hook or by crook: Multifaceted DNA-binding properties of MeCP2. *Cell* **152,** 940–942 (2013).

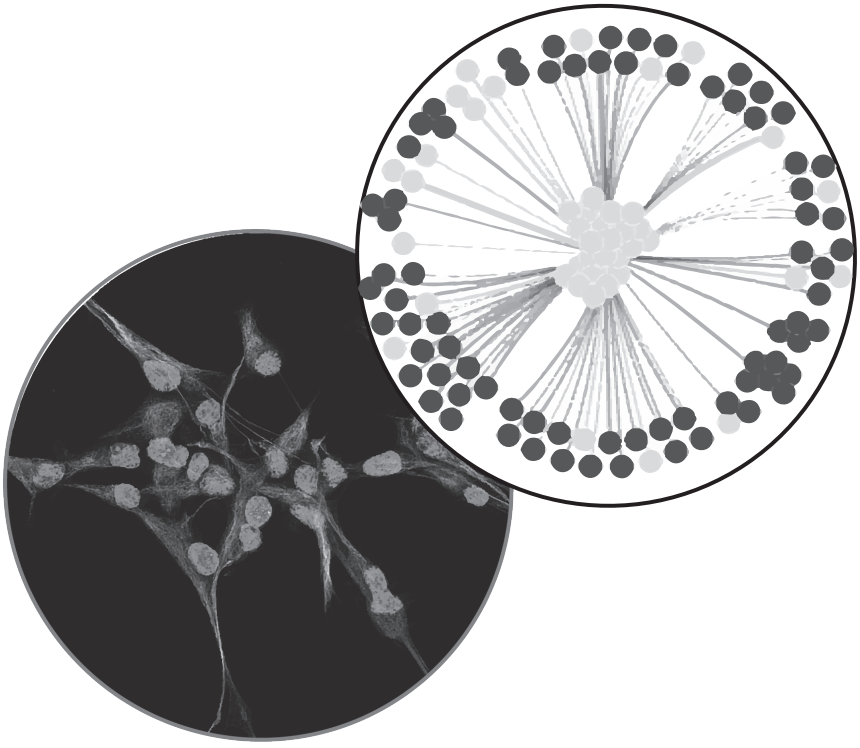
255. Klein, M. E. *et al.* Homeostatic regulation of MeCP2 expression by a CREB-induced microRNA. *Nat. Neurosci.* **10,** 1513–1514 (2007).

256. Lyu, J.-W., Yuan, B., Cheng, T.-L., Qiu, Z.-L. & Zhou, W.-H. Reciprocal regulation of autism-related genes MeCP2 and PTEN via microRNAs. *Sci. Rep.* **6,** 20392 (2016).

257. Zhang, Y. *et al.* MiR-130a regulates neurite outgrowth and dendritic spine density by targeting MeCP2. *Protein Cell* **7,** 489–500 (2016).

Chapter 2

Mediator complex interaction partners organize the transcriptional network that defines neural stem cells



- *Manuscript under revision at Nature Communications* -

Mediator complex interaction partners organize the transcriptional network that defines neural stem cells

Marti Quevedo[1], Mike R. Dekker[1], Dick H.W. Dekkers[2], Johannes H. Brandsma[1], Debbie L.C. van den Berg[1], Zeliha Ozgür[3], Wilfred F.J. van IJcken[3], Jeroen Demmers[2], Maarten Fornerod[1,#], Raymond A. Poot[1,*]

[1]Department of Cell Biology, Erasmus MC, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands
[2]Center for Proteomics, Erasmus MC, The Netherlands
[3]Center for Biomics, Erasmus MC, The Netherlands
[#] Correspondence for computational biology, email: m.fornerod@erasmusmc.nl
[*]Corresponding author

ABSTRACT

Mediator complex regulates transcription by connecting enhancers to promoters. High Mediator binding density defines super enhancers, which regulate cell-identity genes and oncogenes. Protein interactions of Mediator may explain its role in these processes but have not been identified comprehensively. Here we purified Mediator from neural stem cells (NSCs) and identified 75 novel protein-protein interaction partners. We identified super enhancers in NSCs and show that Mediator-interacting chromatin modifiers colocalize with Mediator at enhancers and super enhancers. Transcription factor families with high affinity for Mediator dominate enhancers and super enhancers and can explain genome-wide Mediator localization. We identified E-box transcription factor Tcf4 as a key regulator of NSCs. Tcf4 interacts with Mediator, colocalizes with Mediator at super enhancers and regulates neurogenic transcription factor genes with super enhancers and broad H3K4me3 domains. Our data suggest that high binding-affinity for Mediator is an important organizing feature in the transcriptional network that determines NSC identity.

INTRODUCTION

The Mediator complex is a complex of ~ 30 subunits that is important for transcriptional regulation and is conserved from yeast to human[1–4]. The Mediator complex provides communication between active enhancers and promoters by interacting with proteins that bind to either of these two classes of regulatory DNA elements[4–6]. Accordingly, identified Mediator-interacting proteins include many transcription factors, RNA polymerase II (RNApol2) and transcription elongation factors[7]. Recently, Mediator content was used to rank enhancers in embryonic stem cells (ESCs) and enhancers with the highest Mediator content were postulated as super enhancers (SEs)[8], a class of enhancers that regulates key genes in cell identity and oncogenes[8–10]. Related enhancer types such as stretch enhancers and anti-pause enhancers were described independently[11,12]. There is debate on whether SEs act mechanistically different from typical enhancers[13]. Arguments in favor of the functional distinction of SEs is their ability to drive high levels of transcription and their selective sensitivity to inhibitors of Brd4, a chromatin-binding protein enriched at SEs[10,14,15]. Besides Mediator and Brd4, a number of chromatin modifiers, such as Ep300, Chd7, Smc1a (Cohesin complex) Brg1 (SWI-SNF complex), Chd4 (NuRD complex) and Kdm1a (LSD1 complex), were found to be enriched at SEs[9]. In a recently proposed model, the constituent enhancers of an SE and their regulated promoter(s) would group together to form a phase-separated assembly[16]. Such an assembly would rely on interactions between transcriptional and chromatin regulators[16].

Cell-type specific master TFs colocalise with Mediator at SEs[8,9]. However, evidence for interactions between master TFs and Mediator, which would underpin their role in recruiting Mediator to SEs, is scarce. For example, among SE-binding master TFs Oct4, Sox2 and Nanog (ESCs), Pu.1 (pro-B cells), MyoD (Myotubes) and C/EBPα (Macrophages)[8], Mediator interactions were only detected in immunoprecipitations of Sox2 and C/EBPα and these were with single Mediator subunits[17,18]. Also our understanding of the recruitment of the above chromatin modifiers to enhancers and SEs and their subsequent maintenance at high levels at SEs is far from complete. Mediator was shown to interact with SE-enriched chromatin modifiers Cohesin[19]and Crebbp[20], suggesting that Mediator could, in principle, provide an anchoring role at enhancers, SEs and the proposed phase-separated assemblies.

To investigate the relevance of Mediator interactors in defining enhancers and SEs, we purified the Mediator complex from neural stem cells (NSCs) and identified its protein-protein interaction partners by mass spectrometry. To prevent recording interactions that are mediated via DNA/chromatin, we purified Mediator from non-treated nuclear extracts, nuclear extracts treated with nuclease benzonase and nuclear extracts treated with ethidium bromide to disrupt protein-DNA interactions and only took interactions with Mediator complex that were not affected by these treatments. Our resulting Mediator interactome contains 95 proteins of which 75 are novel Mediator-interacting proteins. Subsequently, we performed Mediator ChIP-seq in NSCs and defined SEs in NSCs by their Mediator content. Remarkably, we find that the three most frequent motifs in SEs are bound by multiple members of the small set of TFs that we identified as Mediator interactors in NSCs.

We show that one of these TFs, Tcf4, regulates a set of key NSC transcription factor genes with SEs and broad H3K4me3 domain-containing promoters. High Mediator affinity therefore appears an important characteristic of master TFs. Our Mediator interactome contains many known enhancer-binding chromatin modifiers and we show that Mediator-interacting chromatin modifiers Jmjd1c and Carm1 bind genome-wide to enhancers and SEs. Together this suggests thathigh Mediator binding affinity selects proteins that play important roles in establishing and maintainingenhancers and SEs to facilitate the regulation of cell identity.

RESULTS

Purification of the Mediator complex from neural stem cells

We generated a mouse neural stem cell line expressing FLAG-tagged Med15 (F-Med15 NSCs) to enable the purification of the Mediator complex by our FLAG-affinity protocol, which combines high efficiency and low background[21] and was extensively validated in the past for accuracy by independent immunoprecipitations of endogenous proteins[21,22]. F-Med15 NSCs and parental NSCs were grown to large scale and nuclear extracts prepared (see Methods). We were interested in proteins that can bind to the Mediator complex relying solely on protein-protein interactions and not being mediated via chromatin, which may co-purify with a chromatin-binding factor such as the Mediator complex. We reasoned that proteins interacting with Mediator by protein-protein interaction would not show a reduced interaction efficiency when treating the nuclear extract with the DNA-RNA digesting enzyme Benzonase or with ethidium bromide (EtBr), which intercalates in the DNA and disrupts protein-DNA interactions, as compared to untreated nuclear extracts (Figure 1A). The used nuclear extract preparation procedure[23] aims to minimize the amount of DNA/chromatin in the extract by gently douncing the nuclei as a method for lysis. Nevertheless, remnants of DNA/chromatin do get released from the nuclei into the extract (Figure 1B, Untreated). Addition of benzonase completely removed chromatin/DNA from the extract. (Figure 1B, compare Benzonase to Untreated). We purified the Mediator complex by FLAG-affinity from nuclear extracts treated with Benzonase, with EtBr or not treated, as well as from parental NSCs as a control. Purified Mediator samples and control samples were analyzed by mass spectrometry to identify the proteins present in these samples. We selected proteins that were specific for Mediator samples and that did not go down in abundance (less than 2-fold drop in emPAI score) when comparing purifications from nuclear extracts treated with Benzonase or EtBr, to purifications from untreated extracts (see Methods). To be included in our final list of Mediator-interacting proteins (Figure 1C, Table S1), selected proteins also had to be specifically present in an independent replicate of the Mediator purification from Benzonase-treated nuclear extract (Table S1).

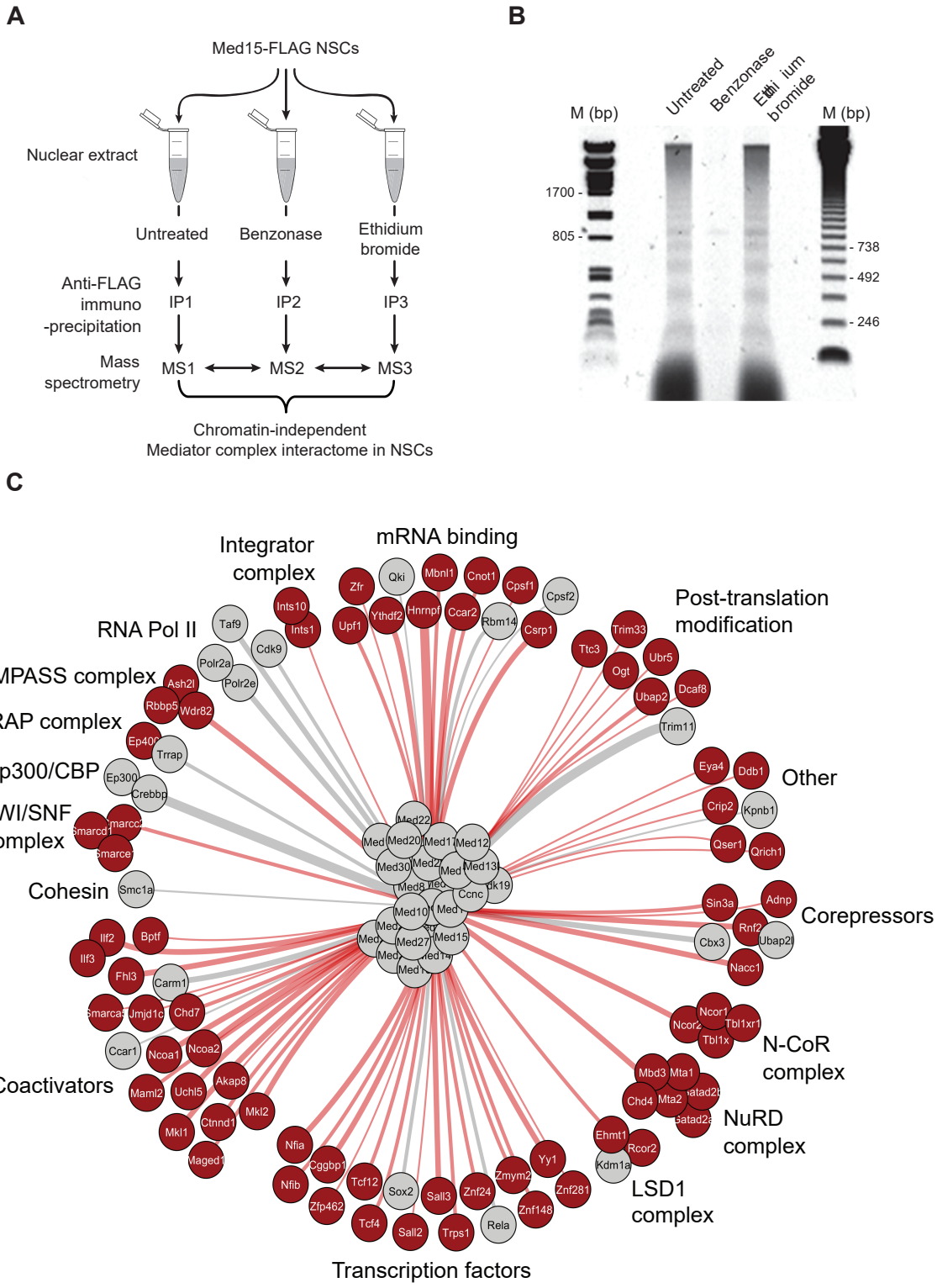


Figure 1 | Mediator complex interactome in neural stem cells

(A) Schematic representation of Mediator complex purifications from neural stem cells (NSCs) expressing Med15-FLAG. Mass spectrometry results of the three conditions were compared to select proteins that do not decrease in abundance upon treatments as chromatin-independent Mediator complex interactors. IP, immunoprecipitation. MS, Mass spectrometry. (B) Agarose gel with DNA from untreated NSC nuclear extract or nuclear extract treated with Benzonase or Ethidium Bromide, as indicated. DNA size markers (M) are indicated. (C) Interactome of the Mediator complex in NSCs. Novel Mediator interaction partners are in red, known Mediator interaction partners are in grey. Thickness of the edges gives an indication of the relative molar protein quantity observed in the purified Mediator complex samples. See also Table S1.

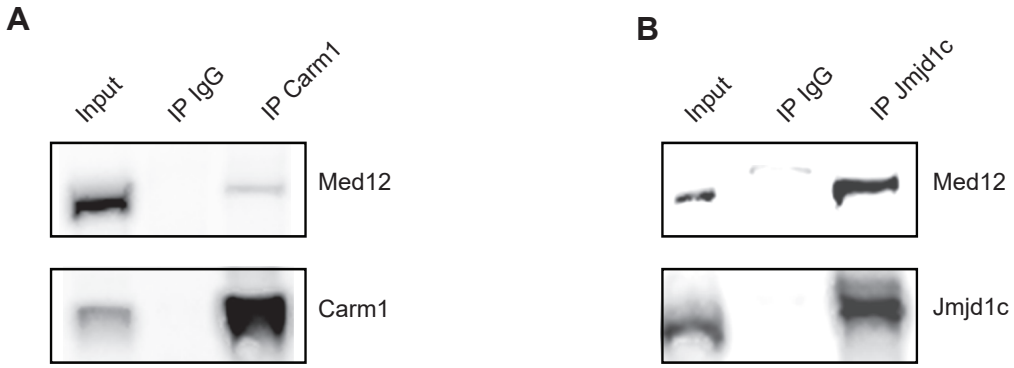
A Mediator interactome in neural stem cells

We identified 119 Med15-interacting proteins from the four FLAG-Med15 purifications (Figure 1C, Table S1), of which 24 proteins are core-subunits of the Mediator complex, leaving 95 proteins that we postulate as Mediator complex-interacting proteins. The vast majority of these Mediator-interacting proteins, 75 proteins, was not previously identified as binding to Mediator and would therefore be novel Mediator interactors (Figure 1C, indicated in red). A number of well-known constituents of enhancers such as Ep300, Chd7, LSD1 complex, NuRD complex and SWI-SNF complex[24–27] were identified as interactors of Mediator (Fig. 1C, Table S1). Cohesin subunit Smc1a[19] was identified, whereas Cohesin subunit Smc3 and Cohesin loader Nipbl were observed in 3 out of 4 Mediator purifications and are therefore not part of the final Mediator interactor list (Table S1). Ep300, Crebbp, Chd7, Kdm1a (LSD1 complex), Chd4 (NuRD complex), Smc1a (Cohesin) and Brg1 (SWI-SNF complex) were recently shown, like Mediator, to have higher binding densities at super enhancers (SEs) in embryonic stem cells, as compared to typical enhancers[9]. Other transcriptional activators and repressors interacting with Mediator included Ncoa1-2, the COMPASS complex, Integrator complex, TRRAP complex and N-CoR complex (Figure 1C). We identified histone demethylase Jmjd1c and arginine demethylase Carm1 as Mediator interactors. Carm1 was recently identified to bind Med9 in a high throughput interaction screen[28]. We independently confirmed the interactions of Jmjd1c and Carm1 with Mediator by reverse co-immunoprecipitations with Carm1 antibodies (Figure 2A) and Jmjd1c antibodies (Figure 2B). Prominent Mediator interactor categories not directly related to transcription were mRNA binding proteins, often functioning in post-transcriptional events, and protein modifiers (Figure 1C).

Mediator has been identified as a co-activator of many DNA sequence-specific transcription factors, often nuclear hormone receptors[29,30]. We identified 16 DNA sequence-specific transcription factors (TFs) of which 14 are novel Mediator interactors (Fig. 1C), including NFI TFs Nfia and Nfib, Sox2 and E-box TFs Tcf4 and Tcf12. The majority of these TFs have an important function in the regulation of NSCs (Figure 2C). To test whether detected Mediator-interacting TFs are the highest expressed TFs in NSC, which could explain their detection by mass spectrometry, we plotted the 16 detected TFs against the 600 highest expressed TFs (by RNA-seq) in our NSCs. We find that Mediator-interacting TFs are not the highest expressed TFs in NSCs (Figure 2D). This suggests that the detection of our Mediator-interacting TFs is primarily related to their high binding-affinity for Mediator, as compared to many other, not detected, TFs.

Brd4 has been shown to strongly colocalise with Mediator at enhancers and promoters. Despite our high sensitivity of detecting Mediator interactors, we did not detect Brd4 in any of our FLAG-Med15 purifications (Table S1 and data not shown). We also did not detect Jmjd6 and Nsd3, functional interaction partners of Brd4[15,31], in any purification.

In conclusion, we expanded the Mediator interactome with many transcription-associated factors and our experimental set-up suggests that these interactions are independent of chromatin.



C

TF name	Function in neural development[a]
Yy1	Enhancer looping in neural stem cells
Nfia	Neural stem cell differentiation and glial precursor self-renewal
Sox2	Neural fate specification and neural stem cell self-renewal
Nfib	Neural stem cell differentiation and glial precursor self-renewal
Sall3	Photoreceptor and retina development and neural stem cell self-renewal
Tcf12	Neural stem cell differentiation
Znf24	Neural stem cell self-renewal
Rela	Neural stem cell self-renewal
Tcf4	Neural stem cell differentiation and neural migration

[a] References in Methods

D

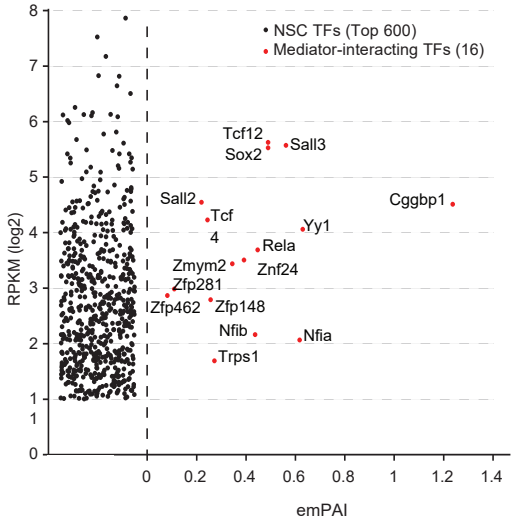


Figure 2 | Mediator complex interactor validation

(A) Immunoprecipitation (IP) of Carm1 and Med12 by a Carm1 antibody from NSC nuclear extract. Western blots are probed with the indicated antibodies. Control IP by rabbit IgG and 5% input are also shown. (B) Immunoprecipitation (IP) of Jmjd1c and Med12 by Jmjd1c antibody from NSC nuclear extract. Western blots are probed with the indicated antibodies. Control IP by rabbit IgG and 5% input are also shown.(C) Function in neural development of identified Mediator-interacting transcription factors in NSCs. References are provided in the Methods.(D) mRNA levels in NSCs of Mediator-interacting transcription factors (TFs) and the Top 600 highest expressed TFs in NSCs. The average emPAI scores, a semi-quantitative mass spectrometry-based measure of molar amounts,in the 4 Mediator complex purifications is shown for Mediator–interacting TFs.

Mediator-based super enhancers in neural stem cells

High Mediator content is a defining feature of so-called super enhancers (SEs)[8]. SEs have not been defined yet in NSCs. We identified SEs in NSCs by ranking NSC enhancers, which were previously defined by the presence of the H3K27ac mark and Ep300[32], by their Med1 ChIP signal using the ROSE algorithm [8,9]. Accordingly, we identified 445 SEs in NSCs and assigned the 9436 remaining enhancers as typical enhancers (Figures 3A and 3B, Table S2). Transcription factors encoded by genes near top SEs include Mediator interactors Nfia, Tcf4, Sox2 and Sall3 (Figure 3B). We find that active genes near SEs (SE genes) in NSCs are, on average, several fold higher expressed than genes near typical enhancers (Figure 3C).

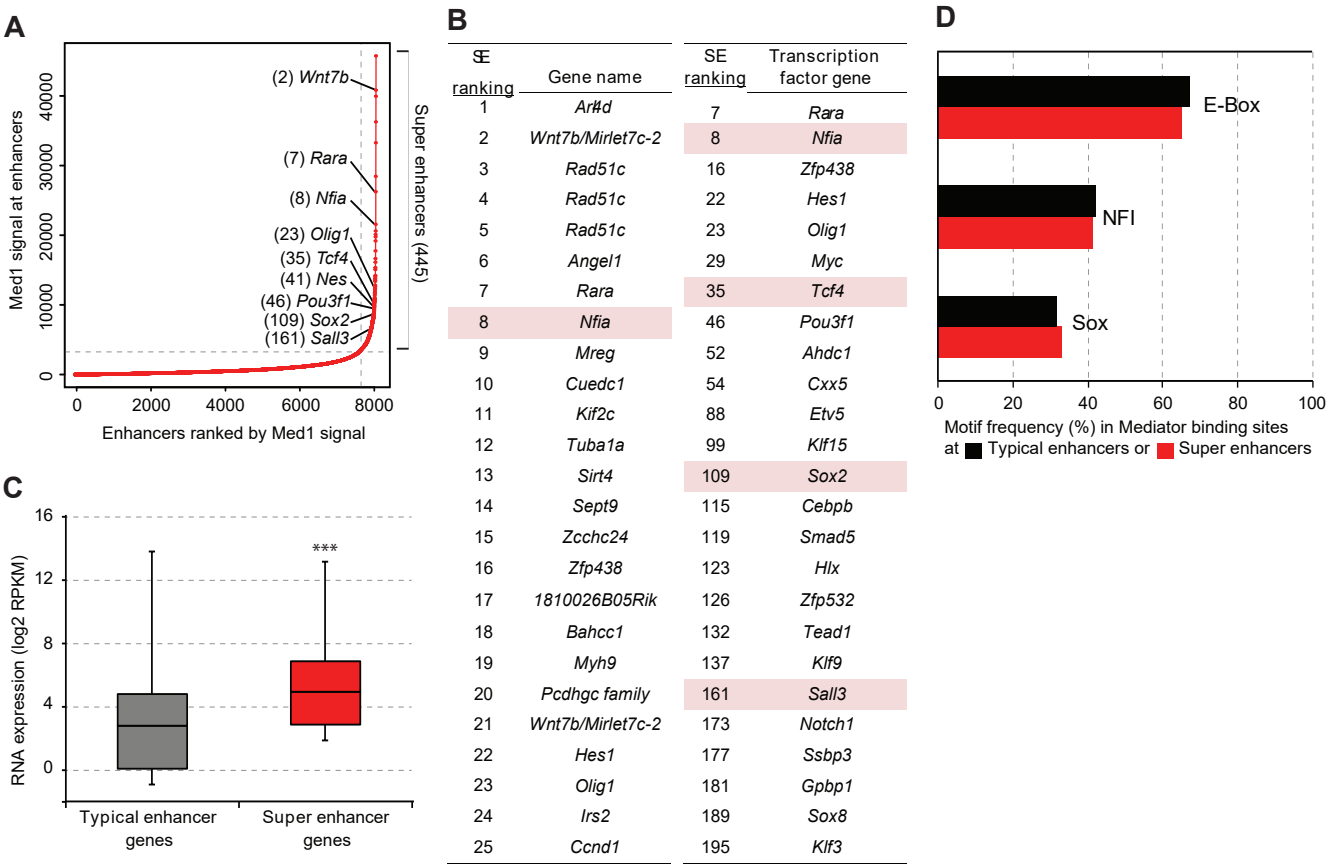


Figure 3 | Super enhancers in neural stem cells

(A) Distribution of Med1 ChIP-seq signal (total reads) in enhancer regions in NSCs. 445 enhancers regions in the right upper quadrant are postulated as super enhancers. Examples of genes near super enhancers and the super enhancer rank are indicated. (B) Top 25 super enhancers (SEs) in NSCs, ranked by Mediator content, and their nearest active gene (left panel). Top 25 active transcription factor genes nearest to SEs (Right panel). SE rank is indicated. Genes encoding transcription factors that we identified as Mediator interactors are red-shaded. (C) Distribution of mRNA expression in NSCs of active genes nearest to SEs and active genes nearest to typical enhancers,but not nearest to SEs. Whiskers represent ultimate range. Significanceof the difference in mRNA levels between two gene categories was assessed by student T test (***;p<0.001) (D) Most frequent transcription factor DNA motifs in Mediator binding sites at typical enhancers and SEs. Motif frequency is indicated as the percentage of all Mediator binding sites at typical enhancers or SEs that harbour this motif. See also Table S2.

DNA motif enrichment analysis revealed that E-box, NFI and SOX motifs were the first, second and third most frequent TF DNA binding motifs in Mediator peaks, both within typical enhancers and SEs (Figure 3D). These motifs were also previously observed in NSC enhancers defined by H3K27ac and Ep300[32]. Interestingly, TFs that can bind these motifs are well represented within the select group of TFs that we find interacting with Mediator, with Tcf4 and Tcf12 binding E-box motifs, Nfia and Nfib binding NFI sites and Sox2 binding SOX sites. In summary, we identified SEs in NSCs and find that the E-box motif is the most frequently occurring motif in Mediator peaks within typical enhancers and SEs in NSCs.

Genome-wide overlap of Mediator and its interaction partners outside promoters

The identification of Mediator binding sites in NSCs allowed us to probe its genome-wide overlap with identified Mediator interaction partners. We first focused on Mediator-interacting transcription factors, which with their sequence-specific DNA binding capacity would be candidates for Mediator-recruitment to the genome. Using published ChIP-seq data sets for TFs Nfia and Nfib (combined ChIP-seq; NFI) and Sox2[32], we found that binding sites of NFI and Sox2 highly overlap with Mediator binding sites outside promoters, including at typical enhancers and SEs (Figure 4A). Using our Tcf4 ChIP-seq data set[33], we show that Tcf4 has an even higher overlap with Mediator outside promoters, at typical enhancers and at SEs (Figure 4A), consistent with the finding that the E-box is the most frequent TF motif at Mediator binding sites in enhancers and SEs in NSCs (Figure 3D). The sum of binding sites of Tcf4, Sox2 and NFI (T+S+N) covers nearly 80% of all Mediator binding sites outside promoters and over 80% of Mediator binding sites within typical enhancers and SEs (Figure 4A). The combined binding sites of representatives of three TF families that we find interacting with Mediator, could therefore potentially account for nearly all recruitment of Mediator outside promoters in NSCs. Examples of the overlap of Mediator with Mediator-interacting TFs are shown in Figures 4B and 4C.

Subsequently, we investigated the overlap of Mediator with interacting chromatin modifiers. We performed ChIP-seq for identified Mediator-interactors arginine methylase Carm1 and H3K9 demethylase Jmjd1c. We found that Carm1 and Jmjd1c highly overlap with Mediator outside promoters, at enhancers and at SEs (Figure 4A). Chromatin remodeler Chd7 is known to bind enhancers in ES cells[34] and indeed overlaps with Mediator at enhancers and SEs in NSCs (Figures 4A). As expected, RNApol2 and its associated Integrator complex[35] show a high overlap with Mediator at promoters (Figures 4A). Polycomb protein Cbx8 and insulator protein Ctcf, which we never found interacting with Mediator, show low genome overlaps with Mediator (Figure 4A). Examples of the overlap of Mediator with interacting chromatin modifiers are shown in Figures 4B and 4C. As expected, we also find high overlaps between Mediator-interacting TFs and Mediator-interacting chromatin modifiers (Figure 4D). We conclude that Mediator shows high binding site overlap at enhancers and SEs with interacting TFs Tcf4, NFI and Sox2 and with interacting chromatin modifiers Jmjd1c, Carm1 and Chd7.

Genes with super enhancers and broad H3K4me3 promoters in NSCs are enriched for neurogenic transcriptional regulators

Recently genes with broad H3K4me3 domains at their promoters were identified[36,37], including in NSCs[36]. The top 5% of broadest H3K4me3 domains in promoters (here abbreviated as broad promoters) associated with cell identity genes[36] and tumor-suppressor genes[37]. Mechanistically, broad promoters have increased rates of transcription elongation and higher transcriptional consistency[36,37] and show enhanced DNA looping interactions with SEs[38], compared to their typical counterparts. We found that the complete sets of SE genes and broad promoter genes in NSCs both have Transcription Regulation as their lead Gene Ontology (GO) category (Figure 5A and Table S3).

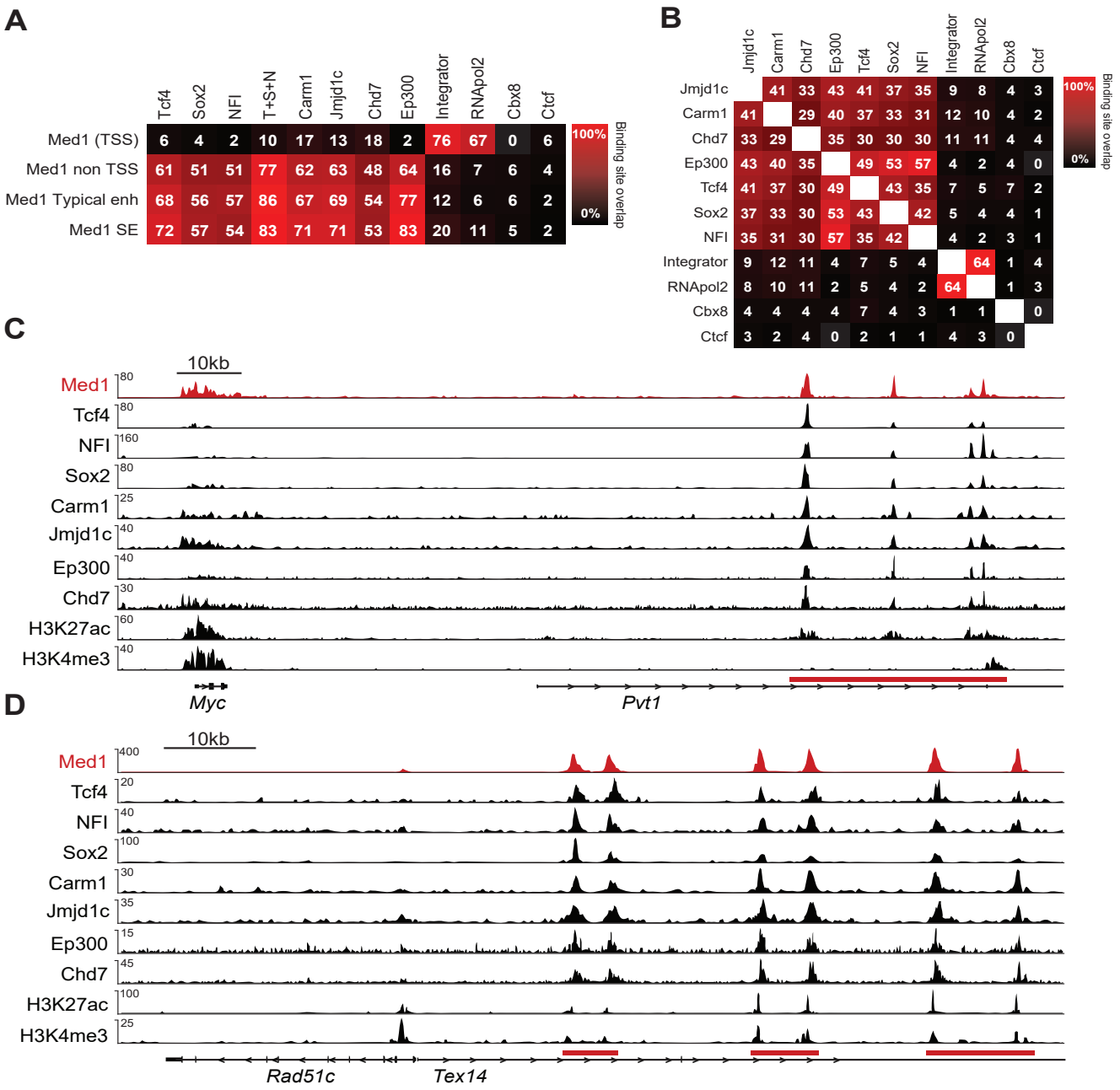


Figure 4 | Binding site overlap of Mediator complex and its interactors

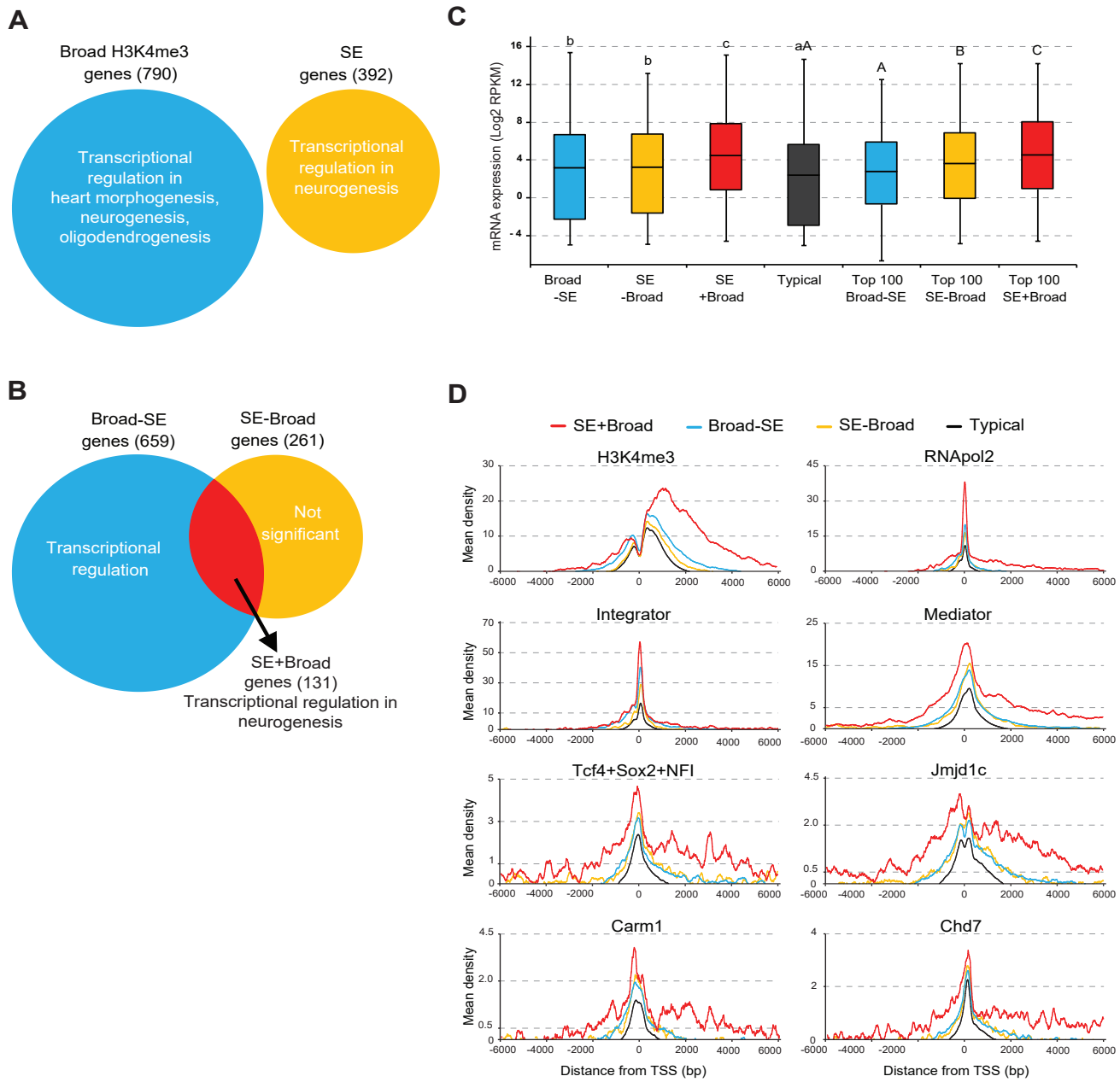
(A) Percentage overlap of genome-wide binding sites of Mediator (Med1) with Mediator interactors Tcf4, Sox2, NFI (Nfia + Nfib), Carm1, Jmjd1c, Chd7, Ep300, Integrator complex (Ints11 subunit), and RNApol2 in NSCs. Cbx8 and Ctcf were not identified as Mediator interactors and serve as negative controls. Percentages overlap of binding sites, as determined by ChIP-seq, are indicated. T+S+N, sum of the binding sites of Tcf4, Sox2 and NFI. TSS, within 1 kb of a transcription start site. (B) Overlap of binding sites of Mediator (Med1) with binding sites of Mediator interactors at the Myc locus in NSCs. ChIP-seq tracks for the indicated proteins and histone modifications at the Myc gene are shown. The Myc SE in the adjacent (inactive) Pvt gene is indicated with a red bar. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated (C) Overlap of binding sites of Mediator (Med1) with binding sites of Mediator interactors at the Myc locus in NSCs. ChIP-seq tracks for the indicated proteins and histone modifications at the Myc gene are shown. The Myc SE in the adjacent (inactive) Pvt gene is indicated with a red bar. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated (D) Overlap of binding sites of Mediator (Med1) with binding sites of Mediator interactors at the Rad51c locus in NSCs. ChIP-seq tracks for the indicated proteins and histone modifications at the Rad51c gene are shown. The Rad51c SEs in the adjacent (inactive) Tex10 gene are indicated with red bars. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated.

Transcriptional regulator genes within the SE category showed neurogenesis as the only significant GO term, whereas transcriptional regulator genes within the broad promoter category included neurogenesis as one of three significant GO terms (Figure 5A and Table S3). The observed enrichment in transcriptional regulators acting in neurogenesis is in line with the association with cell identity genes that has been postulated for genes with SEs[8,9] or genes with broad promoters[36]. We find that genes with broad promoters partially overlap with SE-associated genes in NSCs (Figure 5B). Genes with SEs and broad promoters (SE+Broad) strongly enrich for transcriptional regulators acting in neurogenesis (Figure 5B, Tables S3 and S4). Remarkably, both left-over categories of genes, genes with broad promoters but without SEs (Broad-SE) and genes with SEs but without broad promoters (SE-Broad) lose transcriptional regulators acting in neurogenesis as a GO term, whereas SE-Broad genes lose Transcriptional Regulation as a GO term altogether (Figure 5B, Table S3). Indeed, Mediator-interacting TFs Tcf4, Sox2, Sall3, Nfia and Nfib, as well as other well-known neural TFs including Olig1-2, Pou3f1, Pou3f3 and Npas3 and oncogene Myc have broad promoters and SEs (Table S4). We find that SE+Broad genes are, on average, higher expressed than SE-Broad genes or Broad-SE genes, even when comparing the top 100 of each category (Figure 5C). We conclude that in NSCs, genes with both SEs and broad H3K4me3 promoters account for the association of the separate categories of SE genes and broad promoter genes with transcriptional regulators acting in neurogenesis. Broad promoters and SEs appear to act synergistically to give higher expression in NSCs, as compared to genes with only one of these regulatory elements.

Binding of Mediator and its interaction partners at promoters

We investigated transcriptional regulators binding around promoters of Broad+SE genes. We found that Broad+SE genes had higher and broader promoter signals for H3K4me3, RNApol2 and Integrator than SE-Broad and Broad-SE genes (Figure 5D).

Figure 5 (in next page) | Mediator complex and its interactors at promoters
(A) Predominant Gene Ontology terms for genes with broad H3k4me3 promoters and for active genes nearest to SEs (SE genes) in NSCs. Numbers of genes in each category are indicated between brackets.
(B) Overlap of genes with broad H3K4me3 promoters and SE genes in NSCs. Venn diagram with the two categories of genes, their overlap and their predominant Gene Ontology terms is shown. Numbers of genes in each category are indicated between brackets. (C) Distribution of mRNA levels in NSCs of the different categories of active genes. Box plots based on RNA-seq data are shown. Broad-SE, broad H3K4me3 promoter genes not nearest to SE. SE-Broad, SE genes without broad H3K4me3 promoter. SE+Broad , SE genes with broad H3K4me3 promoter. Typical, genes nearest to a typical enhancer but not nearest to an SE and without a broad H3K4me3 promoter. mRNA levels of all genes and top 100 genes within each category are shown. Statistically significant differences between groups are indicated as separate letters above the box plots, as assessed by student T tests comparing all gene subsets (lower case letters) or top 100 subsets (upper case letters).$p<0.001$ except for B, $p<0.05$.If the letters are the same, the difference between these groups is not significant. (D) ChIP-seq density plots around promoters of the different categories of genes for the indicated factors and histone modifications. Mean ChIP-seq density (y-axis) and distance to TSS (x-axis) are shown. (E) Overlap of binding sites of Med1 with binding sites of Mediator interactors at the Hes1 broad H3K4me3 promotor area in NSCs. ChIP-seq tracks for the indicated proteins and histone modifications at the Hes1 gene are shown. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated. See also Figure S1 and Tables S3 and S4.


A
Broad H3K4me3 genes (790)
SE genes (392)
Transcriptional regulation in heart morphogenesis, neurogenesis, oligodendrogenesis
Transcriptional regulation in neurogenesis
B
Broad-SE genes (659)
SE-Broad genes (261)
Transcriptional regulation
Not significant
SE+Broad genes (131) Transcriptional regulation in neurogenesis
C
mRNA expression (Log2 RPKM)
Broad -SE
SE -Broad
SE +Broad
Typical
Top 100 Broad-SE
Top 100 SE-Broad
Top 100 SE+Broad
D
SE+Broad
Broad-SE
SE-Broad
Typical
H3K4me3
RNApol2
Integrator
Mediator
Tcf4+Sox2+NFI
Jmjd1c
Carm1
Chd7
Mean density
Distance from TSS (bp)
E
1kb
Med1
Tcf4
NFI
Sox2
Carm1
Jmjd1c
Chd7
RNApol2
Integrator
H3K4me3
Hes1

Mediator complex binding to promoters has not yet been analyzed genome-wide at broad promoters or genes nearest to SEs. We found that Mediator has a much higher and broader ChIP signal at Broad+SE genes than at SE-Broad, Broad-SE and typical genes (Figure 5D). Interestingly, we observed the same for Mediator interactors T+S+N, Jmjd1c, Carm1 and Chd7 (Figure 5D). The shape of Mediator signal tracked closely to that of its interactors with a shoulder upstream of the TSS and a long tail into the gene (Figure 5D). As the SE+Broad definition appears to select for genes with the broadest and highest H3K4me3 signal (Figure 5D), we also tested the top 100 SE+Broad, top 100 SE-Broad and top 100 Broad-SE genes to have more equal signals. Indeed top 100 SE+Broad and top 100 Broad-SE have more similar H3K4me3 signals (Figure S1A) and showed more similar signals for Mediator and its interactors at the TSS and upstream of the TSS. However, Mediatorand its interactors have a higher signal downstream of the TSS in SE+Broad genes, as compared to all other categories. Top 100 SE-Broad genes have a narrower signal for all these factors (Figure S1A). The close similarity between the Mediator signal and the signals of its interactors Tcf4, Sox2, NFI, Jmjd1c, Carm1 and Chd7 is also apparent at individual broad promoter regions (Figures 5E and S1B). Top 100 SE+Broad promoters have more RNApol2 and Integrator signal than top 100 Broad-SE and top 100 SE-Broad promoters (Figure S1A), suggesting more efficient recruitment of RNApol2 and Integrator as a potential explanation for their higher expression (Figure 5C). We conclude that broad promoters have higher and broader signals for Mediator that is closely tracked by all its tested interacting factors.

Tcf4 regulates neurogenic transcription factor genes with super enhancers and broad H3K4me3 promoters

Tcf4 showed the highest overlap with Mediator at enhancers and SEs of the tested Mediator-interacting TFs (Figure 4A) prompting us to further investigate a possible role of Tcf4 in regulating genes near SEs. We find that Tcf4 content followed Mediator content at enhancers and SEs (Figure 6A).

To test to what extent Tcf4 regulates genes with or without SEs and/or broad H3K4me3 promotors, we used our RNA-seq data set from RNA isolated 44 hours after Tcf4 knock-down or control knock-down in NSCs[33]. We found that Tcf4 depletion down-regulates nearly two-thirds of all SE+Broad genes (Figure 6B) and also has the strongest down-regulating effect on SE-containing genes (Figure 6C). Genes without SEs, either Broad-SE genes or genes with typical enhancers, are significantly less affected by Tcf4 depletion (Figure 6C). This suggests that Tcf4 predominantly regulates genes via SEs. Indeed, Tcf4 is present on nearly all SEs of SE+Broad genes (Figure 6D). Tcf4-bound and activated SE+Broad genes include 15 transcription factor genes (Figure 6E) of which Bahcc1, Hes1,Myc, Nfib, Sall1 and Sall3, Olig2, Thra and Npas3 encode known regulators of neural progenitors and/or neurogenesis [36,39–44].Tcf4 protein has protein-protein interactions in NSCs with 6 TFs that are part of this set of Tcf4-activated TF genes, including Nfib and Olig2[33] (Figure 6E). This allows for a potential feed-forward circuit (Figure 6E) where Tcf4 maintains the expression of its own co-factors, which then subsequently may aid Tcf4 in the regulation of other target genes and its own expression. In line with this possibility, NFI and Olig2 colocalize with Tcf4 and Mediator on SEs in all 15 TF genes, for example at the Olig2 gene (Figure 6F), the Sall3 gene and the Notch1 gene (Figure S2). Tcf4, Mediator,

NFI and Olig2 also co-localise at the SE in the Tcf4 gene itself (Figure 6G). Nfib expression has the second-best spatial-temporal correlation (0.56 Pearson coefficient) with Tcf4 expression in pre-natal development of the mouse brain (out of 1104 genes[45] and the second best spatial temporal correlation (0.90 Pearson coefficient)with TCF4 in pre-natal human brain development (out of 19700 genes)[46], suggesting that a Tcf4-Nfib co-regulatory partnership could be widespread in mammalian brain development.

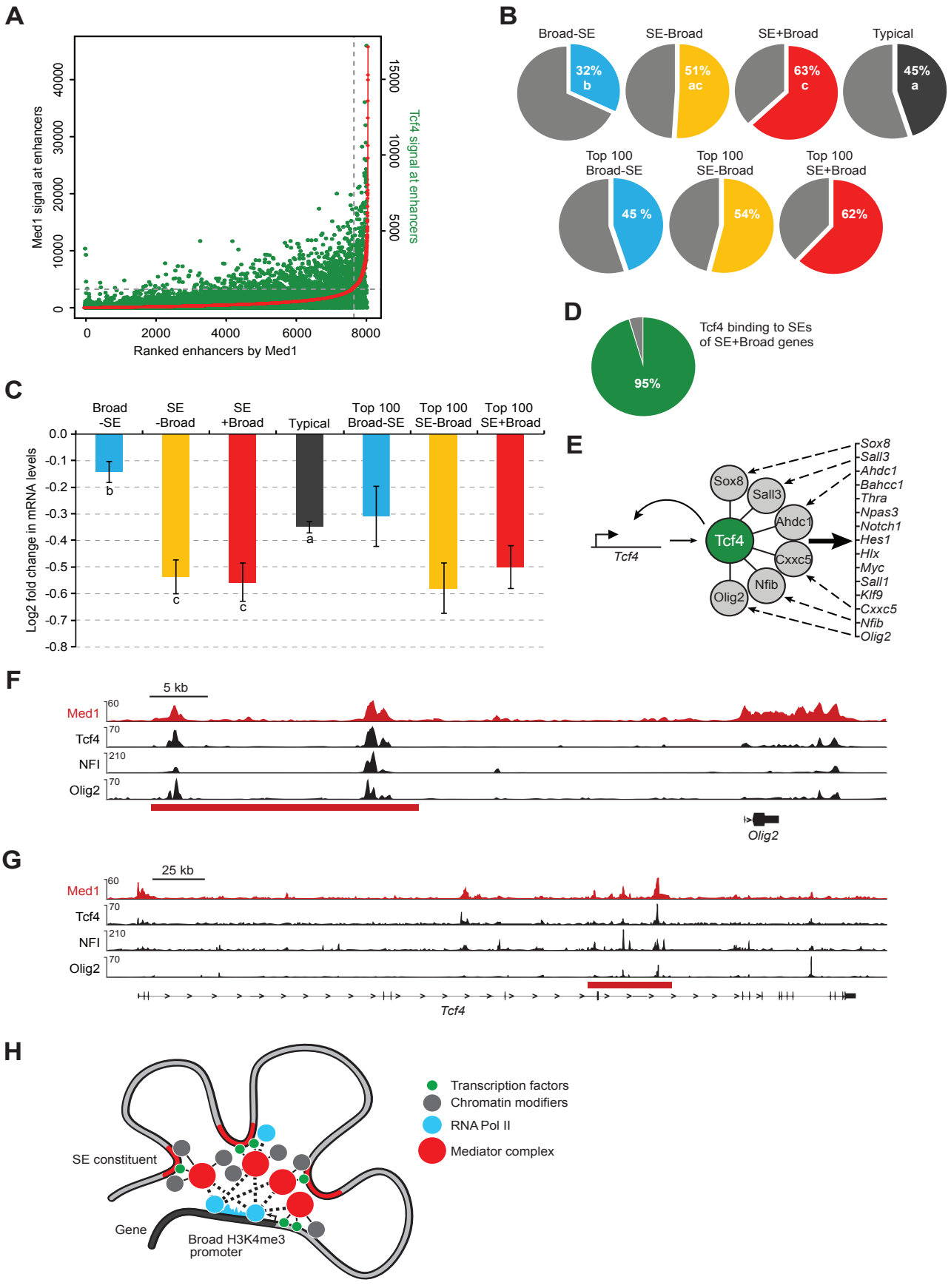
DISCUSSION

A Mediator interaction network of enhancer- and promoter-binding proteins

We have expanded the protein-protein interaction network of the Mediator complex with many proteins and complexes that reside at enhancers, super enhancers or promoters and thereby established the potential of the Mediator complex as a major interaction hub at enhancer-promoter assemblies.

Figure 6 (in next page) | Tcf4 regulates neurogenic transcription factor genes with super enhancers and broad H3K4me3 promoters.

(A) Tcf4 signal at enhancers ranked by Med1 content. Tcf4 ChIP-seq read content (in green) is plotted against at enhancers ranked by Med1 ChIP-seq read content (in red).Enhancers in the right upper quadrant are super enhancers, analogous to Figure 3A. (B) Percentages of down-regulated genes in the different categories upon Tcf4 knock-down in NSCs.Broad-SE: broad H3K4me3 promoter genes not nearest to SE. SE-Broad: SE genes without broad H3K4me3 promoter. SE+Broad: SE genes with a broad H3K4me3 promoter. Typical, genes nearest to a typical enhancer but not nearest to an SE and without a broad H3K4me3 promoter. Counted down-regulated genes are at least 1.5-fold down-regulated. Percentages of down-regulated genes in all genes and top 100 genes within each category are shown.Statistically significant differences between groups are indicated as separate letters in the pie charts, $p<0.001$as assessed by student T tests.If the letters are the same, the difference between these groups is not significant. (C) Changes in mRNA levels of the different categories of genes upon Tcf4 knock-down versus control knock down in NSCs. Log2 fold change, based on RNA-seq data, is shown. Error bars indicate the standard error of the mean (SEM) based on the RNA-seq triplicates. Statistically significant differences between groups are indicated as separate letters below the box plots, $p<0.001$as assessed by student T tests. (D) Percentage of SEs of SE+Broad genes in NSCs bound by Tcf4. SEs nearest to SE+Broad genes with or without significant Tcf4 binding sites, as determined by ChIP-seq, were counted. (E) Model of a Tcf4-driven feed-forward transcriptional circuit of SE+Broad TF genes in NSCs. 15 SE+Broad TF genes bound at their SE by Tcf4 and activated by Tcf4 are indicated. Tcf4 also binds its own SE. TF proteins encoded by 6 target genes also interact with Tcf4 protein and may aid in transcriptional regulation by Tcf4. (F) Overlap of binding sites of Tcf4 and Med1 with binding sites of Tcf4-interactors Olig2 and NFI at the Olig2 locus in NSCs. ChIP-seq tracks for the indicated proteins at and around the Olig2 gene are shown. The Olig2 SE is indicated with a red bar. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated. (G) Overlap of binding sites of Tcf4 and Med1 with binding sites of Tcf4-interactors Olig2 and NFI at the Tcf4 locus in NSCs. ChIP-seq tracks for the indicated proteins at the Tcf4 gene are shown. The Tcf4 SE, inside the Tcf4 gene, is indicated with a red bar. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated. (H) Model of SE-Broad H3K4me3 promoter assemblies. TFs at SE constituents and the Broad H3K4me3 promoter recruit high levels of Mediator complex to SE-Broad assemblies. In turn, Mediator recruits high levels of protein-protein interaction partners such as the RNApol2 complex, Integrator and chromatin modifiers. This would result in efficient pause-release of RNApol2 and high but TF-regulated levels of transcription. See also Figure S2.


A
Med1 signal at enhancers
Tcf4 signal at enhancers
Ranked enhancers by Med1
B
Broad-SE
SE-Broad
SE+Broad
Typical
32%
b
51%
ac
63%
c
45%
a
Top 100 Broad-SE
Top 100 SE-Broad
Top 100 SE+Broad
45 %
54%
62%
C
Log2 fold change in mRNA levels
Broad -SE
SE -Broad
SE +Broad
Typical
Top 100 Broad-SE
Top 100 SE-Broad
Top 100 SE+Broad
D
Tcf4 binding to SEs of SE+Broad genes
95%
E
Tcf4
Sox8
Sall3
Ahdc1
Cxxc5
Nfib
Olig2
Sox8
Sall3
Ahdc1
Bahcc1
Thra
Npas3
Notch1
Hes1
Hlx
Myc
Sall1
Klf9
Cxxc5
Nfib
Olig2
F
5 kb
Med1
Tcf4
NFI
Olig2
Olig2
G
25 kb
Med1
Tcf4
NFI
Olig2
Tcf4
H
SE constituent
Gene
Broad H3K4me3 promoter
Transcription factors
Chromatin modifiers
RNA Pol II
Mediator complex

Mediator binds to enhancers and promoters in close proximity to many other proteins. We believed that chromatin-independent protein-protein interactions of purified Mediator complex, as identified by their detection by mass spectrometry, would be the best indicator of its recruitment capacity. Despite our stringent criteria, 20 years of research on the Mediator complex since its discovery by several labs[29,47–49] and progressing high throughput interaction studies[28,50], we find that 75 of our 95 identified Mediator interactions are novel.

Identified Mediator interactors can be broadly divided into DNA sequence-independent proteins, mostly chromatin modifiers, and sequence-specific transcription factors. The latter category of Mediator interactors would represent potential Mediator recruitment factors. Indeed, NF-kappaB subunit RelA, one of the two known Mediator interactors among the 16 identified transcription factors, recruits Mediator to activate transcription[51]. Whereas Mediator-interacting transcription factors would be more specific for NSCs (see next paragraph), the Mediator-interacting chromatin modifiers and other proteins are mostly ubiquitously expressed and would have general relevance for transcriptional regulation.Supporting this suggestion, our Mediator interactor screen discovered two major enhancer binding proteins. We observed and independently confirmed interactions between Mediator and arginine methylase Carm1 and putative H3K9 demethylase Jmjd1c. Carm1 is a highly studied enzyme and best known in transcriptional regulation as a co-activator of nuclear receptors and NF-kappaB and was shown to act at individual promoters[52,53]. We find that Carm1 is a genome-wide enhancer binding protein in NSCs that closely co-localizes with Mediator. Jmjd1c was identified as a co-activator of the tumor-inducing fusion gene AML1-ETO and shown to be recruited by AML1-ETO to target gene promoters where it lowers the levels of the repressive mark H3K9me2[54]. We show that Jmjd1c marks enhancers genome-wide in NSCs, together with Mediator, where it may perform a similar enzymatic role to maintain enhancer activity.

A recent analysisis[9] showed that chromatin modifiers Brd4, Ep300, Crebbp, Chd7, SWI-SNF complex, LSD1 complex, Cohesin complex and NuRD complex colocalise with Mediator at enhancers and have an increased binding density at SEs, similar to the Mediator complex. With the exception of Brd4, we find that all the above chromatin modifiers as Mediator interactors, which may suggest that Mediator interaction aids in their recruitment to enhancers and SEs. The apparent correlation of having protein-protein interactions with Mediator and co-localizing with Mediator on the genome would predict that other observed Mediator interactors of unknown genomic location also reside at enhancers or promoters. This remains to be tested.

Mediator-interacting transcription factors define the epigenetic landscape in NSCs

We performed Mediator ChIP-seq to identify SEs in NSCs. We find that Mediator-defined SEs in NSCs have as their most frequent TF motifs E-box, NFI and SOX, similar to NSC enhancers in general[32]. Nfia, Nfib, Sox2, Tcf4, and Tcf12, which can bind one of these motifs, are among the small set of 16 TFs that we identified as Mediator interactors. This shows a remarkable synchrony between Mediator-binding TFs and prominent enhancer motifs in NSCs. Our identified Mediator-binding TFs are not the highest expressed TFs in NSCs, suggesting that they have a higher binding affinity for

Mediator than other TFs. The above set of TFs may therefore define enhancers and SEs in NSCs by having high affinity for Mediator and thereby being effective at recruiting Mediator and its interactors to its binding sites. This would suggest Mediator affinity as an important organizing feature in establishing the enhancer landscape in a given cell type. Indeed, the sum of the binding sites of Tcf4, Sox2 and Nfi represents nearly all Mediator binding sites at enhancers, and outside promoters in general, and can therefore explain genome-wide Mediator recruitment outside promoters in NSCs.

Relative promoter occupancy of Mediator has not been analysed genome-wide in higher eukaryotes, to our knowledge. We find that Mediator has higher and especially broader binding signals at promoters with a broad H3K4me3 signal, a class of promoters that was recently discovered[36,37].Tcf4, Sox2 and Nfi show relatively weak occupancy at promoters in general. However, their binding is enhanced at broad promoters and Mediator follows closely their binding pattern in our genome-wide plots, as well as at individual broad promoters. We find that Mediator-interacting chromatin modifiers,such as Carm1 and Jmjd1c, also track Mediator binding at promoters. Relative enrichment of transcription factors at broad promoters was observed before in different cell types[36,37]. Our results suggest that broad promoters may act like proximal enhancers in recruiting TFs, which in turn can recruit Mediator and its interactors. The close resemblance of the Mediator genome-wide binding sites with the binding sites of its interacting TFs is highly suggestive of Mediator recruitment by these TFs.

Tcf4 is part of a regulatory circuit of neurogenic transcription factors that regulate NSC identity

We find that Tcf4 preferentially regulates SE-containing genes in NSCs, including a set of neurogenic transcription factor genes that have SEs and broad promoters. Intriguingly, we find that a number of the TFs encoded by these genes have protein-protein interactions with the Tcf4 protein. Some of these Tcf4-interacting TFs co-localize with Tcf4 at SEs in this set of target genes, as well as on the Tcf4 gene itself, suggesting a feed-forward circuit that maintains the expression of these TFs in NSCs. Feed-forward circuits of key TFs in embryonic stem cells (ESCs), such as Oct4, Sox2, Nanog, Esrrb and Klf4, were shown to regulate pluripotency and follow the same above criteria[8,55,56]. Analogous to the ESC TF circuit, many of the TFs in our NSC circuit are essential for NSC self-renewal or their neuronal differentiation capacity. Together, this suggests that we have uncovered a TF circuit that would be central to the regulation of NSC identity. TCF4 heterozygosity in humans leads to Pitt Hopkins syndrome with severe intellectual disability[57,58], whereas SNPs in the TCF4 locus were among the first significant schizophrenia risk variants to be discovered[59] and are the most significant schizophrenia risk SNPs to date[60]. These genetic data suggest that TCF4 plays an important role in brain development and needs to be tightly regulated to prevent neurodevelopmental disease. Our TF circuit may facilitate this regulation.

Mediator complex as a recruitment hub that facilitates the regulation of cell identity genes

Mediator complex binding signal was used as one parameter to postulate SEs[8], which were subsequently shown to regulate cell identity genes and oncogenes in many cell types([8–10]. More

recently, promoters with a broad H3K4me3 domain were postulated to regulate cell identity genes[36,37]. As was shown before in other cell types[36,37], we find that SE genes and Broad genes partially overlap in NSCs. However, we show that the link to neurogenic transcriptional regulators in SE genes and Broad genes in NSCs is derived from neurogenic transcriptional regulator genes in the overlap of both categories; genes that have both SEs and broad promoters. This suggests that, at least in NSCs, SE+Broad genes represent a special category of genes that is strongly linked to cell identity. These SE+Broad genes have high recruitment of Mediator at their SEs (by definition) and we find that they also recruit high levels of Mediator to their promoters. Increased promoter levels of Mediator are also observed at broad promoters without surrounding SEs may therefore be recruited by Mediator-interacting TFs, which we also find enriched at broad promoters. SEs were recently shown to have increased 3D interactions with broad promoters, as compared to typical promoters[38]. We find that SE+Broad genes in NSCs are the category of genes with highest levels of RNApol2 and Integrator at their promoters. Integrator complex associates with RNApol2 and plays an important role in the transcription-initiation and pause-release of RNApol2[35]. The efficient recruitment of RNApol2 and Integrator at SE+Broad genes thereby provides an explanation for our observation that this category of genes has the highest expression in NSCs.

All together this fits into a model (Figure 6H) where Mediator is recruited by Mediator-interacting TFs to both SEs and Broad promoters. These elements then form relatively stable enhancer-promoter assemblies that have high local concentrations of Mediator and its co-recruited protein-protein interaction partners, including RNApol2, Integrator and chromatin modifiers. Such assemblies would provide an optimal environment for the efficient pause-release of high quantities of RNApol2 and thereby combine the high transcriptional consistency and the high transcriptional efficiency that have been shown for broad promoters and SE genes, respectively[8,9,36]. SE-broad promoter assemblies and our identified Mediator interactions could provide ideal building blocks for the phase-separated complexes that have been recently proposed to drive robust transcription of cell identity genes in mammals[16].

METHODS

Purification of the Mediator complex from neural stem cells

NS-5 neural stem cells (NSCs) were derived from 46C embryonic stem cells[61] and cultured, as described[62] and regularly tested for mycoplasma contamination and for authenticity by expressed NSC markers Pax6, Sox2 and Nestin[22]. NSC lines with stable expression of C-terminally FLAG-tagged Med15 were created by electroporation with pCAG promoter-driven plasmids containing Med15 cDNA and puromycin selection for individual clones with moderate expression of the tagged proteins, as compared to endogenous levels [33,22]. Nuclear extract was prepared from NSCs expressing FLAG-Med15 and from control NSCs[23]. FLAG-tagged Mediator complex was purified from 1.5 ml nuclear extract, equivalent to 2*108 NSCs, by FLAG-affinity purification, as described[22,64]. Mediator complex purifications were performed from nuclear extract with Benzonase (150U per ml nuclear extract) added or Ethidium bromide (50 μg per ml) added at the start of the 3-hour incubation period

of the anti-FLAG antibody beads with the nuclear extract. Alternatively, Mediator complex purification was performed from untreated nuclear extract. In one experiment, Mediator complex purifications were performed from nuclear extracts treated with Benzonase, Ethidium bromide or untreated nuclear extract, together with a control purification from nuclear extract from control NSCs. In a second, independent, experiment, Mediator complex was purified from nuclear extract treated with Benzonase, together with a control purification. Control purifications were from nuclear extract treated with benzonase. All purifications are shown in Table S1.

Identification of proteins by mass spectrometry was as described[64]. Peptide spectra from purified Mediator samples or control sample were searched against UniProt release 2012-11 for protein identification. Initial inclusion criteria for Mediator-interacting proteins are similar to those as described before[64]. In short, 1) A minimal Mascot score of 50, 2) At least 5-fold enrichment by emPAI score in the Mediator purified sample over the control sample. emPAI score is an estimate of the quantity of the identified protein in the purified protein sample, based on the number of peptide spectra identified by MS, normalized for the number of peptides that theoretically should be identifiable for that protein[65]. 3) At least 3-fold enrichment by Mascot score in the mediator purified sample over the control sample. Of note, of the 95 identified Mediator complex interactors, only 13 are also detected in any of the two control samples (Table S1). Subsequently, recorded Mediator interactors cannot be 2-fold lower or more in emPAI score in the Mediator complex purification from purifications from nuclear extracts treated with Benzonase or Ethidium bromide, as compared to a parallel Mediator complexpurification from untreated nuclear extract. Finally, Mediator interaction partners are only included in the final list (Table S1) if they are specifically present in all 4 Mediator complex purifications. Mediator interaction partners were defined as novel if they did not appear as identified by Affinity Capture or Reconstituted Complex in BioGRID, the most comprehensive protein-protein interaction database[66]. Interaction network graphics were made with Cytoscape[67]. Thickness of the edges in the interaction network (Figure 1C) gives an indication of the relative molar protein quantity (based on emPAI score) in purified Mediator complex samples with 4 categories of thickness; emPAI> 1.5, thickest edge, 0.75<emPAI ≤ 1.5, one but thickest edge, 0.25<emPAI ≤0.75, one but thinnest edge, emPAI<0.25, thinnest edge.

Immunopreciptations

Immunoprecipitations of Jmjd1c or Carm1 were performed from 1 ml of NSC nuclear extract using 10µg of Jmjd1c antibody (Merck Millipore #17-10262, or 10µg Carm1 (Cell Signaling Technology #12495) or 10µg of control rabbit IgG (Santa Cruz #sc-2027), exactly as described(van den Berg et al., 2010). Resulting western blots were probed with the same antibodies and Med12 antibody (Bethyl Laboratories #A300-774A).

Mediator-interacting TFs

References for function in neural development. Yy1[68], Nfia[69], Sox2[70], Nfib[71], Sall3[36], Tcf12[72], [73], Rela[74,75], Tcf4[76,77]. TF mRNA levels in our NSCs are from our RNA seq data on our wild-type NSCs[33].

Chromatin immunoprecipitation combined with high-throughput sequencing (ChIP-seq)

We adapted protocols previously described [8,33,78]. 1.5 × 108 NSCs were used per chromatin immunoprecipitation (ChIP). Cells were collected in 1xPBS and crosslinked first with 2 mM disuccinimidylglutarate (Thermo Fisher Scientific, Waltham, MA, USA) solution for 45 min and then 1% formaldehyde solution for 15 min at room temperature. Cells were washed twice with 1X PBS and flash frozen in liquid nitrogen. Chromatin was prepared for sonication with 20 mM Tris-HCl pH8, 150 mMNaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100. We used 15 cycles of 30 seconds ON, 30 seconds OFF on a Bioruptor Pico sonication device (Diagenode Cat# B01060001) to shear chromatin to 150-200bp fragments. The resulting chromatin extract was incubated overnight at 4°C with 100ul of Dynal Protein G magnetic beads that had been pre-incubated with 10 μg of the appropriate antibody. We used the following antibodies: Med1 (Bethyl Labs #A300-793A), Carm1 (Cell Signaling Technology #12495), Jmjd1c (Merck Millipore #17-10262), IgG (Normal Rabbit IgG: Santa Cruz #sc-2027). Beads were washed 1X with the sonication buffer, 1X with 20 mM Tris-HCl pH8, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1%Triton X-100, 1X with 10 mM Tris-HCl pH8, 250 mM LiCl, 2 mM EDTA, 1% NP40 and 1X with TE containing 50 mM NaCl. Bound complexes were eluted from the beads in 50 mM Tris-HCl, pH 8.0, 10 mM EDTA and 1% SDS by heating at 65°C for 1 hr with occasional vortexing and crosslinking was reversed by overnight incubation at 65°C. ChIP-seq sample preparation and sequencing on Illumina GAII or HiSeq2500 (San Diego, CA, USA) platforms was performed at the Erasmus MC Center for Biomics, as described[79].

Genomic data analyses

All ChIP-seq data sets were mapped to the mouse mm9 reference genome using Bowtie v0.12.7[80], where we used a seed length of 36 in which we allowed a maximum of two mismatches. If a read had multiple alignments only the best matching read was reported. ChIP-seq data sets with multiple replicates were merged. Duplicated reads were removed. MACS46 v1.4.2 was used for peak calling using default settings, using IgG ChIP-seq as background control for our Med1, Carm1, Jmjd1c, Tcf4, Olig2 and Chd7 ChIP-seq data. For external ChIP-seq data sets either IgG ChIP-seq or sequenced chromatin input was used as background control. For histone modifications we used HOMER findPeaks[81] using -region -size 1000 -minDist 2500 parameters. Genomic datasets that are reported and/or used in this study are summarized in Table S5. The accession number for the data reported in this paper is GEO: GSE109043.

Enhancers in mouse NSCs were defined with the same criteria as described previously[82] but recalling Ep300 and H3K27ac peaks using HOMER, function REGION, and using Bedtools[83] to generate overlaps between Ep300 peaks and H3K27ac peaks. SEs were identified using the ROSE algorithm[8], ranking defined enhancers by Med1 ChIP-seq signal. 435 super enhancers were identified and the rest were assigned as typical enhancers. Plotting was performed using hockeyfunction in R. We used the already described list of mouse NSCs broad H3K4me3 promoters[36].

For mRNA levels in our mouse NSCs, we used our published RNA sequencing data set[33] consisting of three replicates to calculate the mean mRNA expression levels. Super enhancer (SE) genes and typical enhancer genes are defined as the closest active gene, RKPM > 0.5 in our NSC RNA-seq data[33], to an SE or a typical enhancer, respectively.

Motif analyses were performed using HOMER[81] and selecting the most frequent motifs found at Med1 binding sites at SE constituents and typical enhancers.

For genome-wide binding site overlaps, we used the 5000 most significant binding sites for each factor to determine the percentage of overlap between two factors. Two binding sites were considered overlapping if their summits were within 200 bp. Promoters were defined as the regions within 1.5 kb of a transcription start site (TSS). Top 5000 peaks from Mediator and its interactors were separated in the TSS, nonTSS, typical enhancer and super enhancer categories and the percentage of overlap recalculated for each subset.

Generation of histograms documenting ChIP-seqsignal density at specific sets of promoters in the NSC genome was performed by HOMER annotatePeaks with 10bp bins and 12000bp around the TSS. By default, HOMER normalizes the output histogram such that the resulting units are per bp per peak, on top of the standard total mapped tag normalization of 10 million tags. For each promoter, directionality was extracted from TSS annotation and each subset was split between plus or minus strand. Subsequently, split lists were then remerged taking into account directionality to finally calculate the ChIP-signal density values. Enhancer-annotated expressed genes not present on the super enhancer gene list or the broad H3K4me3 promoter gene list were used as typical genes.

Gene Ontology analyses on the different gene categories were performed using DAVID version 6.7[84] using default categories. SE genes are defined as the closest active gene, RKPM > 0.5 in our NSC RNA-seq data[33], to an SE. Broad H3K4me3 genes are defined as having the top 5% broadest H3K4me3 domains[36]. Additionally, we performed GO ontology biological process analysis of the transcription regulators found in each subset. Benjamini-corrected p-value was used for ranking Gene Ontology terms.

Tcf4-regulated genes were derived from an RNA-seq. experiment performed in triplicate 48 hours after Tcf4 shRNA-transfection or control shRNA-transfection in mouse NSCs[33]. Gene expression values with significant triplicates were assigned to the different subsets. The effect ofTcf4 knock down was indicated by plotting the mean fold change vs the scrambled shRNA condition for each subset.

Data availability

Genomic datasets that are reported and/or used in this study are summarized in Table S5. The accession number for the data reported in this paper is GEO: GSE109043 (link for reviewers: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE109043 token inwbuwianjozjud).

Acknowledgements

We thanks Christel Kockx for technical assistance. M.Q. was supported by a graduate programme grant (nr. 022.004.002) from thethe Netherlands Organisation for Scientific Research (NWO). M.R.D was supported by the DevRepair (P7/07) IAP-VII network, J.H.B. was supported by an ALW-open program grant (No 821.02.004) from NWO and R.A.P. and D.L.C.v.d.B.by a grant from the Dutch government to the Netherlands Institute for Regenerative Medicine (NIRM, grant No. FES0908). D.H.D and J.D were funded by The Netherlands Proteomics Centre (Project Number 184.032.201), financed by NWO.

Author contributions

M.Q. performed nearly all experiments and designed and performed most of the bioinformatic analyses. M.R.D. provided technical assistance. D.H.W.D and J.D. performed the mass spectrometry analyses.D.L.C.v.d.B. performed plasmid constructions and preliminary experiments. Z.O. and W.F.J.van IJ. performed labeling and Illumina sequencing of ChIPmaterial. M.F. designed, performed and supervised part of the bioinformatic analyses. R.A.P. conceived the study and designed experiments. R.A.P. wrote the manuscript with help from co-authors.

REFERENCES

1. Conaway, R. C. & Conaway, J. W. Origins and activity of the Mediator complex. Semin. Cell Dev. Biol. 22, 729–734 (2011).
2. Jeronimo, C. et al. Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. Mol. Cell 64, 455–466 (2016).
3. Jeronimo, C. & Robert, F. The Mediator Complex: At the Nexus of RNA Polymerase II Transcription. Trends Cell Biol. xx, 1–19 (2017).
4. Malik, S. & Roeder, R. G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat. Rev. Genet. 11, 761–772 (2010).
5. Borggrefe, T. & Yue, X. Interactions between subunits of the Mediator complex with gene-specific transcription factors. Semin. Cell Dev. Biol. 22, 759–768 (2011).
6. Ansari, S. A. & Morse, R. H. Mechanisms of Mediator complex action in transcriptional activation. Cell. Mol. Life Sci. 70, 2743–2756 (2013).
7. Takahashi, H. et al. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell 146, 92–104 (2011).
8. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319 (2013).
9. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. Cell 155, (2013).
10. Lovén, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell 153, 320–334 (2013).
11. Liu, W. et al. Brd4 and JMJD6-Associated Anti-Pause Enhancers in Regulation of Transcriptional Pause Release. Cell 155, 1581–1595 (2013).
12. Parker, S. C. J. et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc. Natl. Acad. Sci. 110, 17921–17926 (2013).
13. Pott, S. & Lieb, J. D. What are super-enhancers? Nat. Genet. 47, 8–12 (2014).
14. Chapuy, B. et al. Discovery and Characterization of Super-Enhancer-Associated Dependencies in Diffuse Large B Cell Lymphoma. Cancer Cell 24, 777–790 (2013).
15. Liu, W. et al. Brd4 and JMJD6-Associated Anti-Pause Enhancers in Regulation of Transcriptional Pause Release. Cell 155, 1581–1595 (2013).
16. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. Cell 169, 13–23 (2017).
17. Cox, J. L. et al. The SOX2-Interactome in Brain Cancer Cells Identifies the Requirement of MSI2 and USP9X for the Growth of Brain Tumor Cells. PLoS One 8, e62857 (2013).
18. Grebien, F. et al. Pharmacological targeting of the Wdr5-MLL interaction in C/EBPα N-terminal leukemia. Nat. Chem. Biol. 11, 571–578 (2015).
19. Kagey, M. H. et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature 467, 430–435 (2010).
20. Cho, H. et al. A human RNA polymerase II complex containing factors that modify chromatin structure. Mol. Cell. Biol. 18, 5355–63 (1998).
21. van den Berg, D. L. C. et al. An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells. Cell Stem Cell 6, 369–381 (2010).

22. Engelen, E. et al. Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. Nat. Genet. 43, 607–611 (2011).

23. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. 11, 1475–89 (1983).

24. Ho, L. et al. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc. Natl. Acad. Sci. 106, 5187–5191 (2009).

25. Schnetz, M. P. et al. Genomic distribution of CHD7 on chromatin tracks H3K4 methylation patterns. Genome Res. 19, 590–601 (2009).

26. Visel, A. et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457, 854–858 (2009).

27. Whyte, W. A. et al. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature 482, 221–225 (2012).

28. Huttlin, E. L. et al. Architecture of the human interactome defines protein communities and disease networks. Nature 545, 505–509 (2017).

29. Rachez, C. et al. A novel protein complex that interacts with the vitamin D3 receptor in a ligand-dependent manner and enhances VDR transactivation in a cell-free system. Genes Dev. 12, 1787–800 (1998).

30. Yuan, C. X., Ito, M., Fondell, J. D., Fu, Z. Y. & Roeder, R. G. The TRAP220 component of a thyroid hormone receptor- associated protein (TRAP) coactivator complex interacts directly with nuclear receptors in a ligand-dependent fashion. Proc. Natl. Acad. Sci. U. S. A. 95, 7939–44 (1998).

31. Shen, C. et al. NSD3-Short Is an Adaptor Protein that Couples BRD4 to the CHD8 Chromatin Remodeler. Mol. Cell 60, 847–859 (2015).

32. Martynoga, B. et al. Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. Genes Dev. 27, 1769–1786 (2013).

33. Moen, M. J. et al. An interaction network of mental disorder proteins in neural stem cells. Transl. Psychiatry 7, e1082 (2017).

34. Schnetz, M. P. et al. CHD7 Targets Active Gene Enhancer Elements to Modulate ES Cell-Specific Gene Expression. PLoS Genet. 6, e1001023 (2010).

35. Gardini, A. et al. Integrator Regulates Transcriptional Initiation and Pause Release following Activation. Mol. Cell 56, 128–139 (2014).

36. Benayoun, B. A. et al. H3K4me3 Breadth Is Linked to Cell Identity and Transcriptional Consistency. Cell 158, 673–688 (2014).

37. Chen, K. et al. Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. Nat. Genet. 47, 1149–1157 (2015).

38. Thibodeau, A., Márquez, E. J., Shin, D.-G., Vera-Licona, P. & Ucar, D. Chromatin interaction networks revealed unique connectivity patterns of broad H3K4me3 domains and super enhancers in 3D chromatin. Sci. Rep. 7, 14466 (2017).

39. Dominguez, M. H., Ayoub, A. E. & Rakic, P. POU-III Transcription Factors (Brn1, Brn2, and Oct6) Influence Neurogenesis, Molecular Identity, and Migratory Destination of Upper-Layer Cells of the Cerebral Cortex. Cereb. Cortex 23, 2632–2643 (2013).

40. López-Juárez, A. et al. Thyroid Hormone Signaling Acts as a Neurogenic Switch by Repressing Sox2 in the Adult Neural Stem Cell Niche. Cell Stem Cell 10, 531–543 (2012).

41. Mateo, J. L. et al. Characterization of the neural stem cell gene regulatory network identifies OLIG2 as a multifunctional regulator of self-renewal. Genome Res. 25, 41–56 (2015).

42. Nagao, M. et al. Coordinated control of self-renewal and differentiation of neural stem cells by Myc and the p19 ARF –p53 pathway. J. Cell Biol. 183, 1243–1257 (2008).

43. Nakamura, Y. et al. The bHLH gene hes1 as a repressor of the neuronal commitment of CNS stem cells. J. Neurosci. 20, 283–93 (2000).

44. Pieper, A. A. et al. The neuronal PAS domain protein 3 transcription factor controls FGF-mediated adult hippocampal neurogenesis in mice. Proc. Natl. Acad. Sci. U. S. A. 102, 14052–7 (2005).

45. Xie, W. et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell 153, 1134–1148 (2013).

46. Miller, J. A. et al. Transcriptional landscape of the prenatal human brain. Nature 508, 199–206 (2014).

47. Fondell, J. D., Ge, H. & Roeder, R. G. Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. Proc. Natl. Acad. Sci. U. S. A. 93, 8329–33 (1996).

48. Jiang, Y. W. et al. Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. Proc. Natl. Acad. Sci. U. S. A. 95, 8538–43 (1998).

49. Näär, A. M. et al. Chromatin, TAFs, and a novel multiprotein coactivator are required for synergistic activation by Sp1 and SREBP-1a in vitro. Genes Dev. 12, 3020–31 (1998).

50. Hein, M. Y. et al. A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. Cell 163, 712–723 (2015).

51. Näär, A. M. et al. Composite co-activator ARC mediates chromatin-directed transcriptional activation. Nature 398, 828–832 (1999).

52. Covic, M. et al. Arginine methyltransferase CARM1 is a promoter-specific regulator of NF-κB-dependent gene expression. EMBO J. 24, 85–96 (2005).

53. Stallcup, M. R. et al. The roles of protein-protein interactions and protein methylation in transcriptional activation by nuclear receptors and their coactivators. J. Steroid Biochem. Mol. Biol. 85, 139–45 (2003).

54. Chen, M. et al. JMJD1C is required for the survival of acute myeloid leukemia by functioning as a coactivator for key transcription factors. Genes Dev. 29, 2123–2139 (2015).

55. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. Cell 132, 1049–1061 (2008).

56. Wang, J. et al. A protein interaction network for pluripotency of embryonic stem cells. Nature 444, 364–368 (2006).

57. Amiel, J. et al. Mutations in TCF4, Encoding a Class I Basic Helix-Loop-Helix Transcription Factor, Are Responsible for Pitt-Hopkins Syndrome, a Severe Epileptic Encephalopathy Associated with Autonomic Dysfunction. Am. J. Hum. Genet. 80, 988–993 (2007).

58. Zweier, C. et al. Haploinsufficiency of TCF4 Causes Syndromal Mental Retardation with Intermittent Hyperventilation (Pitt-Hopkins Syndrome). Am. J. Hum. Genet. 80, 994–1001 (2007).

59. Stefansson, H. et al. Common variants conferring risk of schizophrenia. Nature 460, 744–7 (2009).

60. Aberg, K. A. et al. A Comprehensive Family-Based Replication Study of Schizophrenia Genes. JAMA Psychiatry 70, 573 (2013).

61. Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat. Biotechnol. 21, 183–186 (2003).

62. Conti, L. et al. Niche-Independent Symmetrical Self-Renewal of a Mammalian Tissue Stem Cell. PLoS Biol. 3, e283 (2005).

63. Engelen, E. et al. Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. Nat. Genet. 43, 607–611 (2011).

64. van den Berg, D. L. C. et al. An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells. Cell Stem Cell 6, 369–381 (2010).

65. Ishihama, Y. et al. Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. Mol. Cell. Proteomics 4, 1265–1272 (2005).

66. Chatr-aryamontri, A. et al. The BioGRID interaction database: 2017 update. Nucleic Acids Res. 45, D369–D379 (2017).

67. Shannon, P. et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 13, 2498–2504 (2003).

68. Beagan, J. A. et al. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. Genome Res. 27, 1139–1152 (2017).

69. Piper, M. et al. NFIA Controls Telencephalic Progenitor Cell Differentiation through Repression of the Notch Effector Hes1. J. Neurosci. 30, 9127–9139 (2010).

70. Pevny, L. H. & Nicolis, S. K. Sox2 roles in neural stem cells. Int. J. Biochem. Cell Biol. 42, 421–424 (2010).

71. Betancourt, J., Katzman, S. & Chen, B. Nuclear factor one B regulates neural stem cell differentiation and axonal projection of corticofugal neurons. J. Comp. Neurol. 522, 6–35 (2014).

72. Mesman, S. & Smidt, M. P. Tcf12 Is Involved in Early Cell-Fate Determination and Subset Specification of Midbrain Dopamine Neurons. Front. Mol. Neurosci. 10, 353 (2017).

73. Khalfallah, O. et al. Zinc Finger Protein 191 (ZNF191/Zfp191) is necessary to maintain neural cells as cycling progenitors. Stem Cells 27, 1643–1653 (2009).

74. Methot, L. et al. Interaction and Antagonistic Roles of NF- B and Hes6 in the Regulation of Cortical Neurogenesis. Mol. Cell. Biol. 33, 2797–2808 (2013).

75. Young, K. M., Bartlett, P. F. & Coulson, E. J. Neural progenitor number is regulated by nuclear factor-κB p65 and p50 subunit-dependent proliferation rather than cell survival. J. Neurosci. Res. 83, 39–49 (2006).

76. Chen, T. et al. Tcf4 Controls Neuronal Migration of the Cerebral Cortex through Regulation of Bmp7. Front. Mol. Neurosci. 9, 94 (2016).

77. Flora, A., Garcia, J. J., Thaller, C. & Zoghbi, H. Y. The E-protein Tcf4 interacts with Math1 to regulate differentiation of a specific subset of neuronal progenitors. Proc. Natl. Acad. Sci. 104, 15382–15387 (2007).

78. Boyer, L. A. et al. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. Cell 122, 947–956 (2005).

79. Soler, E. et al. The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. Genes Dev. 24, 277–289 (2010).

80. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).

81. Heinz, S. et al. Finding Enriched Peaks, Regions, and Transcripts (findPeaks). (2017).

82. Martynoga, B. et al. Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. Genes Dev. 27, 1769–1786 (2013).

83. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–2 (2010).

84. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13 (2009).

SUPPLEMENTAL INFORMATION

Figure Supplemental 1 | Mediator complex and its interactors at promoters, related to Figure 5

Figure Supplemental 2 | Tcf4 and its interactors at super enhancers and broad H3k4me3 promoters, related to Figure 6

Table Supplemental 1 | Mediator curated interactome in NSCs (proteomics)

Table Supplemental 2 | SEs in NSCs (genomic coordinates)

Table Supplemental 3 | GO terms of SE, Broad and SE+Broad genes

Table Supplemental 4 | Genes with SE(s) and broad H3K4me3 promoters in NSCs

Table Supplemental 5 | Datasets used in the study

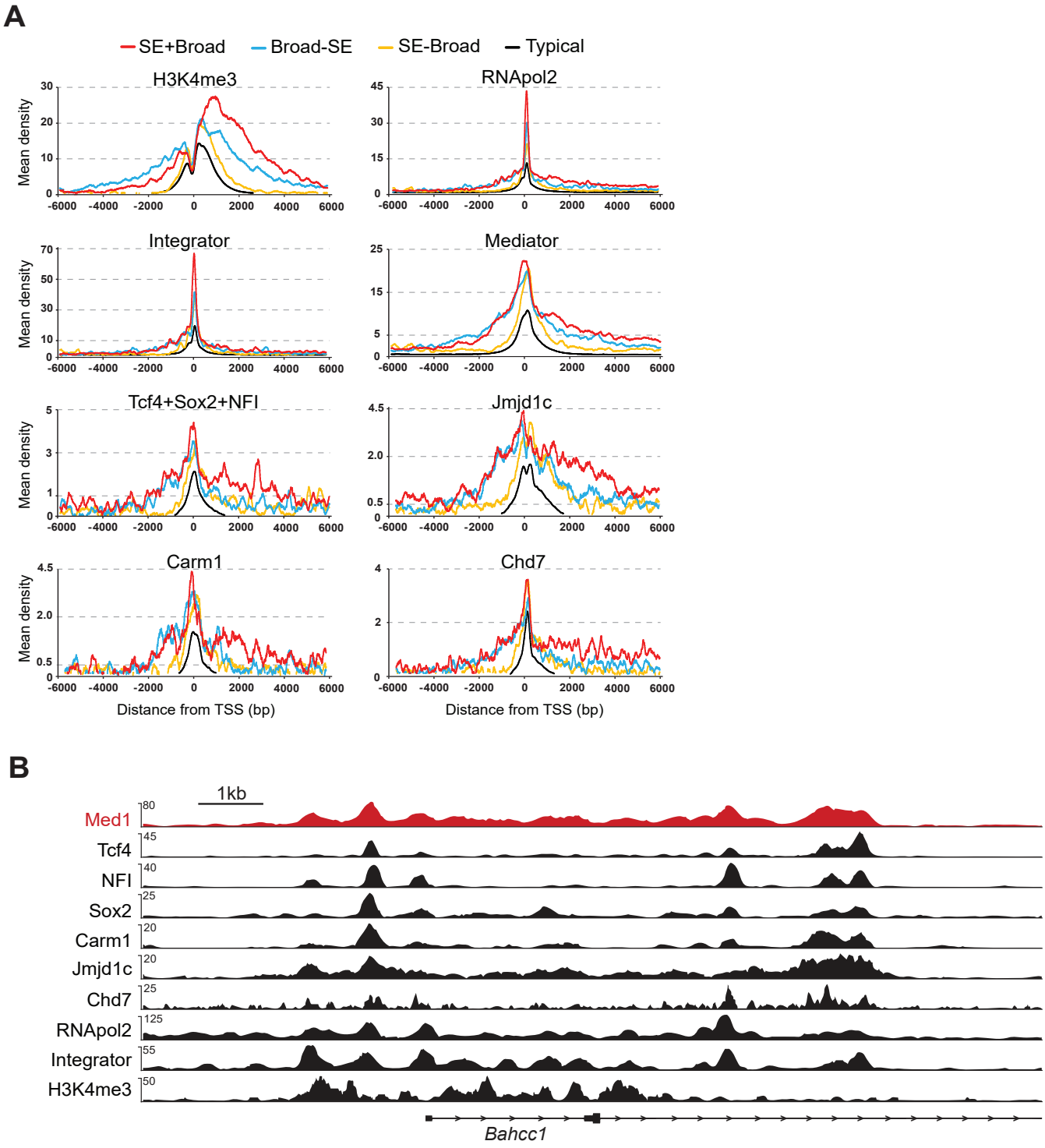


Figure S1 | Mediator complex and its interactors at promoters, related to Figure 5

(A) ChIP-seq density plots around top 100 promoters of Broad+SE genes, top 100 Broad-SE genes, top 100 SE-Broad genes and all typical genes for the indicated factors and histone modifications. Mean ChIP-seq density (y-axis) and distance to TSS (x-axis) are shown. (B) Overlap of binding sites of Med1 with binding sites of Mediator interactors at the Bahcc1 broad H3K4me3 promotor area NSCs. ChIP-seq tracks for the indicated proteins and histone modifications at the Hes1 gene are shown. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated.

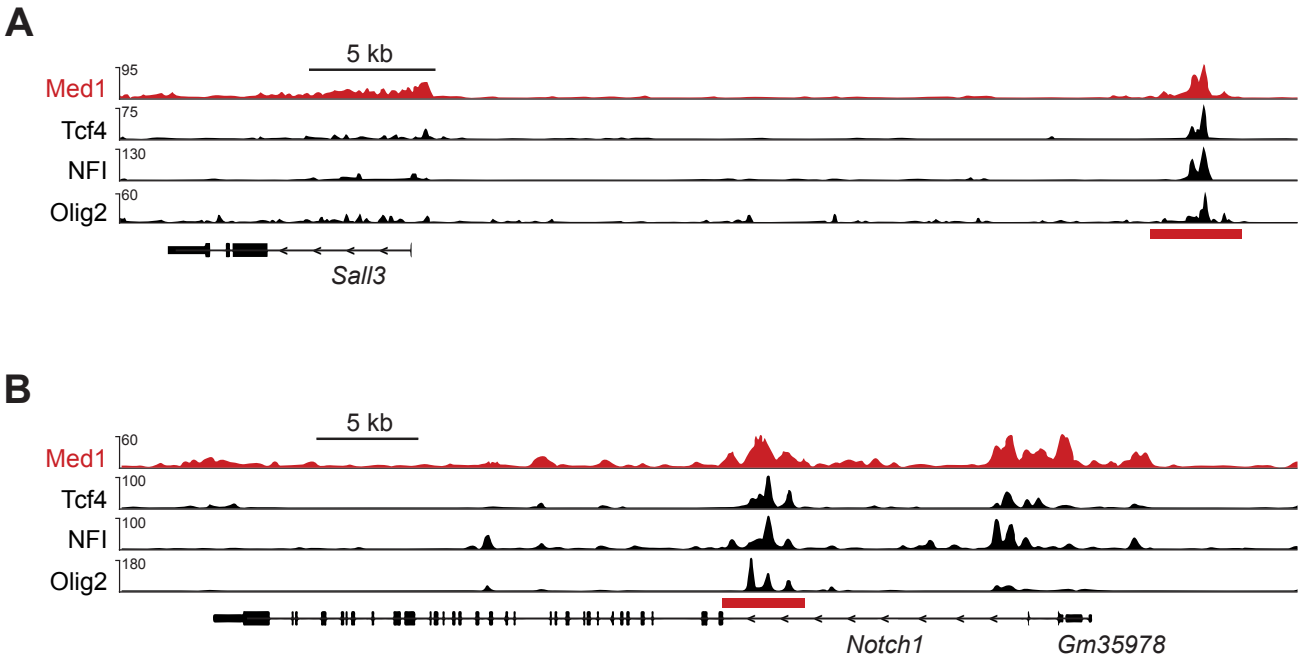


Figure S2 | Tcf4 and its interactors at super enhancers and broad H3k4me3 promoters, related to Figure 6
(A) Overlap of binding sites of Tcf4 and Med1 with binding sites of Tcf4-interactors Olig2 and NFI at the Sall3 locus in NSCs. ChIP-seq tracks for the indicated proteins at the Sall3 gene are shown. The Sall3 SE is indicated with a red bar. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated. (B) Overlap of binding sites of Tcf4 and Med1 with binding sites of Tcf4-interactors Olig2 and NFI at the Notch1 locus in NSCs. ChIP-seq tracks for the indicated proteins at the Notch1 gene are shown. The Notch1 SE is indicated with a red bar. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated.

Table Supplemental 1 | Mediator curated interactome in NSCs (proteomics)

accession	symbol	description	Med15 benzo - unique pepts	Med15 NO benzo - unique pepts	Med15 EtBr - unique pepts	Control benzo - unique pepts	Med15 benzo rep - unique pepts	Control benzo rep - unique pepts	**Average unique peptides**	Med15 benzo - emPAI	Med15 NO benzo - emPAI	Med15 EtBr - emPAI	Control benzo - emPAI	Med15 benzo rep - emPAI	Control benzo rep- emPAI	**Average emPAI**	Med15 benzo - Mascot score	Med15 NO benzo - Mascot score	Med15 EtBr - Mascot score	Control benzo - Mascot score	Med15 benzo rep - Mascot score	Control benzo rep- Mascot score	**Average Mascot score**
Mediator complex																							
G3X8S4	Med15	Mediator of RNA polymerase II transcription subunit 15	29	30	29	0	35	0	**31**	7,03	7,37	6,70	0,00	36,51	0,00	**14,40**	1927	1971	1939	0	2489	0	**2082**
Q9CQA5	Med4	Mediator of RNA polymerase II transcription subunit 4	10	10	8	0	7	0	**9**	4,99	3,72	3,72	0,00	2,71	0,00	**3,79**	508	498	394	0	337	0	**434**
Q9DB40	Med27	Mediator of RNA polymerase II transcription subunit 27	13	14	12	0	10	0	**12**	4,01	4,54	3,10	0,00	2,35	0,00	**3,50**	623	752	649	0	661	0	**671**
Q9R0X0	Med20	Mediator of RNA polymerase II transcription subunit 20	6	6	6	0	8	0	**7**	3,46	2,31	2,31	0,00	4,18	0,00	**3,07**	294	286	303	0	497	0	**345**
Q9DAY7	Med8	Mediator of RNA polymerase II transcription subunit 8	9	9	9	0	9	0	**9**	2,39	2,39	2,82	0,00	2,82	0,00	**2,61**	364	450	440	0	458	0	**428**
Q921D4	Med6	Mediator of RNA polymerase II transcription subunit 6	9	8	9	0	11	0	**9**	2,07	2,07	2,07	0,00	4,06	0,00	**2,57**	414	403	449	0	528	0	**449**
A2ABV5	Med14	Mediator of RNA polymerase II transcription subunit 14	35	35	38	0	45	0	**38**	1,56	1,81	1,68	0,00	2,93	0,00	**2,00**	1907	1808	1909	0	2270	0	**1974**
Q920D3	Med28	Mediator of RNA polymerase II transcription subunit 28	4	5	4	0	5	0	**5**	1,45	1,92	1,45	0,00	1,92	0,00	**1,69**	182	209	175	0	252	0	**205**
Q9CXU0	Med10	Mediator of RNA polymerase II transcription subunit 10	5	4	3	0	3	0	**4**	2,02	1,42	0,94	0,00	0,94	0,00	**1,33**	246	234	146	0	133	0	**190**
Q62276	Med22	Mediator of RNA polymerase II transcription subunit 22	4	5	3	0	3	0	**4**	1,43	1,91	1,04	0,00	0,71	0,00	**1,27**	196	268	186	0	179	0	**207**
Q8VCD5	Med17	Mediator of RNA polymerase II transcription subunit 17	16	19	16	0	20	0	**18**	1,21	1,57	1,32	0,00	2,13	0,00	**1,56**	825	974	820	0	1169	0	**947**
Q9CZ82	Med18	Mediator of RNA polymerase II transcription subunit 18	4	4	4	0	8	0	**5**	1,44	1,44	0,81	0,00	4,18	0,00	**1,97**	159	151	131	0	497	0	**235**
A2AGH9	Med12	Mediator of RNA polymerase II transcription subunit 12	46	49	44	0	50	0	**47**	1,03	1,22	1,00	0,00	1,35	0,00	**1,15**	2354	2436	2210	0	2635	0	**2409**
Q99K74	Med24	Mediator of RNA polymerase II transcription subunit 24	22	19	19	0	19	0	**20**	1,11	1,11	0,98	0,00	1,11	0,00	**1,08**	1085	1110	1057	0	1074	0	**1082**
G3UW74	Med16	Mediator of RNA polymerase II transcription subunit 16	18	19	15	0	23	0	**19**	1,03	1,19	0,96	0,00	1,28	0,00	**1,12**	863	964	824	0	1038	0	**922**
E9QNV2	Med23	Mediator of RNA polymerase II transcription subunit 23	28	29	32	0	26	0	**29**	0,86	0,95	1,19	0,00	0,86	0,00	**0,97**	1483	1577	1681	0	1270	0	**1503**
Q3UXL9	Ccnc	Cyclin-C	5	6	8	0	8	0	**7**	0,75	0,96	1,19	0,00	1,61	0,00	**1,13**	255	240	329	0	400	0	**306**
Q9DB91	Med29	Mediator of RNA polymerase II transcription subunit 29	4	3	3	0	5	0	**4**	1,28	0,93	0,64	0,00	2,73	0,00	**1,40**	178	184	167	0	245	0	**194**
Q9CQI9	Med30	Mediator of RNA polymerase II transcription subunit 30	2	3	3	0	7	0	**4**	0,67	0,98	0,98	0,00	2,32	0,00	**1,24**	148	178	190	0	326	0	**211**
Q8BWD8	Cdk19	Cyclin-dependent kinase 19	7	10	6	0	6	0	**7**	0,66	1,01	0,56	0,00	0,56	0,00	**0,70**	285	355	219	0	323	0	**296**
Q5SWW4	Med13	Mediator of RNA polymerase II transcription subunit 13	31	30	28	0	32	0	**30**	0,67	0,67	0,57	0,00	0,70	0,00	**0,65**	1498	1540	1427	0	1670	0	**1534**
Q925J9	Med1	Mediator of RNA polymerase II transcription subunit 1	20	21	15	0	28	0	**21**	0,61	0,68	0,48	0,00	1,23	0,00	**0,75**	964	966	773	0	1518	0	**1055**
Q8VCB2	Med25	Mediator of RNA polymerase II transcription subunit 25	6	5	5	0	6	0	**6**	0,38	0,32	0,26	0,00	0,32	0,00	**0,32**	316	304	300	0	300	0	**305**
Q7TN02	Med26	Mediator of RNA polymerase II transcription subunit 26	3	3	3	0	4	0	**3**	0,18	0,18	0,18	0,00	0,25	0,00	**0,20**	140	133	117	0	174	0	**141**
E9QLJ3	Med13l	Mediator of RNA polymerase II transcription subunit 13-like	8	10	8	0	6	0	**8**	0,13	0,18	0,14	0,00	0,08	0,00	**0,13**	346	421	386	0	276	0	**357**
p300/CBP																							
F8VPR5	Crebbp	CREB-binding protein	55	51	48	0	75	0	**57**	1,32	1,32	1,14	0,00	3,46	0,00	**1,81**	2523	2576	2311	0	4240	0	**2913**
B2RWS6	Ep300	Histone acetyltransferase p300	46	51	48	0	64	0	**52**	1,10	1,41	1,18	0,00	2,83	0,00	**1,63**	2122	2346	2202	0	3447	0	**2529**
NurD complex																							
Q9Z2D8	Mbd3	Methyl-CpG-binding domain protein 3	5	6	7	0	8	0	**7**	0,73	1,41	2,01	0,00	2,36	0,00	**1,63**	275	357	380	0	431	0	**361**
E9QAS5	Chd4	Chromodomain-helicase-DNA-binding protein 4	32	25	31	0	50	1	**35**	0,79	0,54	0,82	0,00	1,76	0,00	**0,98**	1334	1005	1301	0	2338	44	**1495**
Q9R190	Mta2	Metastasis-associated protein MTA2	10	11	11	2	23	5	**14**	0,61	0,69	0,69	0,10	2,47	0,27	**1,12**	454	535	502	85	1194	164	**671**
E9QMN5	Gatad2a	Transcriptional repressor p66 alpha	6	6	4	0	14	2	**8**	0,38	0,38	0,24	0,00	1,35	0,11	**0,59**	289	279	177	0	681	95	**357**
Q8VHR5	Gatad2b	Transcriptional repressor p66-beta	5	5	5	0	18	4	**8**	0,32	0,32	0,32	0,00	2,54	0,25	**0,88**	260	220	278	0	858	175	**404**
F8WHY8	Mta1	Metastasis-associated protein MTA1	5	8	5	0	19	2	**9**	0,25	0,37	0,25	0,00	1,48	0,09	**0,59**	227	314	228	0	1012	61	**445**
RNA pol II																							
Q80UW8	Polr2e	DNA-directed RNA polymerases I, II, and III subunit RPABC1	4	4	5	0	2	0	**4**	1,05	1,05	1,37	0,00	0,33	0,00	**0,95**	157	252	254	0	86	0	**187**
A0A087WR0	Taf9	Transcription initiation factor TFIID subunit 9	3	2	2	0	6	0	**3**	0,42	0,42	0,42	0,00	1,08	0,00	**0,59**	94	98	106	0	219	0	**129**
P08775	Polr2a	DNA-directed RNA polymerase II subunit RPB1	7	3	2	0	12	0	**6**	0,12	0,05	0,03	0,00	0,22	0,00	**0,11**	244	124	75	0	516	0	**240**
Q99J95	Cdk9	Cyclin-dependent kinase 9	3	2	2	0	7	0	**4**	0,28	0,18	0,18	0,00	0,79	0,00	**0,36**	117	98	113	0	410	0	**185**
Compass complex																							
Q8BFQ4	Wdr82	WD repeat-containing protein 82	5	7	5	0	9	0	**7**	0,66	1,03	0,83	0,00	2,72	0,00	**1,31**	179	300	246	0	436	0	**290**
Q8BX09	Rbbp5	Retinoblastoma-binding protein 5	3	2	2	0	2	0	**2**	0,20	0,13	0,13	0,00	0,13	0,00	**0,15**	109	84	92	0	52	0	**84**
D3YYA0	Ash2l	Set1/Ash2 histone methyltransferase complex subunit ASH2	1	2	2	0	3	0	**2**	0,07	0,14	0,14	0,00	0,21	0,00	**0,14**	66	109	102	0	128	0	**101**
SWI/SNF complex																							
Q61466	Smarcd1	SWI/SNF-related matrix-associated actin-dependent regulator of chro	5	5	8	0	8	0	**7**	0,36	0,36	0,64	0,00	0,74	0,00	**0,53**	190	247	275	0	388	0	**275**
O54941	Smarce1	SWI/SNF-related matrix-associated actin-dependent regulator of chro	4	6	3	0	7	1	**5**	0,36	0,59	0,36	0,00	0,59	0,08	**0,48**	204	276	227	0	393	78	**275**
Q3UID0	Smarcc2	5. SWI/SNF complex subunit SMARCC2	8	6	8	0	8	0	**8**	0,23	0,23	0,30	0,00	0,30	0,00	**0,27**	486	371	349	0	433	0	**410**
Ncor complex																							
Q8BHJ5	Tbl1xr1	F-box-like/WD repeat-containing protein TBL1XR1	9	11	9	0	15	1	**11**	0,78	1,02	0,78	0,00	2,38	0,07	**1,24**	485	651	530	0	1016	62	**671**
Q9QXE7	Tbl1x	F-box-like/WD repeat-containing protein TBL1X	6	6	7	0	9	0	**7**	0,46	0,46	0,55	0,00	0,87	0,00	**0,59**	349	354	384	0	557	0	**411**
F8VQL9	Ncor2	Nuclear receptor corepressor 2	16	11	8	0	44	0	**20**	0,24	0,14	0,11	0,00	0,82	0,00	**0,33**	665	495	356	0	1930	0	**862**
Q5RIM6	Ncor1	Nuclear receptor co-repressor 1, isoform CRA_a	9	6	7	0	16	0	**10**	0,11	0,07	0,13	0,00	0,26	0,00	**0,14**	431	254	276	0	755	0	**429**
Lsd1 complex																							
Q6ZQ88	Kdm1a	Lysine-specific histone demethylase 1A	9	9	7	0	15	3	**10**	0,36	0,41	0,31	0,00	0,84	0,12	**0,48**	372	507	379	0	660	97	**480**
Q8C796	Rcor2	REST corepressor 2	3	3	3	0	6	0	**4**	0,21	0,28	0,21	0,00	0,45	0,00	**0,29**	164	179	165	0	303	0	**203**
Z4YJZ7	Ehmt1	Histone-lysine N-methyltransferase EHMT1	3	3	3	0	7	0	**4**	0,11	0,11	0,11	0,00	0,19	0,00	**0,13**	97	118	121	0	246	0	**146**
Coactivator																							
Q9WVG6	Carm1	Histone-arginine methyltransferase CARM1	11	10	11	0	13	0	**11**	0,84	0,75	0,75	0,00	1,15	0,00	**0,87**	573	578	571	0	678	0	**600**
Q61026	Ncoa2	Nuclear receptor coactivator 2	18	23	21	0	32	0	**24**	0,58	0,95	0,73	0,00	2,64	0,00	**1,23**	953	1229	1107	0	1941	0	**1308**
Q9CXY6	Ilf2	Interleukin enhancer-binding factor 2	5	7	6	0	10	0	**7**	0,51	0,79	0,65	0,00	1,71	0,00	**0,92**	315	414	370	0	558	0	**414**
Q9R059	Fhl3	Four and a half LIM domains protein 3	4	4	4	0	7	0	**5**	0,53	0,70	0,53	0,00	1,10	0,00	**0,72**	193	257	203	0	313	0	**242**
G3X8R8	Mkl2	MCG123888	15	15	12	0	25	0	**17**	0,63	0,63	0,44	0,00	1,75	0,00	**0,86**	681	674	542	0	1337	0	**809**
Q9WUP7	Uchl5	Ubiquitin carboxyl-terminal hydrolase isozyme L5	4	5	4	0	4	1	**4**	0,46	0,61	0,46	0,00	0,61	0,10	**0,54**	141	252	177	0	163	44	**183**
P70365	Ncoa1	Nuclear receptor coactivator 1	13	14	15	0	30	0	**18**	0,41	0,41	0,44	0,00	1,34	0,00	**0,65**	747	719	713	0	1823	0	**1001**
A2AJK6	Chd7	Chromodomain-helicase-DNA-binding protein 7	28	24	21	0	58	0	**33**	0,36	0,30	0,26	0,00	1,03	0,00	**0,49**	1270	1137	911	0	2824	0	**1536**
E9Q8Z6	Ctnnd1	Catenin delta-1	6	9	7	0	13	0	**9**	0,23	0,44	0,34	0,00	0,57	0,00	**0,40**	279	455	360	0	569	0	**416**
Q45VK5	Ilf3	Interleukin enhancer-binding factor 3	5	8	7	0	7	0	**7**	0,21	0,35	0,30	0,00	0,30	0,00	**0,29**	237	348	248	0	262	0	**274**
D3YUG5	Mkl1	MKL/myocardin-like protein 1	5	4	5	0	22	0	**9**	0,27	0,21	0,21	0,00	1,51	0,00	**0,55**	232	192	225	0	1039	0	**422**
F6U238	Maml2	Protein Maml2	5	4	6	0	11	0	**7**	0,18	0,18	0,22	0,00	0,54	0,00	**0,28**	239	148	239	0	572	0	**300**
Q9QYH6	Maged1	Melanoma-associated antigen D1	4	5	4	0	11	0	**6**	0,18	0,24	0,14	0,00	0,59	0,00	**0,29**	142	157	161	0	618	0	**270**
Q91ZW3	Smarca5	SWI/SNF-related matrix-associated actin-dependent regulator of chro	4	5	7	0	12	0	**7**	0,13	0,16	0,23	0,00	0,43	0,00	**0,24**	152	171	221	0	456	0	**250**
Q8CH18	Ccar1	Cell division cycle and apoptosis regulator protein 1	2	2	5	0	2	0	**3**	0,06	0,06	0,15	0,00	0,06	0,00	**0,08**	79	76	152	0	79	0	**97**
A0A0A0MQ9	Jmjd1c	Probable JmjC domain-containing histone demethylation protein 2C	5	6	3	0	24	0	**10**	0,07	0,08	0,04	0,00	0,38	0,00	**0,14**	238	288	179	0	1074	0	**445**
Q9DBR0	Akap8	A-kinase anchor protein 8	8	6	7	0	14	0	**9**	0,46	0,39	0,39	0,00	1,03	0,00	**0,57**	393	329	381	0	683	0	**447**
A2A655	Bptf	Protein Bptf	6	5	2	0	36	0	**12**	0,07	0,06	0,02	0,00	0,51	0,00	**0,17**	264	200	113	0	1626	0	**551**

Table Supplemental 1 | *continuation*

TRRAP complex																								
E9PZA7	Trrap	Transformation/transcription domain-associated protein	24	26	16	0	41	0	**27**	0,23	0,25	0,14	0,00	0,43	0,00	**0,26**	1013	1124	694	0	1850	0	**1170**
Q8CHI8	Ep400	E1A-binding protein p400	6	6	4	0	26	0	**11**	0,06	0,07	0,23	0,00	1,80	0,00	**0,54**	214	243	173	0	1238	0	**467**
Cohesin																								
Q9CU62	Smc1a	Structural maintenance of chromosomes protein 1A	7	7	5	0	14	0	**8**	0,19	0,19	0,13	0,00	0,43	0,00	**0,24**	237	297	197	0	559	0	**323**
Transcription Factors																								
Q8BHG9	Cggbp1	CGG triplet repeat-binding protein 1	4	3	4	0	5	0	**4**	1,09	0,74	1,09	0,00	2,03	0,00	**1,24**	122	129	145	0	185	0	**145**
Q00899	Yy1	Transcriptional repressor protein YY1	5	6	4	0	9	0	**6**	0,49	0,61	0,37	0,00	1,05	0,00	**0,63**	211	255	145	0	434	0	**261**
B1AUC0	Nfia	Nuclear factor 1	6	6	5	0	11	0	**7**	0,49	0,49	0,40	0,00	1,09	0,00	**0,62**	293	304	246	0	561	0	**351**
Q60I23	Sox2	SRY-box containing gene 2	3	5	4	0	2	0	**4**	0,37	0,68	0,68	0,00	0,22	0,00	**0,49**	101	204	148	0	82	0	**134**
A2BG76	Nfib	Nuclear factor 1	5	5	6	0	9	0	**6**	0,33	0,33	0,41	0,00	0,68	0,00	**0,44**	287	275	310	0	458	0	**333**
Q62255	Sall3	Sal-like protein 3	10	9	9	0	24	2	**13**	0,40	0,36	0,26	0,00	1,23	0,05	**0,56**	517	451	387	0	1216	94	**643**
Q9CU65	Zmym2	Zinc finger MYM-type protein 2	12	11	12	0	16	2	**13**	0,32	0,26	0,32	0,00	0,48	0,05	**0,35**	548	482	621	0	754	69	**601**
Q61286	Tcf12	Transcription factor 12	4	4	3	0	16	0	**7**	0,30	0,30	0,22	0,00	1,14	0,00	**0,49**	176	190	135	0	905	0	**352**
Q91VN1	Znf24	Zinc finger protein 24	3	3	2	0	9	0	**4**	0,29	0,18	0,18	0,00	1,14	0,00	**0,45**	135	128	80	0	392	0	**184**
F7ARK3	Sall2	Sal-like protein 2	6	4	5	0	7	0	**6**	0,25	0,17	0,17	0,00	0,29	0,00	**0,22**	270	239	183	0	375	0	**267**
Q04207	Rela	Transcription factor p65	2	2	4	0	12	0	**5**	0,13	0,13	0,27	0,00	1,04	0,00	**0,39**	97	86	197	0	507	0	**222**
Q61624	Znf148	Zinc finger protein 148	3	2	3	0	12	0	**5**	0,13	0,08	0,13	0,00	0,69	0,00	**0,26**	158	94	155	0	716	0	**281**
Q925H1	Trps1	Zinc finger transcription factor Trps1	4	5	4	0	24	0	**9**	0,08	0,11	0,11	0,00	0,79	0,00	**0,27**	158	188	145	0	1079	0	**393**
Q99LI5	Znf281	Zinc finger protein 281	2	2	2	0	5	0	**3**	0,08	0,08	0,08	0,00	0,20	0,00	**0,11**	114	125	87	0	178	0	**126**
B1AWL2	Zfp462	Protein Zfp462	5	6	7	0	8	0	**7**	0,07	0,08	0,09	0,00	0,09	0,00	**0,08**	165	213	256	0	256	0	**223**
E9Q8G4	Tcf4	Transcription factor 4	1	1	1	0	11	0	**4**	0,05	0,05	0,05	0,00	0,83	0,00	**0,25**	53	72	76	0	487	0	**172**
Post-translation modification																								
Q99PQ2	Trim11	E3 ubiquitin-protein ligase TRIM11	13	16	16	0	16	0	**15**	2,14	2,84	2,84	0,00	3,10	0,00	**2,73**	955	1070	1024	0	1093	0	**1036**
A2AMY5	Ubap2	Ubiquitin-associated protein 2	6	4	2	0	12	0	**6**	0,20	0,13	0,06	0,00	0,59	0,00	**0,25**	236	162	74	0	614	0	**272**
E9Q2H1	Ubr5	E3 ubiquitin-protein ligase UBR5	10	11	8	0	26	0	**14**	0,12	0,15	0,09	0,00	0,40	0,00	**0,19**	464	456	357	0	1148	0	**606**
Q8CGY8	Ogt	UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase	4	3	3	0	12	0	**6**	0,13	0,10	0,10	0,00	0,49	0,00	**0,21**	158	116	91	0	667	0	**258**
Q8N7N5	Dcaf8	DDB1- and CUL4-associated factor 8	2	2	1	0	2	0	**2**	0,11	0,11	0,06	0,00	0,11	0,00	**0,10**	72	75	50	0	78	0	**69**
E9QME5	Trim33	E3 ubiquitin-protein ligase TRIM33	4	3	2	0	3	0	**3**	0,12	0,09	0,06	0,00	0,09	0,00	**0,09**	177	178	85	0	160	0	**150**
O88196	Ttc3	E3 ubiquitin-protein ligase TTC3	3	3	2	0	7	0	**4**	0,05	0,05	0,03	0,00	0,12	0,00	**0,06**	134	160	65	0	250	0	**152**
mRNA binding																								
Q9Z2X1	Hnrnpf	Heterogeneous nuclear ribonucleoprotein F	11	10	10	2	8	0	**10**	1,98	1,76	1,98	0,00	1,76	0,00	**1,87**	640	530	559	57	472	0	**550**
P97315	Csrp1	Cysteine and glycine-rich protein 1	4	3	5	0	4	0	**4**	0,93	0,39	1,28	0,00	0,93	0,00	**0,88**	172	144	199	0	191	0	**177**
Q8VDP4	Ccar2	Cell cycle and apoptosis regulator protein 2	15	13	12	0	11	0	**13**	0,63	0,58	0,52	0,00	0,47	0,00	**0,55**	581	492	498	0	470	0	**510**
A0A0A6YW8	Mbnl1	Muscleblind-like protein 1	3	3	3	0	4	0	**3**	0,29	0,29	0,29	0,00	0,41	0,00	**0,32**	102	98	99	0	157	0	**114**
Q91YT7	Ythdf2	YTH domain-containing family protein 2	5	4	4	0	14	0	**7**	0,34	0,26	0,26	0,00	1,53	0,00	**0,60**	211	196	185	0	720	0	**328**
Q9EPU0	Upf1	Regulator of nonsense transcripts 1	9	10	6	0	15	2	**10**	0,30	0,34	0,19	0,00	0,55	0,06	**0,35**	472	507	289	0	743	76	**503**
Q8C2Q3	Rbm14	RNA-binding protein 14	5	5	4	0	13	0	**7**	0,30	0,30	0,23	0,00	1,07	0,00	**0,48**	251	239	173	0	730	0	**348**
O88532	Zfr	Zinc finger RNA-binding protein	4	5	5	0	11	0	**6**	0,13	0,17	0,17	0,00	0,45	0,00	**0,23**	145	281	196	0	592	0	**304**
Q9QYS9	Qki	Protein quaking	2	2	2	0	6	0	**3**	0,21	0,21	0,21	0,00	0,94	0,00	**0,39**	98	99	81	0	249	0	**132**
B7ZWL1	Cnot1	CCR4-NOT transcription complex subunit 1	13	10	13	0	26	0	**16**	0,19	0,15	0,19	0,00	0,46	0,00	**0,25**	623	555	616	0	1282	0	**769**
O35218	Cpsf2	Cleavage and polyadenylation specificity factor subunit 2	3	2	2	0	2	0	**2**	0,13	0,08	0,08	0,00	0,10	0,00	**0,10**	106	93	92	0	95	0	**97**
Q9EPU4	Cpsf1	Cleavage and polyadenylation specificity factor subunit 1	5	5	3	0	9	0	**6**	0,12	0,12	0,05	0,00	0,20	0,00	**0,12**	216	194	97	0	366	0	**218**
Integrator complex																								
Q8K2A7	Ints10	Integrator complex subunit 10	2	2	2	0	2	0	**2**	0,09	0,09	0,09	0,00	0,14	0,00	**0,10**	86	97	77	0	106	0	**92**
Q6P4S8	Ints1	Integrator complex subunit 1	4	4	3	0	7	0	**5**	0,06	0,06	0,04	0,00	0,11	0,00	**0,07**	157	135	126	0	295	0	**178**
Corepressors																								
P23198	Cbx3	Chromobox protein homolog 3	4	2	4	0	5	0	**4**	0,96	0,65	1,32	0,00	2,24	0,00	**1,29**	169	120	217	0	275	0	**195**
Q7TSZ8	Nacc1	Nucleus accumbens-associated protein 1	5	5	7	0	13	0	**8**	0,37	0,37	0,55	0,00	1,74	0,00	**0,76**	247	287	303	0	720	0	**389**
Q80X50	Ubap2l	Ubiquitin-associated protein 2-like	11	9	12	0	21	0	**13**	0,45	0,36	0,45	0,00	1,39	0,00	**0,66**	515	427	495	0	1129	0	**642**
Q9CQJ4	Rnf2	E3 ubiquitin-protein ligase RING2	5	3	3	0	7	0	**5**	0,60	0,33	0,33	0,00	1,33	0,00	**0,65**	252	147	137	0	363	0	**225**
Q60520	Sin3a	Paired amphipathic helix protein Sin3a	11	10	12	0	6	0	**10**	0,32	0,32	0,42	0,00	0,16	0,00	**0,31**	443	386	528	0	269	0	**407**
A2BDX0	Adnp	Activity-dependent neuroprotector homeobox protein	4	4	3	0	3	0	**4**	0,12	0,12	0,09	0,00	0,09	0,00	**0,11**	136	180	119	0	137	0	**143**
other																								
A2BIE1	Qser1	Protein Qser1	19	15	15	0	33	0	**21**	0,45	0,34	0,34	0,00	1,28	0,00	**0,60**	892	784	704	0	1857	0	**1059**
Q3UA37	Qrich1	Glutamine-rich protein 1	6	8	6	0	11	0	**8**	0,28	0,40	0,34	0,00	0,87	0,00	**0,47**	296	315	242	0	538	0	**348**
Q9DCT8	Crip2	Cysteine-rich protein 2	2	2	3	0	2	0	**2**	0,35	0,35	0,57	0,00	0,57	0,00	**0,46**	75	138	178	0	127	0	**130**
P70168	Kpnb1	Importin subunit beta-1	6	7	8	0	8	0	**7**	0,25	0,29	0,34	0,00	0,34	0,00	**0,31**	331	362	326	0	417	0	**359**
Q3U1J4	Ddb1	DNA damage-binding protein 1	16	11	14	0	15	0	**14**	0,67	0,37	0,45	0,00	0,49	0,00	**0,50**	810	479	577	0	668	0	**634**
Q9Z191	Eya4	Eyes absent homolog 4	2	1	2	0	5	0	**3**	0,11	0,06	0,11	0,00	0,24	0,00	**0,13**	72	68	73	0	193	0	**102**

Table Supplemental 2 | SEs in NSCs (genomic coordinates)

Chr	Start	End	Nearest active gene	SE Ranking
chr11	101502299	101519975	Arl4d	1
chr15	85482117	85519888	Wnt7b/Mirlet7c-2	2
chr11	87275110	87285154	Rad51c	3
chr11	87256127	87262339	Rad51c	4
chr11	87235422	87240416	Rad51c	5
chr12	88110704	88167882	Angel1	6
chr11	98786982	98833230	Rara	7
chr4	97268244	97333957	Nfia	8
chr1	72272346	72302938	Mreg	9
chr11	87931123	87976574	Cuedc1	10
chr4	116861522	116875583	Kif2c	11
chr15	98798082	98824335	Tuba1a	12
chr5	115938824	115942514	Sirt4	13
chr11	117156345	117182248	Sept9	14
chr14	26543039	26577940	Zcchc24	15
chr18	5130392	5165290	Zfp438	16
chr7	80729113	80766141	1810026B05Rik	17
chr11	120112669	120126041	Bahcc1	18
chr15	77611812	77655136	Myh9	19
chr18	37982390	38012569	Pcdhgc family	20
chr15	85429940	85446290	Wnt7b/Mirlet7c-2	21
chr16	30171292	30201194	Hes1	22
chr16	91300931	91316898	Olig1	23
chr8	10943368	10992655	Irs2	24
chr7	152278676	152309068	Ccnd1	25
chr5	125614247	125647398	Ncor2	26
chr4	8835933	8882787	Chd7	27
chr11	35731471	35749952	Wwc1	28
chr15	61908412	61941902	Myc	29
chr10	59352662	59379935	Dnajb12	30
chr2	116797035	116807250	D330050G23Rik	31
chr6	39465192	39499349	Dennd2a	32
chr13	60011111	60050908	Zcchc6	33
chr1	88386927	88408896	Ptma	34
chr18	69719430	69759581	Tcf4	35
chr18	5927357	5950006	Zeb1/Arhgap12	36
chr10	39364940	39397688	Rev3l	37
chr10	21918559	21946849	Raet1c	38
chr4	81370973	81411114	Mpdz	39
chr15	78454098	78469246	Tst	40
chr3	87784757	87795603	Nes	41
chr6	52007076	52047390	Skap2	42
chr5	135728452	135752267	Fzd9	43
chr9	65043416	65044793	Parp16	44
chr17	10069869	10104111	Qk	45
chr4	124269570	124304820	Pou3f1	46
chr4	62242178	62291719	Rgs3	47
chr5	125681568	125710489	Ncor2	48
chr14	26189541	26199113	Zmiz1	49
chr17	87758354	87796822	Calm2	50
chr1	87926058	87953470	2810459M11Rik	51
chr4	132493828	132520993	Ahdc1	52
chr15	8605991	8618845	Slc1a3	53
chr18	36007350	36036753	Cxxc5	54
chr10	120933693	120946080	Rassf3	55
chr18	39047596	39090160	Fgf1	56
chr11	60021829	60025403	Mir6922	57
chr1	135955951	135964641	Btg2	58
chr9	13471785	13509560	Mtmr2	59
chr4	132611406	132651895	Ahdc1	60
chr11	20813571	20824144	Lgalsl	61
chr15	103128946	103138067	Copz1	62
chr7	87333320	87344538	Sema4b	63
chr15	59002559	59031582	Mtss1	64
chr3	37655131	37671882	Gm5148	65
chr11	8342981	8369889	Tns3	66
chr1	133357056	133391143	Srgap2	67
chr15	25681462	25705837	Myo10	68
chr3	21695113	21736102	Tbl1xr1	69
chr15	78498507	78523432	Cdc42ep1	70
chr19	5809441	5828444	Malat1	71
chr3	105323427	105351914	Kcnd3	72
chr6	125415483	125442499	Cd9	73
chr11	88336246	88372755	Msi2	74
chr18	5805306	5823613	Zeb1/Arhgap12	151
chr11	119566837	119577289	Rptor	152
chr2	170110374	170135828	Bcas1	153
chr8	129347420	129364215	Tomm20	154
chr11	88387914	88408876	Msi2	155
chr11	77963873	77970757	Traf4	156
chr8	93385734	93401737	Chd9	157
chr9	64438337	64461582	Rab11a	158
chr10	122320493	122332643	Mirlet7i	159
chr1	39977143	40012275	Map4k4	160
chr18	81244290	81251873	Sall3	161
chr11	16716012	16745711	Egfr	162
chr5	114319446	114334043	Coro1c	163
chr11	97665648	97688033	Lasp1	164
chr1	90551466	90565911	Arl4c	165
chr14	100338322	100345888	Klf12	166
chr11	49574375	49580983	Gfpt2	167
chr2	30284316	30304337	Ppp2r4	168
chr17	87320926	87338653	Rhoq	169
chr5	24101501	24118490	Smarcd3	170
chr11	103001805	103006975	Hexim2	171
chr12	16625392	16671585	Lpin1	172
chr2	26386212	26407938	Notch1	173
chr7	143236009	143243275	Mki67	174
chr6	53984690	54024937	Chn2	175
chr5	121012463	121019436	Tpcn1	176
chr4	106508605	106523148	Ssbp3	177
chr6	114214004	114248479	Slc6a1	178
chr13	64001745	64022125	Ercc6l2	179
chr8	13613545	13626651	Rasa3	180
chr13	112074423	112085101	Gpbp1	181
chr15	77188962	77199095	Rbfox2	182
chr1	34719932	34737116	AA619741	183
chr11	76458691	76462174	Gosr1	184
chr7	86501160	86513711	Rlbp1	185
chr11	45875589	45883899	Adam19	186
chr15	76290520	76297397	Bop1	187
chr9	49784575	49814700	Ncam1	188
chr17	25791100	25818928	Sox8	189
chr2	28376738	28396020	Ralgds	190
chr8	122845943	122864224	Gse1	191
chr18	38900865	38928521	Spry4	192
chr14	106019938	106032695	Spry2	193
chr11	107186922	107221669	Pitpnc1	194
chr5	65091139	65103112	Klf3	195
chr11	100247060	100255613	Jup	196
chr19	21797403	21818695	Tmem2	197
chr6	143990591	144008324	Sox5	198
chr17	29847980	29875911	Ccdc167	199
chr16	35587713	35625492	Sema5b	200
chr2	166495405	166519996	Mir6364	201
chr12	87982084	88002414	Vash1	202
chr6	72460631	72471111	Mat2a	203
chr5	141011853	141033240	Chst12	204
chr4	104840775	104857349	Plpp3	205
chr14	79581611	79609988	Zfp957	206
chr6	136479697	136505087	Atf7ip	207
chr1	171388527	171417062	Nuf2	208
chr15	38397915	38404625	Mir6951	209
chr13	112716978	112735932	Gm15326	210
chr13	34238485	34247378	Tubb2b	211
chr11	117488035	117490540	2900041M22Rik	212
chr12	74509264	74518368	D830013O20Rik	213
chr16	91172132	91196351	Olig2	214
chr15	58871421	58886972	Mtss1	215
chr13	26659373	26672488	Hdgfl1	216
chr16	22541970	22549876	Etv5	217
chr11	68849115	68855239	Aurkb	218
chr8	97870915	97902135	Mmp15	219
chr3	122163561	122187231	Bcar3	220
chr7	130172936	130191822	Rbbp6	221
chr11	117383807	117396713	Tnrc6c	222
chr15	27706763	27728254	Fam105a	223
chr1	74433011	74444840	Ctdsp1	224
chr4	140469593	140486963	Padi2	301
chr1	135991387	136001493	Chit1	302
chr4	135648973	135663687	Id3	303
chr18	5970822	5983663	Zeb1/Arhgap12	304
chr1	157075334	157089623	Stx6	305
chr9	116121027	116141991	Tgfbr2	306
chr8	124445860	124461019	Car5a	307
chr1	161591032	161618515	Tnr	308
chr17	48562894	48577727	Apobec2	309
chr6	29384145	29386998	Ccdc136	310
chr18	81346834	81353180	Sall3	311
chr5	113560943	113582451	2900026A02Rik	312
chr4	8936157	8947227	Clvs1	313
chr14	65736523	65759961	Extl3	314
chr12	60140818	60141650	Trappc6b	315
chr9	35137770	35150197	4933422A05Rik	316
chr4	134852801	134859097	Clic4	317
chr12	81062489	81073185	9430078K24Rik	318
chr7	53762869	53772311	Kcnc1	319
chr11	62779165	62781844	Cdrt4	320
chr11	77598914	77607555	Myo18a	321
chr9	107512784	107518245	Sema3b	322
chr18	35832971	35839090	Spata24	323
chr14	27978449	27991717	Arhgef3	324
chr12	83376347	83390351	Sipa1l1	325
chr14	101146771	101153365	Prr30	326
chr11	6005975	6023534	Camk2b	327
chr15	73262364	73292509	Ptk2	328
chr8	26573553	26578214	Fgfr1	329
chr12	70597916	70604630	5830428M24Rik	330
chr8	37205364	37223732	D8Ertd82e	331
chr11	101337567	101341639	Rnd2	332
chr12	80779237	80798075	9430078K24Rik	333
chr13	102568876	102584348	Pik3r1	334
chr14	52623496	52632429	Zfp219	335
chr16	37682996	37686232	Ndufb4	336
chr5	135933897	135944363	Pom121	337
chr15	30368065	30388874	Ctnnd2	338
chr6	36786558	36811119	Ptn	339
chr9	72475905	72489493	Nedd4	340
chr1	9543368	9561036	Adhfe1	341
chr4	106450160	106460814	Acot11	342
chr2	38678183	38681875	Mir181a-2	343
chr11	102068207	102081095	Mir8101	344
chr12	55076846	55087701	Npas3	345
chr18	65721577	65727797	Zfp532	346
chr10	59656782	59673902	Chst3	347
chr12	107197439	107214545	Vrk1	348
chr4	151444093	151454584	Plekhg5	349
chr8	87251095	87268911	Lyl1	350
chr12	73531504	73541049	Lrrc9	351
chr7	71095640	71116327	Klf13	352
chr11	5898824	5904126	Ykt6	353
chr17	29477951	29494001	Mtch1	354
chr6	114661667	114672322	Atg7	355
chr11	20376341	20378512	Sertad2	356
chr3	109101057	109113595	Vav3	357
chr17	35886598	35892597	Ier3	358
chr1	64331347	64353823	Klf7	359
chr11	54829010	54838136	Anxa6	360
chr7	87294161	87317692	Sema4b	361
chr12	83152113	83159247	Sipa1l1	362
chr17	35913022	35921493	Ier3	363
chr4	124122620	124130349	Pou3f1	364
chr15	85384277	85400995	Wnt7b	365
chr11	117268050	117271978	Sept9	366
chr4	82327709	82351755	Nfib/Zdhhc21	367
chr15	78231658	78247788	Mpst	368
chr2	165238990	165247308	Ocstamp	369
chr10	58061925	58068273	Edar	370
chr10	98659891	98668241	Dusp6	371
chr1	133312343	133335917	Srgap2	372
chr1	133274183	133283339	Srgap2	373
chr11	69839610	69843883	Dlg4	374

Table Supplemental 2 | *continuation*

chr10	39252947	39275332	Fyn	75
chr9	110186503	110201403	Elp6	76
chr6	36702984	36750809	Ptn	77
chr13	48640358	48663417	Mirlet7a-1	78
chr2	30501723	30516698	Ier5l	79
chr11	31727738	31748918	Cpeb4	80
chr12	101526618	101566851	Calm1	81
chr15	78661793	78667613	Cdc42ep1	82
chr1	137050543	137092644	Gpr37l1	83
chr13	99866942	99873487	Tnpo1	84
chr3	154353144	154365670	Tyw3	85
chr13	98516438	98548119	Enc1	86
chr15	79153194	79166581	Pla2g6	87
chr16	22514952	22528945	Etv5	88
chr2	30591868	30616196	Ntmt1	89
chr11	69221609	69233544	Kdm6b	90
chr14	32472471	32484205	Btd	91
chr1	184011739	184049049	Srp9	92
chr6	94579953	94606822	Lrig1	93
chr2	13305156	13337097	Vim	94
chr1	39789060	39821375	Map4k4	95
chr10	82579250	82609544	Chst11	96
chr15	58806194	58810009	Ndufb9	97
chr2	91793586	91827535	Dgkz	98
chr6	90477513	90492911	Klf15	99
chr1	139794284	139805162	Mir181a-1	100
chr7	86664292	86680964	Polg/Mir9-3	101
chr14	55668547	55676284	Ngdn	102
chr17	15125416	15130870	9030025P20Rik	103
chr19	44355124	44360183	Scd2	104
chr9	34791299	34805746	St3gal4	105
chr3	105365912	105396277	Kcnd3	106
chr9	121266324	121282187	Trak1	107
chr9	56714927	56731970	Snx33	108
chr3	34338789	34366158	Sox2	109
chr9	44286307	44304119	Bcl9l	110
chr17	15155439	15162263	Gm3435	111
chr9	62356315	62381012	Coro2b	112
chr13	25817776	25847367	Gm11351	113
chr19	44385015	44390289	Scd2	114
chr2	167569381	167603828	Cebpb	115
chr18	38944592	38972979	Fgf1	116
chr7	104572058	104597317	Ndufc2	117
chr11	31611849	31635912	Bod1	118
chr13	56682238	56703022	Smad5	119
chr9	57266721	57282574	Ppcdc	120
chr17	29186499	29200742	Srsf3/Cdkn1a	121
chr1	137540473	137585381	Nav1	122
chr1	186264104	186280978	Hlx	123
chr14	106127780	106142495	Spry2	124
chr15	59445607	59469858	Trib1	125
chr18	65674684	65691615	Zfp532	126
chr1	88329973	88340197	Ncl	127
chr14	26125981	26141394	Zmiz1	128
chr8	13528802	13551745	Rasa3	129
chr11	119376092	119385608	Endov	130
chr11	107271012	107281765	Pitpnc1	131
chr7	119960982	119986425	Tead1	132
chr8	129189159	129193320	Irf2bp2	133
chr17	9212696	9229211	6530411M01Rik	134
chr15	99162347	99175420	Fmnl3	135
chr15	79021229	79033974	Pick1	136
chr19	22978017	23016587	Klf9	137
chr19	44320244	44324739	Scd2	138
chr11	98642944	98651171	Msl1	139
chr11	93994779	94007015	Tob1	140
chr9	77738396	77753515	Sez6	141
chr10	78335960	78354451	Eef1a1	142
chr7	90642678	90651315	Tmpo	143
chr1	124595576	124619488	Xylt1	144
chr7	93099876	93110430	Ramp1	145
chr6	53853925	53863806	Sergef	146
chr15	114823567	114844354	Vgll4	147
chr5	30200074	30218354	Ctnnd2	148
chr4	135005963	135032236	Clip2	149
	130352907	130372208	Sdc3	150
chr8	122892469	122907958	Gse1	225
chr1	36645343	36665085	Sema4c	226
chrX	162743290	162768051	Gpm6b	227
chr11	98607594	98613816	Thra	228
chr3	121388273	121411290	4930432M17Rik	229
chr14	100854257	100863838	Klf12	230
chr11	107735817	107738929	Cacng5	231
chr15	102485657	102490508	Atp5g2	232
chr13	43624890	43636456	Mcur1	233
chr17	45717443	45723481	Slc29a1	234
chr4	45479393	45503906	Shb	235
chr17	73599158	73608387	Lclat1	236
chr11	88464669	88470954	0610039H22Rik	237
chr15	85287905	85306749	7530416G11Rik	238
chr9	77757413	77760135	Elovl5	239
chr3	126716030	126725976	Ank2	240
chr2	72832158	72844442	Sp3	241
chr12	80832265	80855286	Zfp36l1	242
chr6	86518508	86536863	1600020E01Rik	243
chr4	41562329	41579764	Enho	244
chr8	14989534	15006822	Kbtbd11	245
chr1	74149984	74166326	Tns1	246
chr6	85244303	85271543	Sfxn5	247
chr13	112949010	112960151	Ankrd55	248
chr11	88267894	88289697	Msi2	249
chr2	93720525	93752955	Accsl	250
chr19	36770308	36781653	1500017E21Rik	251
chr8	74804986	74835449	Klf2	252
chr6	136414498	136420436	Mir125b-1/311003	253
chr9	41683624	41699969	Pkm2	254
chr9	59538799	59543968	Gramd2	255
chr11	118058189	118064612	Cyth1	256
chr15	84861897	84873879	Fam118a	257
chr17	84480290	84512248	4933433H22Rik	258
chr1	182307087	182340046	Gm5069	259
chr11	102194399	102201283	Ubtf	260
chr5	135960939	135970246	Hip1	261
chr9	56344605	56365585	Hmg20a	262
chr4	155149195	155159718	Mrpl2	263
chr19	6443147	6466666	Nrxn2	264
chr6	112807023	112841300	Srgap3	265
chr9	107603884	107627578	Sema3f	266
chr2	166167580	166186470	Gm11468	267
chr2	152161301	152177011	Trib3	268
chr5	124443078	124470006	Vps37b	269
chr1	64234470	64249750	Klf7	270
chr10	92737558	92739987	Elk3	271
chr19	46592847	46607119	Trim8	272
chr18	38538643	38546213	Ndfip1	273
chr17	84652297	84675091	Zfp36l2	274
chr9	31072452	31091003	Gm7244	275
chr2	26340749	26345305	Notch1	276
chr6	53397179	53422217	Creb5	277
chr12	73032419	73055731	Daam1	278
chr2	136503440	136513657	Snap25	279
chr19	7024721	7028144	Bad	280
chr8	113185121	113213595	Mtss1l	281
chr7	54240533	54257788	Spty2d1	282
chr8	122939645	122951964	Gse1	283
chr12	55810864	55812120	Eapp	284
chr14	26044478	26049525	4930572O13Rik	285
chr15	12290142	12301766	Golph3	286
chr1	72783015	72800660	Rpl37a	287
chr11	94019856	94023396	Tob1	288
chr1	42584731	42599926	Pou3f3	289
chr8	26154016	26182438	Adam9	290
chr11	83084573	83085029	Pex12	291
chr3	35038826	35053871	Mir6378	292
chr15	95656244	95661884	Ano6	293
chr11	32134474	32145299	Mpg	294
chr3	101569935	101610913	Mab21l3	295
chr17	9913544	9922915	Qk	296
chr3	35231237	35242401	Gm6639	297
chr3	88007793	88012001	Mir3093	298
chr9	89663983	89686399	Tmed3	299
chr6	22740908	22748985	Ptprz1	300
chr11	86405505	86410125	Vmp1	375
chr11	75021681	75028384	Rtn4rl1	376
chr3	88558867	88567366	Syt11	377
chr11	107589197	107605134	Cacng1	378
chr6	28791218	28796815	Lrrc4	379
chr9	80002525	80027585	Myo6	380
chr2	156388332	156411882	Aar2	381
chr13	55428521	55440531	Mxd3	382
chr4	135688796	135693829	Id3	383
chr11	51898872	51901371	Ppp2ca	384
chr8	114239316	114250867	Bcar1	385
chr16	43571139	43579759	Mir568	386
chr7	123410058	123427461	Plekha7	387
chrX	100745666	100760134	Gm9159	388
chr11	20756904	20760894	Lgalsl	389
chr2	165987701	166007365	Sulf2	390
chr13	60151109	60181610	A530065N20Rik	391
chr17	31692614	31718635	Pknox1	392
chr2	30547262	30551415	Ntmt1	393
chr12	55839189	55840100	Eapp	394
chr11	116189998	116194923	Foxj1	395
chr9	106278968	106309549	Dusp7	396
chr17	10525646	10531094	1700110C19Rik	397
chr12	83121484	83128456	Sipa1l1	398
chr15	95735967	95763426	Ano6	399
chr11	53305197	53322737	Sept8	400
chr8	94650727	94659962	Crnde	401
chr2	30961546	30976237	Fnbp1	402
chr12	81927296	81965098	Susd6	403
chr2	131271425	131289448	Smox	404
chr15	99267484	99280734	Nckap5l	405
chr4	62319913	62332185	Rgs3	406
chr5	43745039	43752089	Cpeb2	407
chr3	101504274	101517745	Atp1a1	408
chr14	51707262	51714893	Ang	409
chr11	54565119	54570306	Cdc42se2	410
chr17	26639262	26642523	Dusp1	411
chr4	8299838	8310262	Car8	412
chr6	91123339	91153252	Fbln2	413
chr8	13449303	13468670	Tmem255b	414
chr1	128576891	128584916	Nckap5	415
chr1	139822383	139828925	Mir181a-1	416
chr3	85103159	85115027	1700036G14Rik	417
chr1	196565187	196575574	Plxna2	418
chr15	80502731	80512260	Tnrc6b	419
chr9	43878618	43887871	Usp2	420
chr17	31671150	31679500	Ndufv3	421
chr6	113280566	113288654	Ogg1	422
chr12	55132907	55138973	Npas3	423
chr5	34031087	34049607	Fgfr3	424
chr5	91530384	91546502	Areg	425
chr8	91560010	91567809	Sall1	426
chr14	25756779	25761407	4930572O13Rik	427
chr10	6224636	6242419	Mthfd1l	428
chr17	45885742	45907544	1600014C23Rik	429
chr2	72919925	72927796	Sp3	430
chr16	43502316	43511338	Gm15713	431
chr1	173032153	173042525	Mpz	432
chr11	113043188	113054240	2610035D17Rik	433
chr5	25114042	25123783	1700096K18Rik	434
chr6	142815022	142828647	Gm7457	435
chr4	124385034	124388644	Sf3a3	436
chr11	4090272	4104237	Tbc1d10a	437
chr12	16521050	16540843	Lpin1	438
chr5	54486695	54495960	Stim2	439
chr5	125743298	125751794	Scarb1	440
chr15	99250086	99252820	Tmbim6	441
chr7	141436315	141455911	Adam12	442
chr5	138197068	138202120	Tsc22d4	443
chr16	11011201	11022405	Gm4262	444
chr5	22669061	22674497	5031425E22Rik	445

Table Supplemental 3 | GO terms of SE, Broad and SE+Broad genes

Broad H3K4me3 promoter

GO terms	Benjamini	Enrichment
Transcriptional regulation	8,00E-47	41.4
mRNA splicing	1.7E-5	5.0

TFs Broad

GO terms	Benjamini	Enrichment
Ventricular septum morphogenesis	8.2E-6	3.7
Regulation of neurogenesis	4.3E-5	3.3
Oligodendrocyte differentiation	7.2E-5	3.9

SE

GO terms	Benjamini	Enrichment
Transcriptional regulation	5.5E-8	7.1
Semaphorin	0.2; not significant	2.1

TFs SE

GO terms	Benjamini	Enrichment
Regulation of neurogenesis	0.01	2.6

SE+Broad

GO terms	Benjamini	Enrichment
Transcriptional regulation	4.4E-10	9.0

TFs SE+Broad

GO terms	Benjamini	Enrichment
Regulation of neurogenesis	0.002	3.1

Broad-SE

GO terms	Benjamini	Enrichment
Transcriptional regulation	9,00E-36	32
mRNA splicing	1.2E-5	5.1

TFs Broad-SE

GO terms	Benjamini	Enrichment
Fat cell differentiation	0.11; not significant	1.8

SE-Broad

GO terms	Benjamini	Enrichment
Semaphorin	0.35; not significant	2.1

Table Supplemental 4 | Genes with SE(s) and broad H3K4me3 promoters in NSCs

1810026B05Rik	Kdm6b	Rasa3
5031425E22Rik	Klf12	Rassf3
Adam12	Klf13	Rbbp6
Adhfe1	Klf3	Rev3l
Ahdc1	Klf7	Rhoq
Ank2	Klf9	Sall1
Arl4c	Lrig1	Sall3
Aurkb	Malat1	Scd2
Bahcc1	Map4k4	Slc1a3
Bcl9l	Mir125b-1	Snx33
Calm2	Mir181a-1	Sox2
Camk2b	Mir3093	Sox8
Ccnd1	Mir9-3	Spry2
Cdkn1a	Mirlet7c	Spry4
Chd7	Mki67	Srgap2
Chst11	Mmp15	Srsf3
Cpeb4	Msi2	Stim2
Ctnnd2	Msl1	Sulf2
Cxxc5	Mtss1	Tbl1xr1
D8Ertd82e	Mtss1l	Tcf4
Ddr1	Myc	Thra
Dennd2a	Myh9	Tnpo1
Dlg4	Myo18a	Tob1
Dusp1	Nav1	Traf4
Dusp6	Ncor2	Trib1
Dusp7	Nedd4	Trim8
Eef1a1	Nes	Tuba1a
Egfr	Nfia	Tubb2b
Enc1	Nfib	Ubtf
Etv5	Notch1	Vash1
Fam105a	Npas3	Vav3
Fgf1	Olig1	Vgll4
Fgfr3	Olig2	Vim
Fyn	Pcdhgc4	Vps37b
Gm5069	Pik3r1	Wwc1
Golph3	Plxna2	Xylt1
Gpm6b	Polg	Zcchc24
Hes1	Pom121	Zeb1
Hip1	Pou3f1	Zfp36l1
Hlx	Pou3f3	Zfp36l2
Ier5l	Ptk2	Zfp532
Irf2bp2	Ptma	Zmiz1
Irs2	Ptn	Qk
Kcnc1	Ptprz1	

Table Supplemental 5 | Datasets used in the study

ChIP-seq

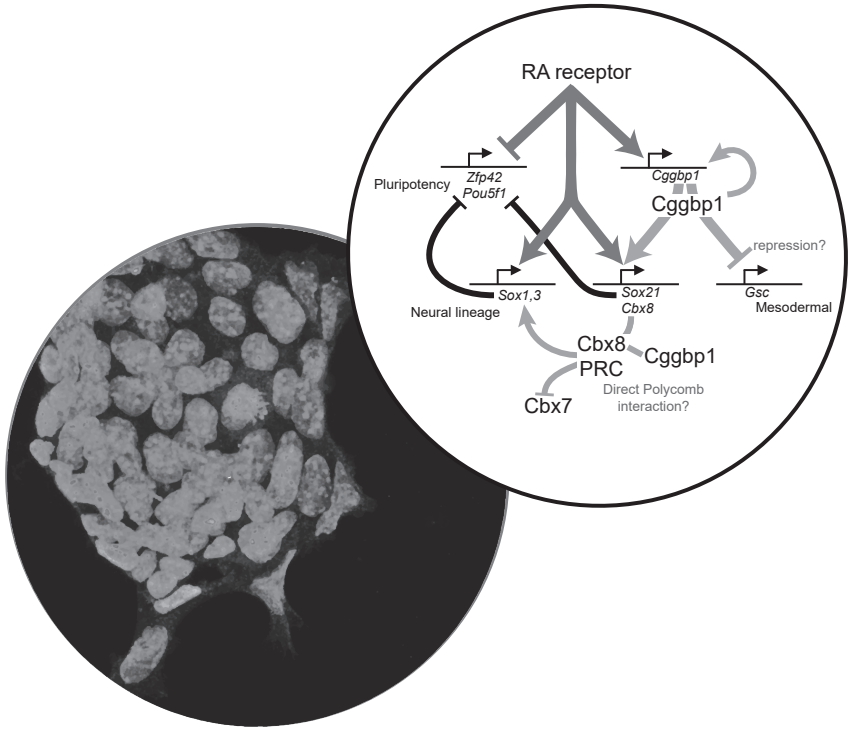
Med1	GSM2928425
Jmjd1c	GSM2928426
Carm1	GSM2928427
IgG	GSM2928428
mNSCs Input	GSM1187180
Tcf4	GSM1820990
Sox2	ERR414096
NFI	ERR414099
Olig2	GSM1820994
Ints11	ERR1173522
RNA pol II	ERR1173526
Ep300	ERR216112
H3K27ac	ERR216108
H3K4me3	SRR006888_SRR006889
Cbx8	GSM2393587
Ctcf	GSM883647

RNA-seq

wild-type NSCs	GSM1861892, GSM1861893, GSM1861894
scrambled shRNA NSCs	GSM1861886, GSM1861887, GSM1861888
Tcf4 knock down NSCs	GSM1861889, GSM1861890, GSM1861891

Chapter 3

Cgg-binding protein 1 regulates neural induction and neural stem cell homeostasis



- Work in progress -

Cgg-binding protein 1 regulates neural induction and neural stem cell homeostasis

Marti Quevedo[1*], Lleches-Padilla A[1], Gordaliza I[1], Mike R. Dekker[1], Dick H.W. Dekkers[2], Zeliha Ozgür[3], Wilfred F.J. van IJcken[3], Jeroen Demmers[2], Raymond A. Poot[1]

[1]Department of Cell Biology, Erasmus MC, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands
[2]Center for Proteomics, Erasmus MC, The Netherlands
[3]Center for Biomics, Erasmus MC, The Netherlands
[*]Corresponding author

ABSTRACT

Neural induction represents the earliest step in the determination of ectodermal cell fates. During early embryogenesis a complex transcriptional network regulates the balance between embryonic stem cells pluripotency versus differentiation towards neural precursors. Among them, Polycomb group proteins regulate the state of many developmental genes by keeping them in a poised state.

In our study we have identified Cggbp1, a small zing finger protein, as a regulator of neural induction and neural progenitor homeostasis. Cggbp1 binds to early neural fate gene promoters together with the Mediator complex and chromatin remodelers to activate transcription. In addition, we find an association between Cggbp1 and neural poised genes in embryonic stem cells and we explore its relation to Polycomb mediated derepression.

INTRODUCTION

The central nervous system originates from the dorsal part of the ectoderm layer. Around embryonic day 7.5 in the mouse a part of the ectoderm thickens and flattens forming the neural plate, which will wrap in on itself to give rise to the neural tube. This developing neuroepithelium will in time generate most of neurons and glia in the body[1].

The combination of morphogenetic signals expressed within the neural ectoderm and among surrounding tissues integrate through several signaling cascades to activate cell-intrinsic factors that cooperate to stabilize neural fate[2]. During this process the balance between pluripotency and differentiation is tightly regulated at the transcriptional level by factors such as the Polycomb group (PcG) proteins. PcG proteins are powerful regulators of cell fate decisions as they dictate the state of many developmental genes by keeping them in a poised state. The selective derepression of poised genes depends on the combination of cell extrinsic morphogens and cell intrinsic factors to which cells are exposed. Recent studies suggest that Polycomb repressive complexes are recruited through a range of different mechanisms, from non-coding RNA, to CpG islands and transcription factors[3].

Cggbp1 is a 167 amino acid long 20 kDa zinc finger protein originally identified in a screen for factors associated with CGG tandem repeats[4]. However, the reported function of Cggbp1 in transcription is ambiguous. Initially, its association with CGG repeats attracted some attention as a regulator of the fragile X mental retardation 1 gene (FMR1) due to its pathology being caused by an expansion of the CGG tripleat[4]. On top of that, several studies have pointed Cggbp1 as a transcription repressor. However, one study postulates a direct transcription effect at specific heat-shock promoters[5], while another study points at a genome-wide effects on Alu/SINE repeats[6].

Here we identify Cggbp1 as an activating transcription factor that regulates the expression of neural specific genes. Cggbp1 binds to GCC repeats in gene promoters of neural stem cells (NSCs) together with the Mediator complex and chromatin remodelers and activates their transcription. Cggbp1 levels are essential not only for the correct expression of key NSCs genes but also for NSCs survival. Furthermore, we identified a Cggbp1 role at early stages of embryonic stem cells (ESCs) neural induction where it regulates transient neural inducers such as *Sox21* in response to retinoic acid (RA). Finally, we associated Cggbp1 early neural fate function to Polycomb poised gene regulation and we preliminarily explore their cooperation.

RESULTS

Cggbp1 is expressed in NSCs and interacts with the Mediator complex

We have identified Cggbp1 as a DNA-independent protein-protein interactor of the Mediator complex in Mediator subunit 15 (Med15)-FLAG purifications from NSCs. Cggbp1 was consistently present in all 4 Mediator purifications, irrespective of the addition of Benzonase or Ethidium Bromide (Figure 1A, Chapter 2 of this thesis). Subsequent Cggbp1-Flag purifications confirmed its interaction with Mediator in NSCs (Figure 1B). In addition, immunohistochemistry staining of Cggbp1 in NSCs

showed endogenous nuclear localization (Figure 1C). Cggbp1 interaction with Mediator, a central core piece of transcription activation[7], and the published interaction with NFIX[4], a transcription factor recently identified as a crucial NSCs regulator[8], lead us hypothesize a wider role in NSC transcription for Cggbp1.

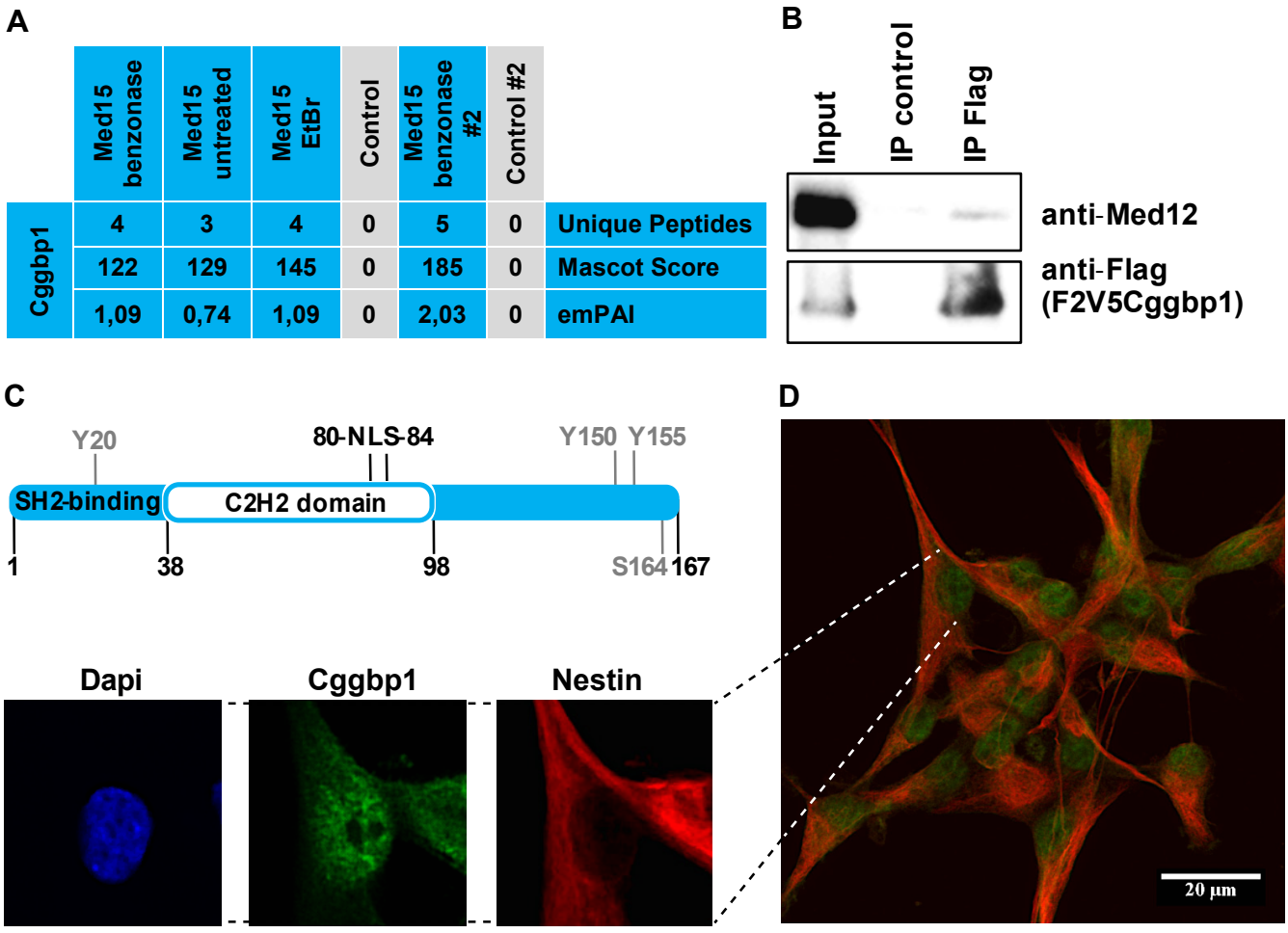


Figure 1 | Cggbp1 is expressed in NSCs and interacts with the Mediator complex
(A) Mass spectrometry counts for Cggbp1 in Med15-Flag purifications in NSCs. (B) Flag-Cggbp1 or control purifications in NSCs analyzed by western blot. (C) Protein structure of Cggbp1 and reportedphosphorylations. (D) Immunocytochemistry with Cggbp1 and Nestin antibodies on NSCs showing nuclear localization of endogenously expressed Cggbp1.

Cggbp1 positively regulates essential neural genes

In order to explore Cggbp1 function in transcriptional regulation, we determined Cggbp1 genomic occupancy by chromatin immunoprecipitation followed by high-throughput sequencing of the bound DNA (ChIP-seq) in NSCs. We identified 1855 specific Cggbp1 binding sites after filtering with a published NSC input background signal. Cggbp1 has a strong preference for promoter sequences (Figures 2A-C) including CpG islands (Figure 2A,D). Analysis of Cggbp1 DNA binding sites retrieved sequences with a highly significant accumulation of 3 times the GCC triplet (Figure 2E).

Cggbp1 bound genes identified by gene annotation of Cggbp1 promoter peaks showed slightly increased expression compared to all NSC expressed genes suggesting a role of Cggbp1 beyond transcriptional repression (Figure Supl. 1A). The amount of Cggbp1 signal in each promoter peak did not correlate with the expression of the assigned gene (Figure Supl. 1B).

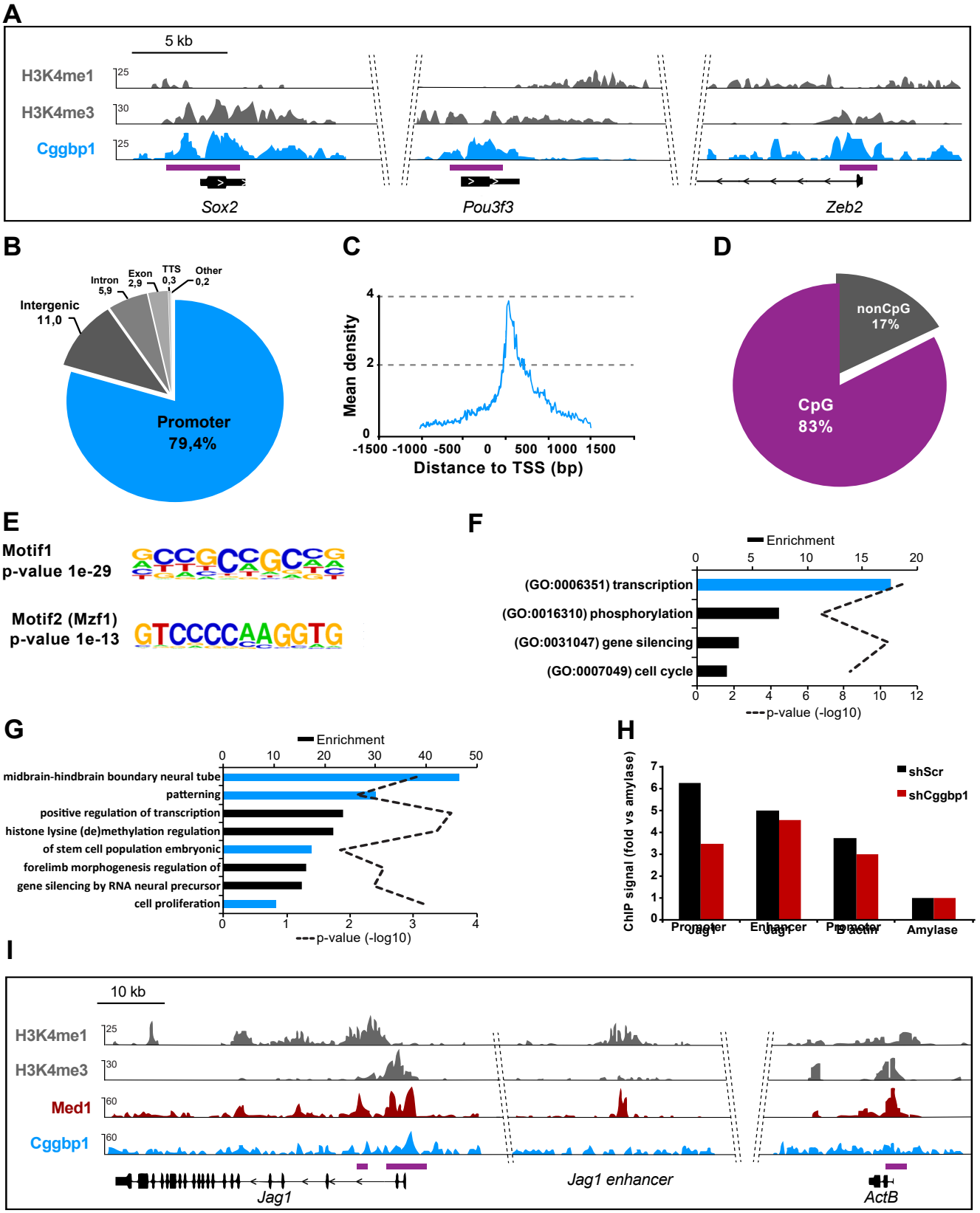


Figure 2 (at previous page)| Cggbp1 activates promoters of neural induction genes
(A) Chip-seq tracks of Cggbp1 and histone modifications H3K4me3 and H3K4me1 at different neural genes loci. Purple bars indicate CpG islands. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated. (B) Distribution of Cggbp1 genomic binding sites. (C) ChIP-seq meta-profile plots around promoters. Mean Cggbp1 ChIP-seq density (y-axis) and distance to TSS (x-axis) are shown. (D) Quantification of the genomic overlap of Cggbp1 peaks to CpG islands. (E)Significant enriched motifs in Cggbp1 peaks. Significant values as reported by HOMER are listed. (F) DAVID cluster analysis on genes bound by Cggbp1. GO terms related to transcription are highlighted. (G) GO analysis on genes from the transcription category in F. Neural GO terms are highlighted. (H) Normalized Med1 ChIP signal values at different loci in NSCs after 40h transfection with scramble or shCggbp1. (I) Chip-seq tracks of Cggbp1, Med1 and histone modifications H3K4me3 and H3K4me1 at different neural genes loci. Purple bars indicate CpG islands. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated.

Despite being associated to the Mediator complex, which in a previous study we mapped to almost all active genes in NSCs (see Chapter 2 of this thesis), we found Cggbp1 to bind to a very small subset of active genes. In fact, GO term analysis of Cggbp1 bound genes revealed a strong enrichment for genes involved in transcription followed by genes involved in phosphorylation, gene silencing by RNA and cell cycle (Figure 2F). Interestingly, Cggbp1 knock down effects have been reported to have substantial effects on transcription regulators in cancer cells[9]. In the same study, effects in cell cycle were also observed, although they were not associated with direct Cggbp1 transcriptional regulation of cell cycle genes, as our results would suggest.

A closer examination on the transcription gene cluster revealed many transcription factors and chromatin remodelers involved in early neural development such as boundary establishment and neural tube patterning (Figure 2G). Moreover, important genes for NSC regulation were also found to be bound by Cggbp1 (Figure 2A and 2G). Next, we investigated the capacity of Cggbp1 to regulate some of its target genes. Depletion of Cggbp1 by shRNA knock down lead to a downregulation of most of the targets tested (Figure 3A and Figure Supl. 2A-B). Importantly, *Fmr1* expression, which recently was stated not to be affected by Cggbp1 reduced levels[10], was found to be also affected in NSCs.

In addition, we examined the capacity of Cggbp1 to recruit activation machinery to one of its targets. After Cggbp1 depletion, we observed a Mediator complex recruitment defect at *Jag1* promoter, a gene target of Cggbp1, but no defects on *Jag1* enhancer neither to *ActB* promoter, a non Cggbp1-bound gene (Figure 2H,F).

As hinted by our genome wide GO analysis, depletion of Cggbp1 affected the homeostasis of NSC and by 72 hours most of Cggbp1 depleted NSCs had died (Figure 3A-C). Changes in proliferation and viability could be explained described function of Cggbp1 in cell cycle regulation[9]. However, FACS cell cycle analysis of misregulated Cggbp1 NSCs showed no significant changes compared to their transfected controls (Figure Supl. 2D). Overexpression of Cggbp1 showed no major effects on its targets except the downregulation of *Sox11* and *Pou3f3* (also known as *Brn1*) neither significantly

affecting NSC viability (Figure 3C-D, Figure Supl 2C). Taken together, these results suggest that Cggbp1 mainly acts as a transcriptional activator that regulates genes essential for NSCs survival.

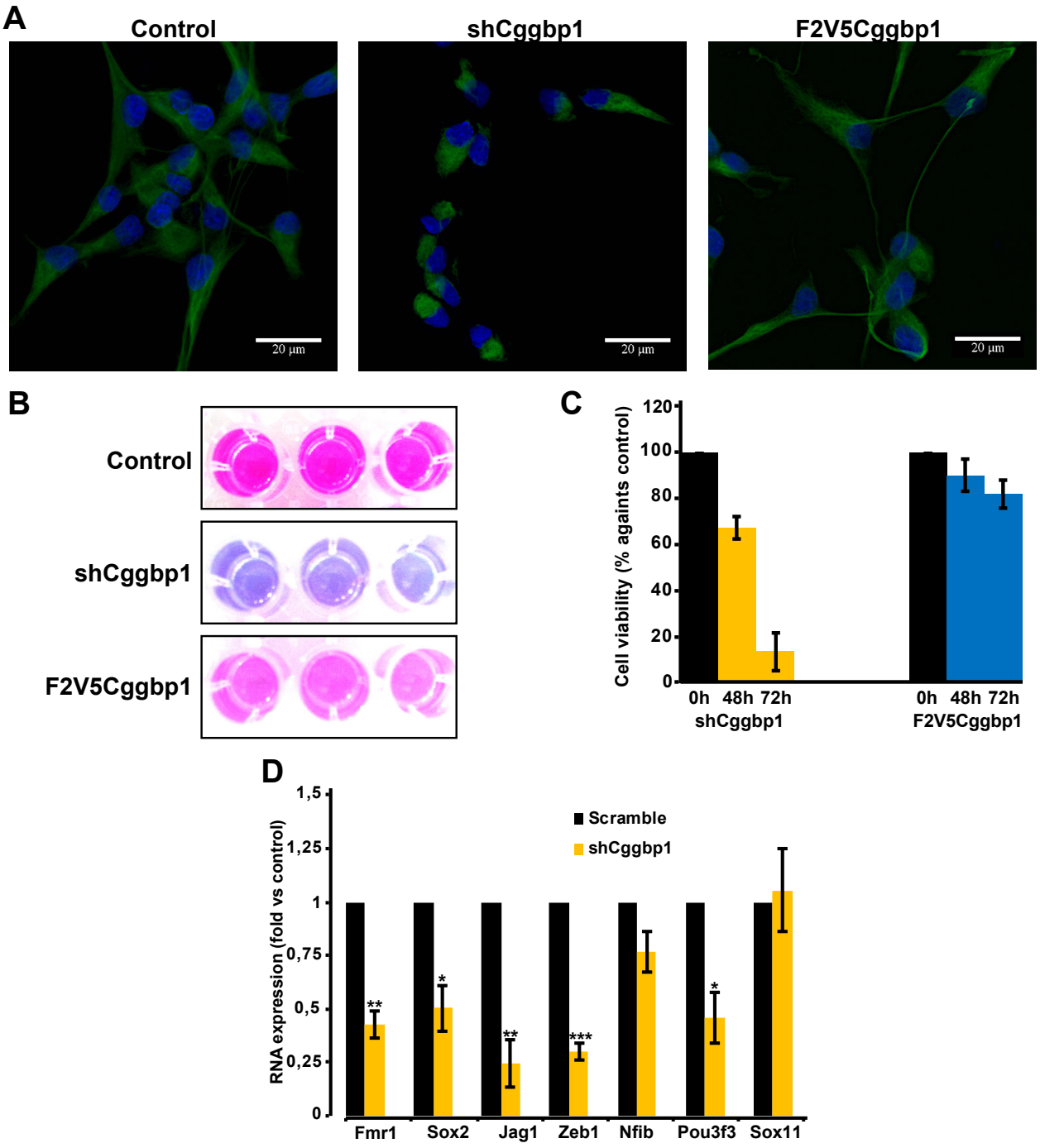


Figure 3 | Cggbp1 regulates genes essential for NSCs survival

(A) Immunocytochemistry with Nestin antibodies on NSCs transfected with shCggbp1, F2V5Cggbp1. Control is representative of scramble and F2V5empty. (B) Plate wells incubated with alamarBlue showing the effects of Cggbp1 misregulation in NSCs . (C) Fluorescence quantifications by alamarBlue incubation (cell viability) of transfected NSCs at different time points. Error bars represent SEM, *p < 0.05, unpaired Student's t test, n=2. (D) qPCR analysis on NSCs transfected with scramble or Cggbp1-targeting shRNA. Error bars represent SEM, *p < 0.05, **p < 0.01, ***p<0.001, unpaired Student's t test, n=3.

Cggbp1 regulates early lineage specification and induces loss of pluripotency

Due to its association with genes involved in early neural specification events, its putative regulation of histone (de)methylases (Figure 2G) together with its increased expression during brain and testis development[11], we hypothesized that Cggbp1 may play a function in early neural development.

Cggbp1-bound promoters in NSCs partially overlap with histone H3 lysine K27 trimethylation (H3K27me3)-marked promoters in ESCs (Figure 4A and 4B). Indeed, genes from the resulting overlap are enriched for neurodevelopment and cell differentiation (Figure 4C).

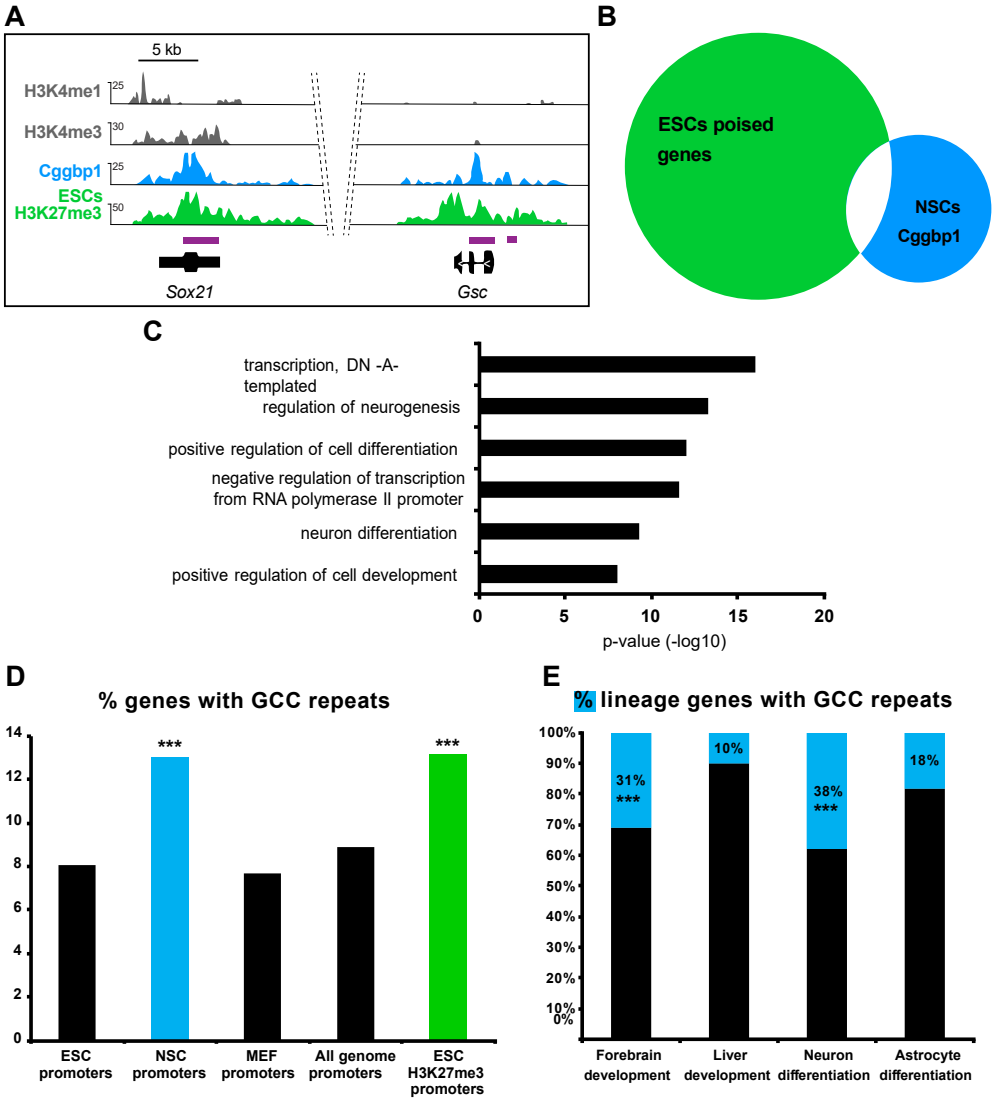


Figure 4 | Cggbp1 binds promoters of neural induction genes poised in ESCs
(A) Chip-seq tracks of Cggbp1, histone modifications H3K4me3, H3K4me1 from NSCs and H3K27me3 from ESCs at different genes loci. Purple bars indicate CpG islands. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated. (B) Venn diagram representing the overlap between ESCs H3K27me3 promoter peaks and NSCs Cggbp1 promoter peaks. (C) GO analysis on overlap genes from B. (D) Genome-wide quantification of GCC(3) repeats at gene subsets promoters (+/- 1,5Kb from TSS). Significance assessed by Chi square test (Yates correction, *** p<0.001). (E) Genome-wide quantification of GCC(3) repeats at gene subsets promoters (+/- 1,5Kb from TSS). Gene subsets extracted by GO term clusters. Significance assessed by Fisher exact test (*** p<0.001).

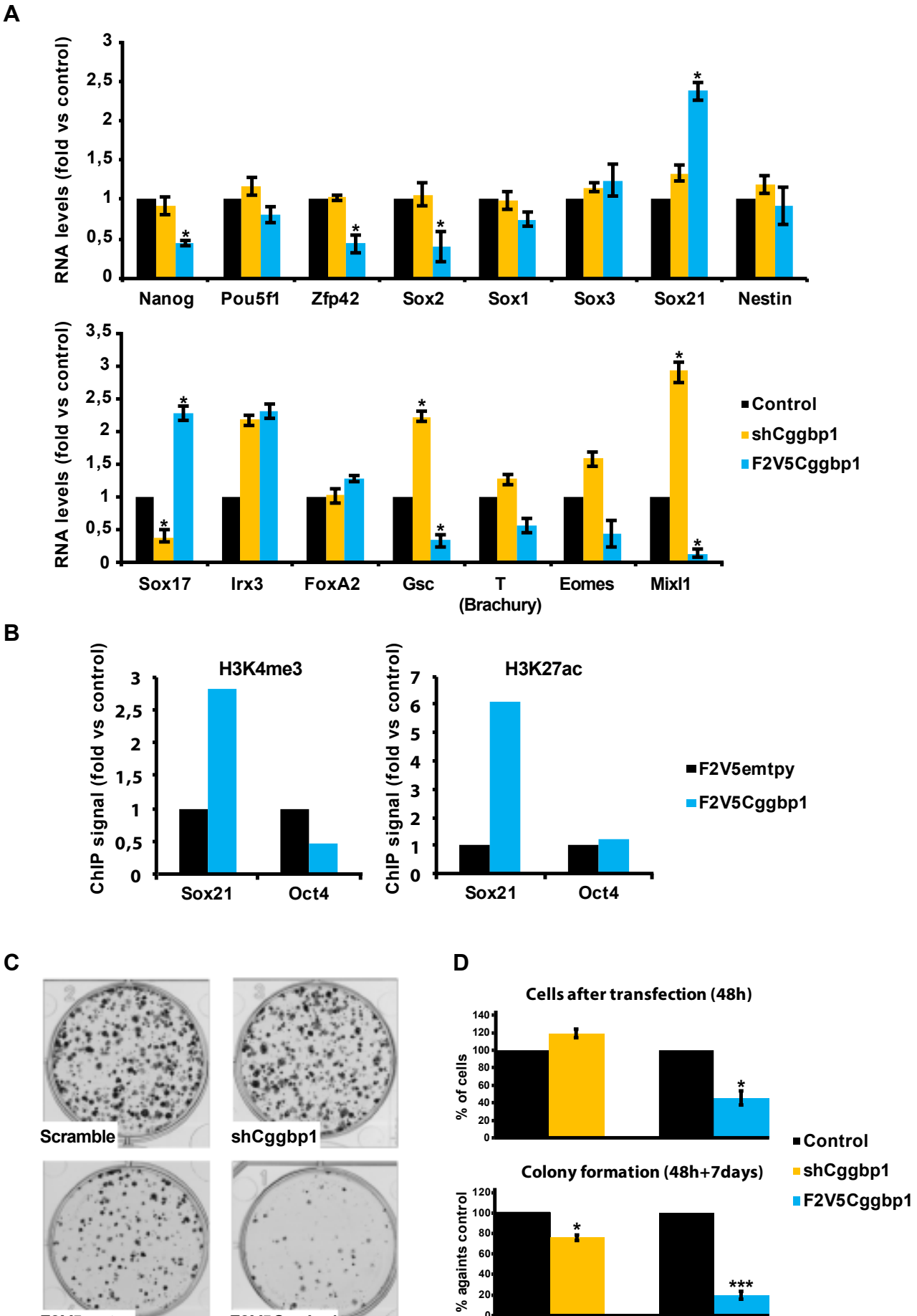
Interestingly, when we analyzed cell specific genes from different lineages, we found a significant enrichment for the Cggbp1 DNA binding motif at ESCs poised promoters at the same levels as in NSCs promoters but not in the whole ESCs population neither at mouse embryonic fibroblasts (MEFs) promoters (Figure 4D). We found that forebrain developmental genes have significantly more $GCC_{(3)}$ motifs than genes from an endodermal lineage subset such as liver development. Moreover, among the neural lineage, neuronal genes were noticeably more enriched with the Cggbp1 motif than astrocyte genes (Figure 4E).

In summary, Cggbp1 motif is overrepresented in neural lineage genes, of which many of them are poised in ESCs.

To characterize the physiological role of Cggbp1 during early ESC differentiation, we evaluated its transcriptional regulation effects on mouse ESCs. After 48hours, ESCs transfected with either shCggbp1 or F2V5Cggbp1 presented a clear misregulation of genes involved in pluripotency and early lineage specification (Figure 5A, upper graphic). While Cggbp1 knock down had no effect on pluripotency genes, its overexpression caused a significant reduction of *Nanog* and *Zfp42* (also known as *Rex1*) together with *Sox2*, all known to be reduced during ESC differentiation. Surprisingly, Cggbp1 overexpression did not trigger the upregulation of neural induction genes found to be bound in NSCs such as *Sox1* and *Sox3* but only *Sox21*. A direct activation effect of Cggbp1 on the *Sox21* promoter could also be observed at the epigenetic level as early as 24hours after F2V5Cggbp1 transfection (Figure 5B). Interestingly, the alteration of the normal levels of Cggbp1 caused a robust regulation of mesodermal markers. Furthermore, *Sox17*, an endoderm marker but found to be bound by Cggbp1 in NSCs, responded positively to Cggbp1 overexpression. Finally, *Irx3* both expressed during mesodermal and neural specification was upregulated in both conditions (Figure 5A, lower graphic).

Considering the drastic effects on pluripotency and lineage markers found in transient transfected ESCs, we proceeded to evaluate their renewal capacity after loss or overexpression of Cggbp1. We observed that the upregulation but not absence of Cggbp1 produced a defect in ESC colony formation (Figure 5C and 5D). On the basis of our findings we propose a role for Cggbp1 in neural lineage specification and pluripotency exit.

Figure 5 (at next page)| Cggbp1 induces loss of pluripotency and neural lineage specification
(A) qPCR analysis on ESCs after 48h transfection with scramble or Cggbp1-targeting shRNA and F2V5empty or F2V5Cggbp1 constructs. Values normalized to their respective controls. Error bars represent SEM, *p < 0.05, unpaired Student's t test, n=2. (B) Normalized histone modifications ChIP signal values at different loci in ESCs after 24h transfection with F2V5empty or F2V5Cggbp1. (C) Plate wells stained with coomassie blue to visualize ESCs 9 days colony formation after transfection with scramble or Cggbp1-targeting shRNA and F2V5empty or F2V5Cggbp1 constructs. (D) Quantification of number of cells after 48h transfection with scramble or Cggbp1-targeting shRNA and F2V5empty or F2V5Cggbp1 constructs or number of colonies after seeding 500 cells from each condition. Values normalized to their respective controls. Error bars represent SEM, *p < 0.05, unpaired Student's t test, n=2.


A
RNA levels (fold vs control)
Nanog
Pou5f1
Zfp42
Sox2
Sox1
Sox3
Sox21
Nestin
Sox17
Irx3
FoxA2
Gsc
T (Brachury)
Eomes
Mixl1
Control
shCggbp1
F2V5Cggbp1
B
ChIP signal (fold vs control)
H3K4me3
H3K27ac
Sox21
Oct4
F2V5emtpy
F2V5Cggbp1
C
Scramble
shCggbp1
F2V5emtpy
F2V5Cggbp1
D
Cells after transfection (48h)
% of cells
Colony formation (48h+7days)
% againts control
Control
shCggbp1
F2V5Cggbp1

Cggbp1 regulates *Sox21* during early ESCs neural induction mediated by retinoic acid

Subsequently, we studied 1 the effect of Cggbp1 in a dynamic system of ESC differentiation. We used the embryonic body (EB) model of ESC differentiation where neural fate is acquired by the removal of leukemia inhibitory factor (LIF) and the addition of retinoic acid (RA)[12]. First, we analyzed the behavior of known markers to evaluate the profile of differentiation of our system. We could observe that pluripotency markers such *Zfp42* and *Oct4* (also named *Pou5f1*) were reduced rapidly while neural markers such as *Nestin* started to be upregulated after RA stimulation (Figure 6A). Intriguingly, *Cggbp1* showed a powerful induction within the first days of RA treatment followed by a downregulation; in contrast with the steady increase of *Nestin* expression.

Focusing on the Cggbp1 locus revealed not only an upstream H3K27me3 poised enhancer in ESCs but also a clear binding by the Retinoic acid receptor alpha (RARa) both at the enhancer and the promoter of the Cggbp1 gene implying a direct activation by RA (Figure 6B). Indeed, *Cggbp1* levels were increased in ESCs cultured in pluripotency conditions 24 hours after stimulation with RA and then returned to initial levels after another day (Figure 6C). These experiments also allowed us to observe the early RA ESC response. First, a primary quick response takes place within the first 6h where *Sox1* is already activated and the first decrease in pluripotency is observed by a decrease in *Zfp42* levels. However, the expression of certain pluripotency markers is partially retained until a secondary response around 12-24 hours where a peak on *Cggbp1* and *Sox21* is observed, suggesting these neural inducers could have a transient upregulation during neural specification that mediates the definitive exit from pluripotency.

As we unraveled the short burst of Cggbp1 levels after RA treatment, we designed a short dynamic differentiation experiment in order to bypass several days of EB formation where the transfection effects would be lost (see methods). Only 24h without LIF was enough to prepare the cells for RA induction, although the effects were not as pronounced as in the EB protocol (Figure 6A). However, this approach supports the study of ESC differentiation towards either the neural or the mesodermal lineages in parallel, using the same induction timing. Indeed, this allowed us to observe that neural induction stimulated *Cggbp1* and *Sox21* expression while mesodermal induction produced opposite effects (Figure 6D). Moreover, the knock down effects were preserved until day 5 (d5) making it possible to observed a Cggbp1 dependency on the RA mediated *Sox21* induction (Figure 6E). Therefore, we conclude that Cggbp1 acts in early neural induction by inducing transient genes such *Sox21* in response to retinoic acid.

Cggbp1 associates with Polycomb

Cggbp1 association with poised promoters in ESCs suggests a role in the activation of poised genes with a preference for the neural lineage. To better understand how Cggbp1 regulated these genes, we have identified its protein-protein interaction partners in ESCs by FLAG affinity purification coupled to mass spectrometry. We found several Polycomb components interacting specifically with Cggbp1

(Figure 7A). On the other hand, we also observed many components of NuRD and SWI/SNF, remodeling complexes more related to activation rather than silencing.

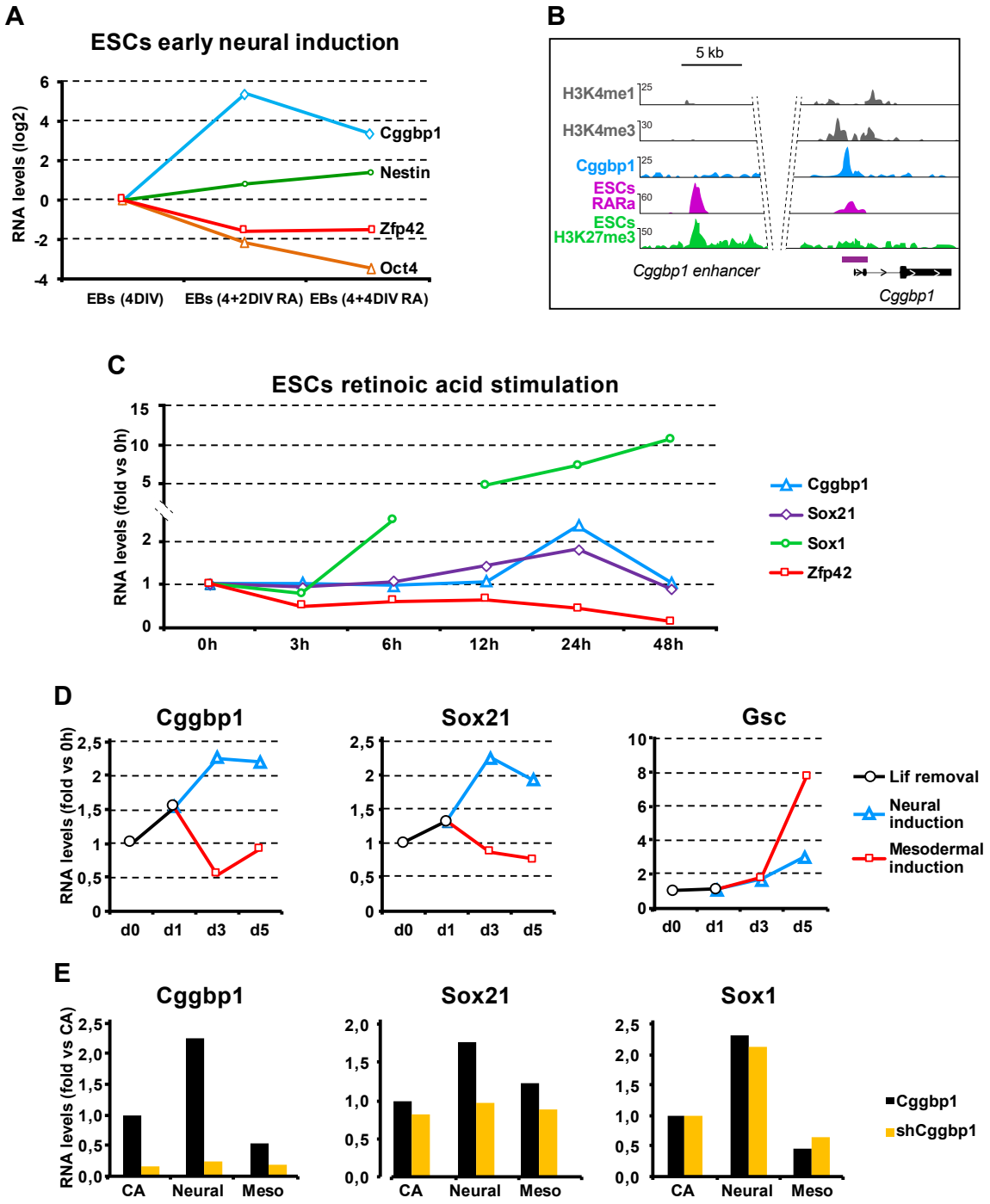


Figure 6 | Cggbp1 regulates Sox21 during early ESCs neural induction mediated by RA

(A) RNA expression profiles of several genes during early ESCs EB differentiation with RA, n=1. (B) Chip-seq tracks of Cggbp1, histone modifications H3K4me3, H3K4me1 from NSCs and H3K27me3 and RARa from ESCs at Cggbp1 locus. Purple bars indicate CpG islands. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated. (C) RNA expression profiles of several genes during ESCs stimulated with retinoic acid, n=1. (D) RNA expression profiles of several genes during short neural or mesodermal induction of EBs. (E) RT-qPCR analysis of ESCs genes after transfection with scramble or Cggbp1-targeting shRNA constructs and treated with CA media (no LIF) or with neural or mesodermal inducer media for 4 days. Values normalized to their respective controls, n=1. EB, embryonic body; RA, retinoic acid; RARa, retinoic acid receptor alpha.

Although Cggbp1 levels are lower in ESCs than NSCs, we were able to ChIP Cggbp1 at poised promoters suggesting that Cggbp1 is already present on some of its targets before its activation by RA (Figure 7B).

A

		F-Cggbp1	Control
Polycomb	Suz12	0,27	0,00
	Rnf2	0,18	0,00
	Eed	0,13	0,00
	L3mbtl3	0,03	0,00

		F-Cggbp1	Control
NuRD complex	Mta2	0,59	0,14
	Gatad2a	0,48	0,16
	Chd4	0,29	0,03
	Gatad2b	0,24	0,00
	Mbd3	0,22	0,00

		F-Cggbp1	Control
SWI/SNF complex	Smarcb1	0,43	0,00
	Smarca4	0,28	0,05
	Smarcc2	0,23	0,00
	Arid2	0,10	0,00
	Smarcad1	0,09	0,00
	Brd7	0,09	0,00

		F-Cggbp1	Control
Transcription factors	Hmgb1	0,86	0,00
	Hmgb2	0,69	0,00
	Mybbp1a	0,46	0,00
	Zbtb39	0,44	0,04
	Tex10	0,32	0,00
	Znf593	0,22	0,00
	Zzef1	0.06	0.00

B

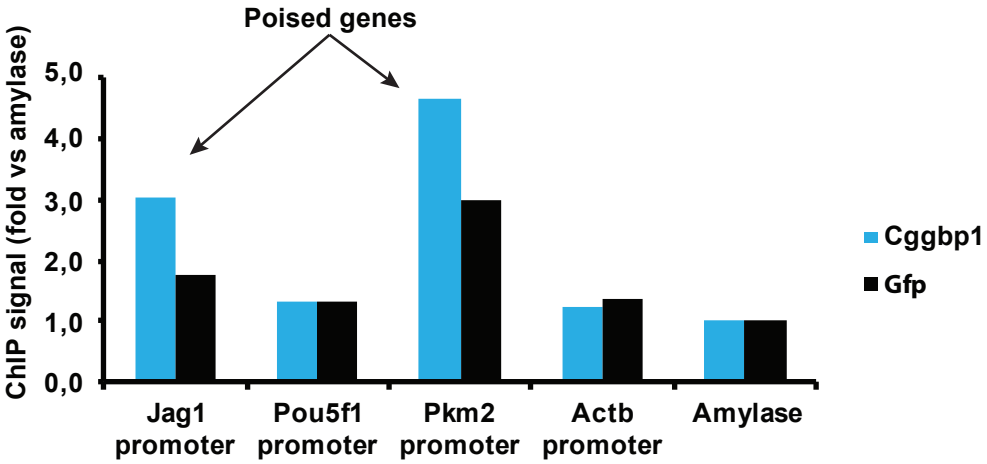


Figure 7. Cggbp1 associates with Polycomb in ESCs.
(A) Mass spectrometry emPAI scores for proteins found in Cggbp1-Flag and control purifications in ESCs after 24h transfection with F2V5Cggbp1 construct. (B) Normalized Cggbp1 and GFP control ChIP signal values at different loci in ESCs. n=1.

The ability to bind both silencing Polycomb proteins but also activating remodelers makes us postulate that Cggbp1 may act as a switch for poised promoters during the transition to the neural lineage.

DISCUSSION

Although Cggbp1 has been poorly studied, several reports from Westermark's lab have indicatednew functions for this protein, ranging from transcription both at heat shock genes and Alu repeats, genomic integrity, cell cycle and DNA methylation[13]. However, despite that its ubiquitous expression is only altered at early development, where it is enriched in the brain and testis, Cggbp1 role in embryogenesis has not been explored.

Here, we have identified Cggbp1 as a transcription factor acting both in NSCs and in early neural induction. Our data indicate that Cggbp1 can act as an activator, in contrast with the repressing roles previously reported[5,9]. We support this claim not only by the downregulation of Cggbp1 targets upon Cggbp1 knock down in NSCs but both the epigenetic activation of the Sox21 after 24h Cggbp1 overexpression in ESC and the dependence on Cggbp1 for the recruitment of the Mediator complex at one of its targets in NSCs. In relation with a previous report denying a role in *Fmr1* gene regulation[10], we hypothesize that the cell type model may be important in order to evaluate Cggbp1 effects. We suggest that our NSCs may be more representative than lymphoblasts or fibroblasts used in that study, in part because Cggbp1 and FMR1 have neural functions and Cggbp1 regulation is highly dependent on post-translation modifications triggered by growth factors.

Although we found a small fraction of Cggbp1 associated with Alu/SINE repeats (data not shown) almost 80% of Cggbp1 occupancy consisted in gene promoters with the $GCC_{(3)}$motif, highly correlated with CpG islands. Interestingly, CpG island promoters are largely found in early developmental genes marked by DNA methylation valleys (DMVs), which are devoid of DNA methylation and use a H3K27me3-Polycomb-based mechanism for silencing[14].

Recently, a wide characterization of *cis* and *trans* components associated to Polycomb repressive complex 2 (PRC2) in *Arabidopsis* has given exciting hints on the Cggbp1 mechanism of early neural induction[15]. Xiao and colleagues identified several *cis* motifs in Polycomb reponse elements (PRE) with CCG/GCC as the second most significant one. Polycomb regulation is widely conserved in metazoans suggesting that Cggbp1 could be one of the paralogue zing finger proteins binding GCC in mammalian PREs.

In addition recent discoveries suggest that enhancer-promoter loops are already established in poised genes[16]. Our finding that low levels of Cggbp1 are already binding ESCs poised genes together with the fact that Cggbp1 defects in ESCs mimics a loss of PRC2 in differentiating ESCs supports a putative role for Cggbp1 in Polycomb response, possibly as a recruiting factor at poised promoters.

Remarkably, Cggbp1 association with Polycombappears not to end at the protein level as we identified PRC1 components *Cbx4* and *Cbx8* as targets of Cggbp1 (Figure Supl. A). In an exciting

recent study, it has been shown that Polycombmediated regulation of poised genes requires a switch in PRC1 component *Cbx7* to be replaced by *Cbx2*, *Cbx4* and/or *Cbx8* in order to resolve poised genes to the activation phase[17–20]. Indeed, we observed a change in expression for several Polycombcomponents depending on the cell type (Figure Suppl. 3B). When tested in our dynamic ESCs induction protocol, we observed that *Cbx4* and *Cbx8* were induced during neural but not mesodermal stimulation and that Cggbp1 depletion greatly affected that induction (Figure Suppl. 3C).

On the verge of more conclusive data, we hypothesize that Cggbp1 acts during the transition from pluripotency to the neural lineage by direct induction of neural gene targets but also by modulating the Polycomb response.

MATERIAL AND METHODS

Neural stem cell culture and transfection

NS-5 neural stem cells (NSCs) derived from 46C embryonic stem cells were cultured as described[21] and regularly tested for mycoplasma contamination and for authenticity by expressed NSC markers Pax6, Sox2 and Nestin[22].

NSCs were electroporated as previously described[23] with pCAG promoter-driven plasmids containing the FLAG-V5 tags and Cggbp1 cDNA in parallel to an empty pCAG FLAG-V5 in overexpression experiments. pSuper-puro constructs encoding Cggbp1 short hairpin RNA (shRNA sequence: 5′- GGTGAGCTTCATGAGGATG-3′), or a control scrambled sequence were used in knock down experiments. Puromycin (2 μg per ml) was added after 24 hours and NSCs were collected for the several analyses at 40 h after electroporation.

Neural stem cell viability quantification

After electroporation, NSCs from each condition were counted and re-seeded in triplicates in p96 plates 1000 cells per well. The next day (48h) or the day after (72h) cells were incubated with alamarBlue (DAL1025, Invitrogen) and viability was quantified by fluorescence-based instrumentation.

Embryonic stem cell culture and transfection

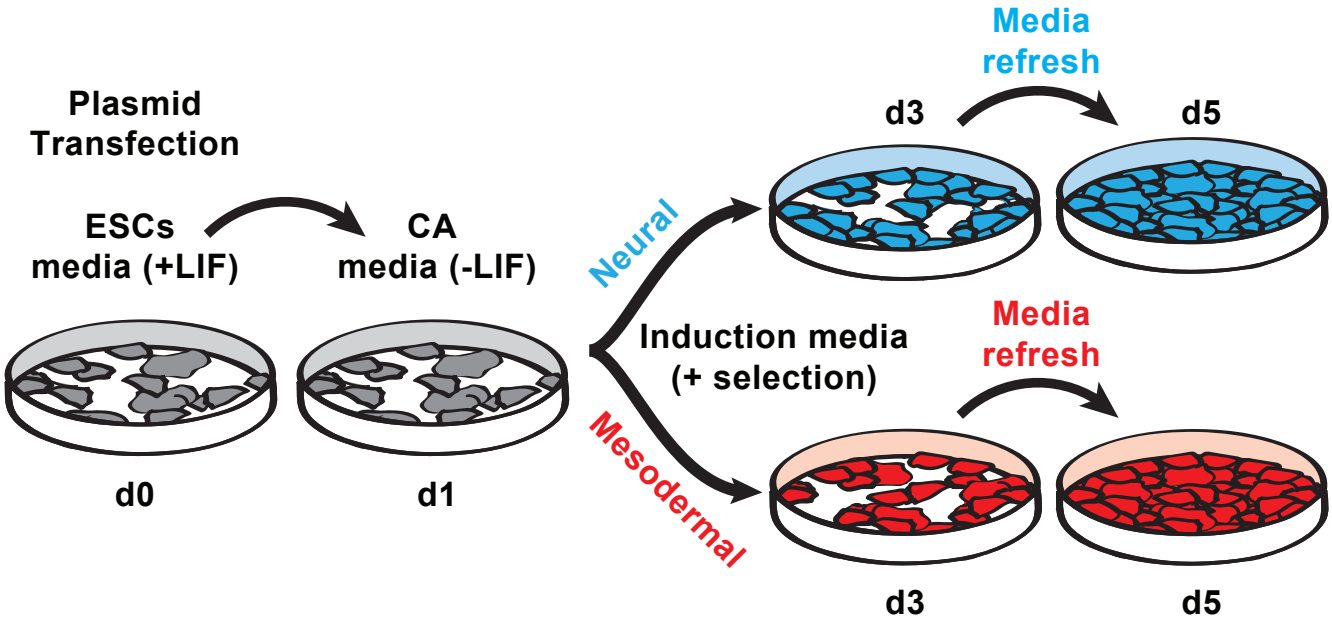
Mouse ESC (46C line)[24] were grown on gelatin-coated dishes without feeders in Glasgow minimal essential medium (GMEM) supplemented with leukemia inhibitory factor (LIF), 15% fetal bovine serum, 0.25% sodium bicarbonate, 1 mM glutamine, 1 mM sodium pyruvate, nonessential amino acids, 50 μM beta-mercaptoethanol, and penicillin-streptomycin, as described previously[25]. Plasmid transfections were performed with Lipofectamine 2000 (Invitrogen) and puromycin (2ug or 1ug per ml for 24 or 48 hour experiments, respectively) was used to select expressing cells.

Embryonic stem cell colony formation

3*10E6 ESCs were transfected and selected with 1ug/ml puromycin for 48 hours. Cells were collected and counted (Figure C, upper graph) and 500 cells were re-seeded in triplicates into p6 plate wells in normal ESCs media. 7 days after re-seeding colonies were stained using Coomassie Brilliant Blueand counted using GelCountTM (Oxford Optronix, version 1.1.2.0).

Embryonic stem cell differentiation protocols

We used a standard embryonic body differentiation protocol described in[26] for Figure 6A. In addition we adapted the same principle plus a mesoderm (heart) differentiation protocol[27] to design a short term neural/mesoderm parallel differentiation protocol compatible with short term transfection (for Figure 6D).



The basis of this protocol is the short preparation phase without LIF to adherent ESCs cultures before the application induction media. After normal lipofectamine transfection the media is replaced the same day into ESCs media without LIF (CA media). After 24 hours, induction media (either neural with RA, or mesodermal with ascorbic acid and activin) are added together with the selection agent (by 1ug/ml puromycin). The same media is used to refresh the culture 48 hours later. For Figure 6E we extended a control condition without induction media (just CA media) for 4 extra days to compensate for the possible effects that a removal of LIF alone could explain.

Immunocytochemistry

Cells were grown on poly-D-lysine–coated cover slips and fixed in 4% formaldehyde, 15min at room temperature. After fixation, they were permeabilized with Triton X-100 in PBS and blocked with blocking buffer (10% FCS, 0.2 M glycine, 0.1% Triton X-100 and 0.05% Deoxycholatein PBS-2%

gelatin) for 1 h at room temperature. To label the cells, the following antibodies or dyes were used: anti-Nestin (mAb1259, R&D systems) and anti-Cggbp1 in blocking buffer for 2 h and with the corresponding secondary antibodies labeled with fluorochromes (Alexafluor 546 or 488, Invitrogen). Samples were mounted in MOWIOL (#324590 Sigma-Aldrich) and nuclei were stained with DAPI (Vector laboratories).

Images were capture by confocal microscopy using a Leica SP5 microscope. Image processing was performed using FIJI (ImageJ)[28].

Affinity protein purification of Cggbp1

Embryonic stem cells were transfected with pCAG promoter-driven plasmids containing C-terminally FLAG-tagged Cggbp1. After 24h cells were harvested and nuclear extract was prepared from ESCs expressing FLAG-Cggbp1 and from control ESCs[29]. Nuclear extracts, equivalent to $2*10^8$ ESCs were incubated with Benzonase (150U per ml nuclear extract) and used in FLAG-affinity purifications together with nuclear extracts from control ESCs, as described [22,25]. Identification of proteins by mass spectrometry was as described[25].

Similarly, 1 ml of nuclear extracts from NSCs containing FLAG-Cggbp1 or parental lines were used to perform large-scale immunoprecipitations. The resulting western blots were probed with antibodies against Med12 antibody (Bethyl Laboratories #A300-774A) and Cggbp1 antibody (Bethyl Laboratories #A304-037A).

Chromatin immunoprecipitation combined with high-throughput sequencing (ChIP-seq)

For Cggbp1, we adapted protocols previously described in Chapter 2 of this thesis with some modifications. We single crosslinked the cells with 1% formaldehyde solution for 30 min at room temperature. We apply the same incubation and washing buffer replacing Tris-HCl for HEPES solutions. Primers used for small scale ChIP-QPCR can be found in [22] in addition to:

Sox21_P_F	CTGATCTCCGAGTTGTGCAT	Pou5f1_P_F	GATCCTCGAACCTGGCTAAG
Sox21_P_R	CAGCATGTCCAAGCCTGT	Pou5f1_P_R	CCAACCTGAGGTCCACAGTA

ChIP-seq data analysis

Together with our own Cggbp1 ChIP-seq data we used the following set of published data to complement our study:

NSCs_Med1	GSM2928425
NSCs_mNSCs_Input	GSM1187180
NSCs_H3K4me3	SRR006888_SRR006889
NSCs_H3K4me1	SRR002250_SRR002251_SRR002252
ESCs_H3K27me3	GSM307619
ESCs_RARa	SRR627785_SRR627786

In addition we download CpG Island coordinates in mm9 mouse reference genome from UCSC genome browser. Finally, RNA-seq data from wild-type mNSCs was extracted from[23].

Most of the ChIP-seq analysis, after the first processing steps similar as in Chapter 2 of this thesis, were performed using HOMER[30] and BEDTOOLS[31]. In order to characterize the enrichment of GCC repeats across the genome in Figure 4D we curated lists of specific genes from ESCs, NSCs and MEFs using the published expression analysis[32] by selecting genes that were expressed 5 times higher to the other two conditions for each subset. Next, we extracted their DNA promoter sequences (1,5 Kb around the TSS) and proceed to quantify $GCC_{(3)}$ occurrence on each subset. We added gene promoters marked with H3K27me3 in ESCs (own analysis) and the whole list of mouse promoters to our analysis. We followed a similar approach for Figure 4E but we subtracted gene lists according to the terms depicted in the figure from the Gene Ontology consortium database[33] and applied the same analysis.

Expression analysis

Quantitative real-time PCR analyses on complementary DNA transcribed from total RNA with Superscript IV Reverse Transcriptase (#18090010 Invitrogen) was performed on a DNA Engine Opticon2/CFX96 (Biorad) and normalized for B-Actin expression. Primer sequences used can be found in [22,23] in addition to:

Cbx4_Forw	TGCTGATCGCCTTCCAGAA
Cbx4_Rev	GGGCCCTCTCTTGCGATATC
Cbx8_Forw	GAGGACCCAAGCCTAAAACC
Cbx8_Rev	CCTGGAAGTAGACGCCAAATC
B-Actin_forw	TCTCCTTCTGCATCCTGTCAGCAA
B-Actin_rev	TCTTGGGTATGGAATCCTGTGGCA
Sox17 _forw	ACTTGCTCCCCACAATCACT
Sox17 _rev	ACCCCGCTGTTTGTGTTTAG
FoxA2 _forw	TGGTCACTGGGGACAAGGGAA
FoxA2 _rev	CTGCAACAACAGCAATAGAGAACAA
T_forw	AAGGAACCACCGGTCATC
T_rev	GTGTGCGTCAGTGGTGTGTAATG
Mixl1_forw	GCACGTCGTTCAGCTCGGAG
mixl1_rev	GTCATGCTGGGATCCGGAACGTG
Eomes_forw	GGTACGGCGTTCAAAACTTC
Eomes_rev	ATGGGAGCAAGGTACTGGAA
Gsc_forw	TCCAGGAGACGAAGTACCCAGACGT
Gsc_rev	CTCGGCGGTTCTTAAACCAGACCT

Sox1_forw	GTTTTTTGTAGTTGTTACCGC
Sox1_Rev	GCATTTACAAGAAATAATAC
Zfp42_forw	TTGGGGCGAGCTCATTACTT
Zfp42_rev	TTGCCACACTCTGCACACAC
Nanog_forw	ATGAAGTGCAAGCGGTGGCAGAAA
Nanog_rev	CCTGGTGGAGTCACAGAGTAGTTC
Sox21_f	CCTAAGATGCACAACTCGGA
Sox21_r	CTTGTAGTCGGGATGCTCC
Pou3f3_F	CTCACCTCCCGTCCATGG
Pou3f3_R	GATGGTGGTGATGATGCTCC
Fmr1_F	agatcaagctggaggtgcca
Fmr1_R	cagagaaggcaccaactgcc
Nfib_Forw	GGACAGCTTTGTAAAATCCGGA
Nfib_Rev	GTGGAGAAGACAGCGACCT
Cggbp1_FW	TGCCATTAGTGACCACCTCA
Cggbp1_RV	GGACCCTCCGTTCTTCACAT

SUPPLEMENTAL INFORMATION

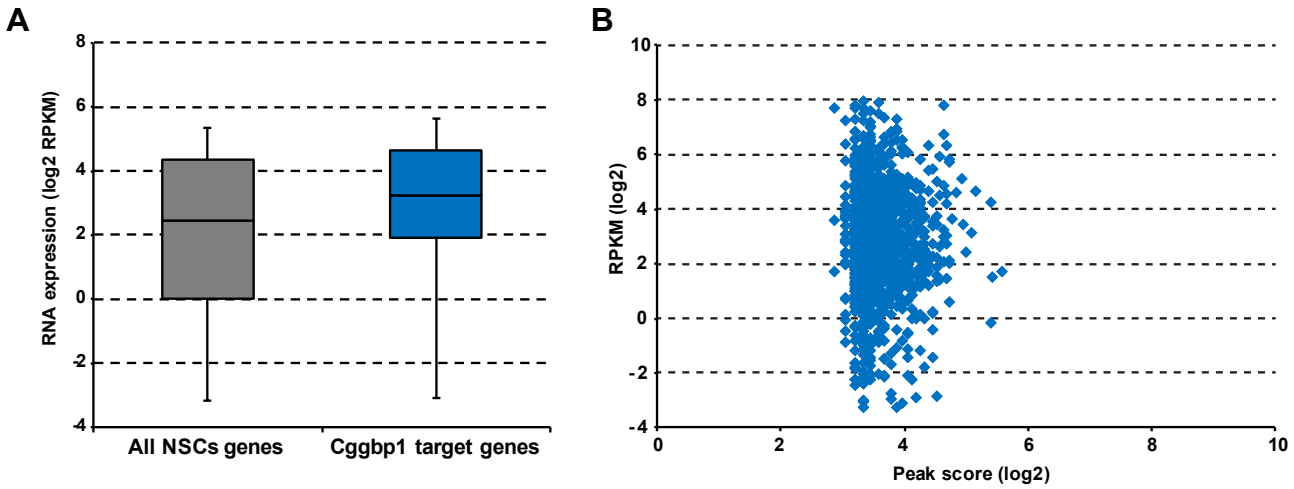


Figure Supplemental 1 | Genome wide RNA correlations
(A) Box plot representation of RNA expression levels of Cggbp1 targets compared to all NSCs expressed genes. Whiskers represent maximum and minimum values. **p < 0.05 unpaired Student's t test. (B) Scatter plot displaying peak score of Cggbp1 Chip-seq data on X axis vs the corresponding RNA expression (RPKM) of the annotated gene on Y axis. Pearson correlation coefficient = 0,01.

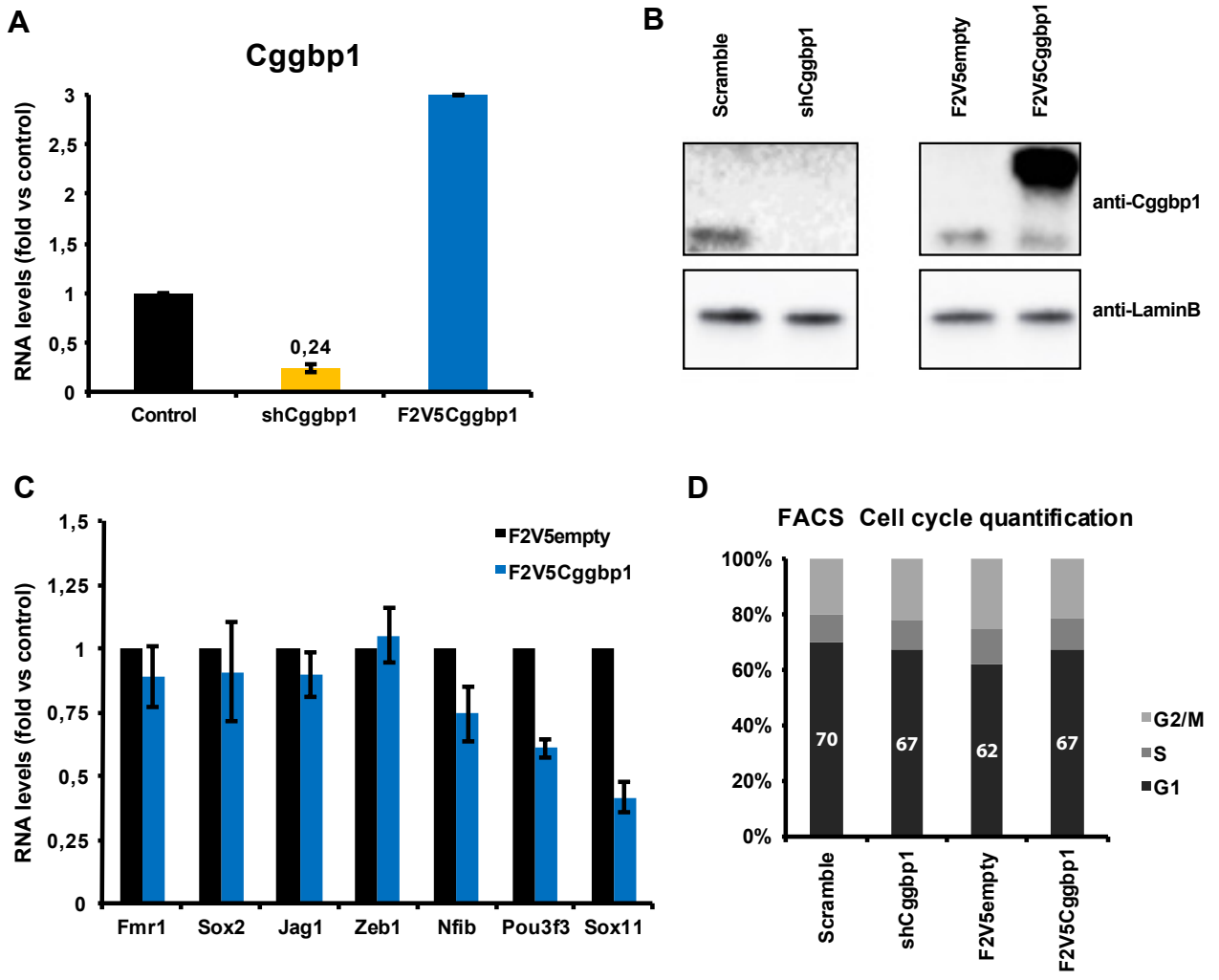


Figure Supplemental 2 (at previous page) | Cggbp1 levels modulation in NSCs.
(A) QPCR analysis on NSCs after 40h transfection with scramble or Cggbp1-targeting shRNA and F2V5empty or F2V5Cggbp1 constructs. Values normalized to their respective controls. Error bars represent STD, **p < 0.01, unpaired Student's t test, n=2. (B) Cggbp1 protein levels analyzed by western blot from NSCs after 40h transfection with scramble or Cggbp1-targeting shRNA and F2V5empty or F2V5Cggbp1 constructs. (C) qPCR analysis on NSCs after 40h transfection F2V5empty or F2V5Cggbp1 constructs. Values normalized to their respective controls. Error bars represent STD, *p < 0.05, unpaired Student's t test, n=2. (D) FACS cell cycle quantification of NSCs stained with propidium iodide after after 40h transfection with scramble or Cggbp1-targeting shRNA and F2V5empty or F2V5Cggbp1 constructs. n=1.

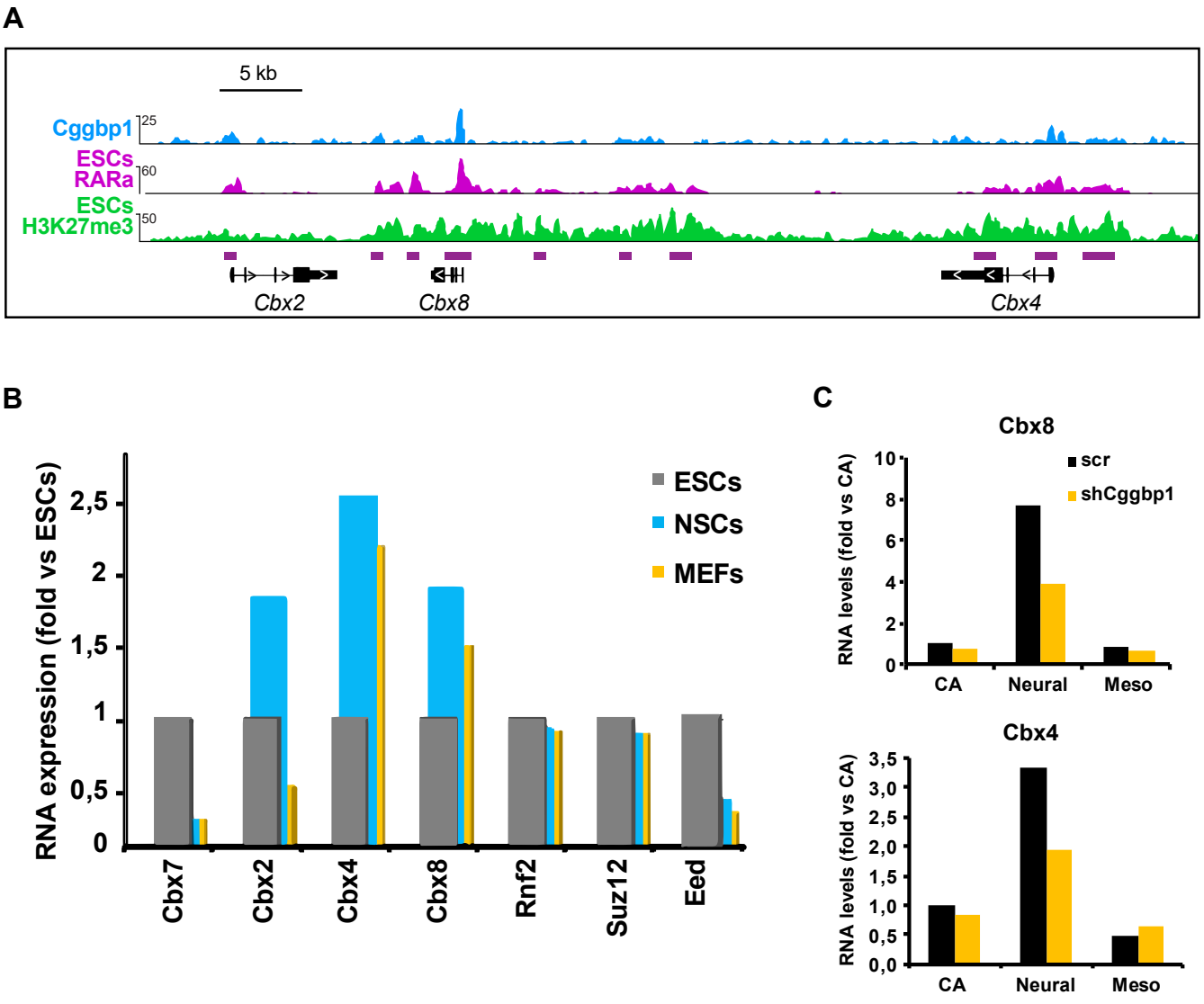


Figure Supplemental 3 | Cggbp1 as a Polycomb protein
(A) Chip-seq tracks of Cggbp1, histone modifications H3K4me3, H3K4me1 from NSCs and H3K27me3 and RARa from ESCs at Polycomb gene loci. Purple bars indicate CpG islands. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated. (B) Normalized RNA expression levels of Cggbp1 and Polycombgenes in different cell types, normalized to ESCs relative expression. Data from[34]. (C) qPCR analysis of ESCs Polycomb genes after 48h transfection with scramble or Cggbp1-targeting shRNA constructs and treated with neural or mesodermal inducers. Values normalized to their respective controls, n=1.

REFERENCES

1. Greene, N. D. E. & Copp, A. J. Development of the vertebrate central nervous system: formation of the neural tube. *Prenat. Diagn.* **29,** 303–311 (2009).

2. Rogers, C. D., Moody, S. A. & Casey, E. S. Neural induction and factors that stabilize a neural fate. *Birth Defects Res. Part C - Embryo Today Rev.* **87,** 249–262 (2009).

3. van Kruijsbergen, I., Hontelez, S. & Veenstra, G. J. C. Recruiting polycomb to chromatin. *Int. J. Biochem. Cell Biol.* **67,** 177–187 (2015).

4. Müller-Hartmann, H. *et al.* The human 20-kDa 5′-(CGG)(n)-3′-binding protein is targeted to the nucleus and affects the activity of the FMR1 promoter. *J. Biol. Chem.* **275,** 6447–6452 (2000).

5. Singh, U., Bongcam-Rudloff, E. & Westermark, B. A DNA sequence directed mutual transcription regulation of HSF1 and NFIX involves novel heat sensitive protein interactions. *PLoS One* **4,** (2009).

6. Agarwal, P. *et al.* Growth signals employ CGGBP1 to suppress transcription of Alu-SINEs. *Cell Cycle* **15,** 1558–1571 (2016).

7. Jeronimo, C. & Robert, F. The Mediator Complex: At the Nexus of RNA Polymerase II Transcription. *Trends Cell Biol.* **xx,** 1–19 (2017).

8. Martynoga, B. *et al.* Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. *Genes Dev.* **27,** 1769–1786 (2013).

9. Singh, U., Roswall, P., Uhrbom, L. & Westermark, B. CGGBP1 regulates cell cycle in cancer cells. *BMC Mol. Biol.* **12,** 28 (2011).

10. Goracci, M. *et al.* Defining the role of the CGGBP1 protein in FMR1 gene expression. *Eur J Hum Genet* **24,** 697–703 (2016).

11. Naumann, F., Remus, R., Schmitz, B. & Doerfler, W. Gene structure and expression of the 5′-(CGG)n-3′-binding protein (CGGBP1). *Genomics* **83,** 106–118 (2004).

12. Lu, J. *et al.* All-trans retinoic acid promotes neural lineage entry by pluripotent embryonic stem cells via multiple pathways. *BMC Cell Biol.* **10,** 57 (2009).

13. Singh, U. & Westermark, B. CGGBP1--an indispensable protein with ubiquitous cytoprotective functions. *Ups J Med Sci* **120,** 219–232 (2015).

14. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153,** 1134–1148 (2013).

15. Xiao, J. *et al.* Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. *Nat. Genet.* (2017). doi:10.1038/ng.3937

16. Cruz-Molina, S. *et al.* PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell* **20,** 689–705.e9 (2017).

17. O'Loghlen, A. *et al.* MicroRNA regulation of Cbx7 mediates a switch of polycomb orthologs during ESC differentiation. *Cell Stem Cell* **10,** 33–46 (2012).

18. Morey, L. *et al.* Nonoverlapping functions of the polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **10,** 47–62 (2012).

19. Klauke, K. *et al.* Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. *Nat. Cell Biol.* **15,** 353–362 (2013).

20. Creppe, C., Palau, A., Malinverni, R., Valero, V. & Buschbeck, M. A Cbx8-Containing Polycomb Complex Facilitates the Transition to Gene Activation during ES Cell Differentiation. *PLoS Genet.* **10,** (2014).

21. Conti, L. *et al.* Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.* **3,** 1594–1606 (2005).

22. Engelen, E. *et al.* Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat. Genet.* **43,** 607–611 (2011).

23. Moen, M. J. *et al.* An interaction network of mental disorder proteins in neural stem cells. *Transl. Psychiatry* **7,** e1082 (2017).

24. Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21,** 183–186 (2003).

25. van den Berg, D. L. C. *et al.* An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells. *Cell Stem Cell* **6,** 369–381 (2010).

26. Bibel, M., Richter, J., Lacroix, E. & Barde, Y.-A. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat. Protoc.* **2,** 1034–43 (2007).

27. Gadue, P., Huber, T. L., Paddison, P. J. & Keller, G. M. Wnt and TGF-beta signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 16806–11 (2006).

28. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9,** 676–682 (2012).

29. Dignam, J. D., Martin, P. L., Shastry, B. S. & Roeder, R. G. Eukaryotic gene transcription with purified components. *Methods Enzymol.* **101,** 582–98 (1983).

30. Heinz, S. *et al.* Finding Enriched Peaks, Regions, and Transcripts (findPeaks). (2017).

31. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–2 (2010).

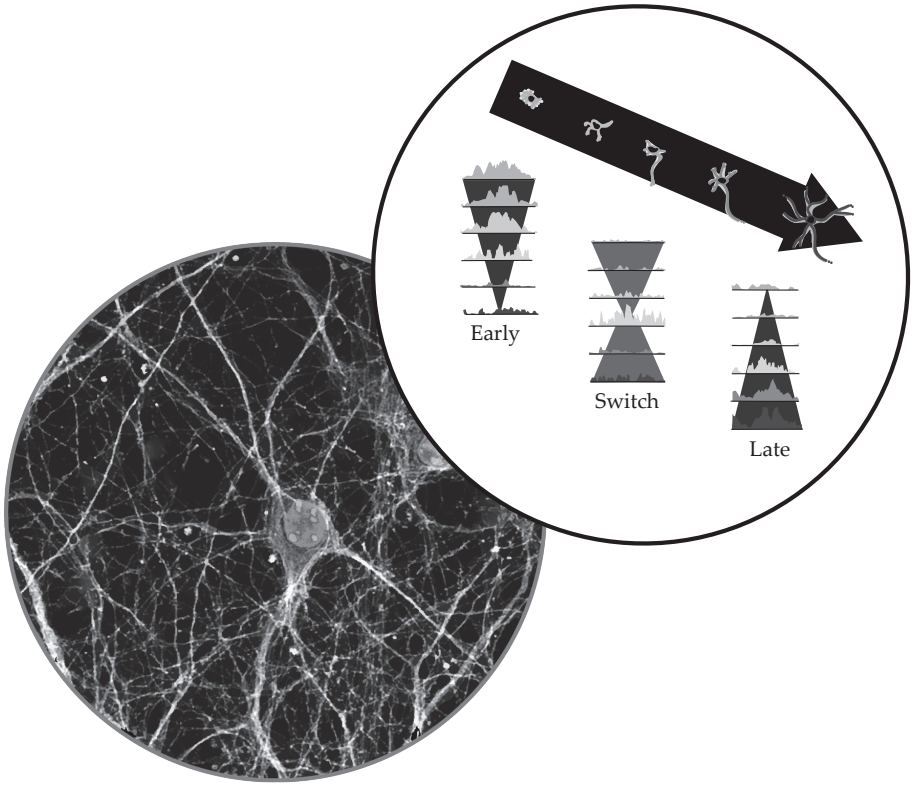
32. Zhu, J. *et al.* Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152,** 642–654 (2013).

33. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29 (2000).

34. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125,** 315–326 (2006).

Chapter 4

A dynamic active chromatin map of neuronal maturation



- Work in progress -

A dynamic active chromatin map of neuronal maturation

Marti Quevedo[1*],Johannes H. Brandsma[1], Zeliha Ozgür[3], Wilfred F.J. van IJcken[3], Raymond A. Poot[1]

[1]Department of Cell Biology, Erasmus MC, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands
[2]Center for Proteomics, Erasmus MC, The Netherlands
[3]Center for Biomics, Erasmus MC, The Netherlands
[*]Corresponding author

ABSTRACT

Neuronal cells endure a long process of maturation before they integrate to functional signaling networks. The correct regulation of the maturation process is essential for the assembly of adult neuronal networks leading to neurodevelopmental disorders when not successfully accomplished. Recent studies have revealed a gene expression switch occurring early in neuron maturation. However, the effectors of these changes remain to be discovered. In the present study we have used chromatin immunoprecipitation coupled to next generation sequencing (ChIP-seq) of the active chromatin mark histone H3 lysine K27 acetylation (H3K27ac) applied to a model of *in vitro* maturation of mouse hippocampal neurons in order to characterize the epigenetic dynamics of neuronal maturation. We have identified several transcription factors specific to different maturation stages and several candidates are suggested based on their *in vivo* expression.

INTRODUCTION

The integration of classic neurophysiology approaches with the next generation of "-omics" disciplines have the potential to surpass the current limitations in nervous system research.

While next-generation techniques have been rapidly implemented in other biological systems, the extreme cellular heterogeneity and complexity of the neural tissue have delayed in great measure their use. Nevertheless, the development of better isolation sorting techniques[1], *in vitro* primary pure neuronal cultures[2], *in vitro* differentiation protocols[3] and single-cell sequencing[4,5] have provided a breach an increasing number of neuroscientists are starting to exploit.

Indeed, the transcriptional networks governing differentiation and functionality of matured neurons have been molecularly explored in depth in the last decade[6–8]. However, the maturation steps from the early post-mitotic phase to a fully synaptic active neuron remains still poorly understood. This transition period has been proven to be essential for the correct assembly of neuronal networks as not only the proper number of cells is important but their development and integration to the system are key variables. Perhaps the best example representing the importance of maturation is Rett syndrome patients, who suffer from several neurodevelopmental defects and present intellectual disability. The major cause for Rett syndrome is the mutation in the chromatin regulator *MeCP2*[9]. MeCP2 levels are tightly controlled during neurogenesis and increase dramatically during neuronal maturation[10]. Recent studies have shown that in matured neurons, MeCP2 can largely substitutes for histone H1 *in vitro* and is distributed throughout the genome affecting nucleosome spacing. In addition, its increased levels allow MeCP2 binding to the epigenetic mark 5hmC in active genes, regulating important synaptic processes[11].

The genome-wide epigenetic changes response during neuronal maturation remain to be defined. Several groups have analyzed the transcriptome changes through well-defined neuronal maturation time points, uncovering a massive switch in the expression profiles of maturing neurons[12–14]. However, the effectors of these changes remain to be discovered.

In this study, we analyze the dynamic changes in the active chromatin regions during hippocampal neuron *in vitro* maturation. Using chromatin immunoprecipitation coupled to next generation sequencing (ChIP-seq) of the active promoter and enhancer mark histone H3 lysine K27 acetylation (H3K27ac) we imply the transcription factor response at several stages of neuronal maturation and propose several transcription factor candidates based to play a role on their *in vivo* expression.

RESULTS

Hippocampal neuronal maturation timing *in vitro*

To establish a comprehensive epigenetic profile during the development of primary mouse hippocampal neurons in culture, we took time points corresponding to the peak periods for which major cellular and physiological events occur during neuronal development in culture as defined by

Dotti et al.[15](Figure 1A). Interestingly, when compared to expression profiles of neurons isolated from developing mousebrains, it was demonstrated that *in vitro* neurons reach maturation sooner[13]. In addition to the described expression profile switch that takes place around 4 days *in vitro* (D IV), *in vivo* brain levels of MeCP2 have been quantified[10] suggesting that its upregulation starts around this switch period. MeCP2 levels inversely correlate to the differential nucleosome repeat length (NRL), or nucleosome spacing, seen in neurons and its shortening represents a macro-signature of the epigenetic changes that these cells undergo[16,17] (Figure 1B).

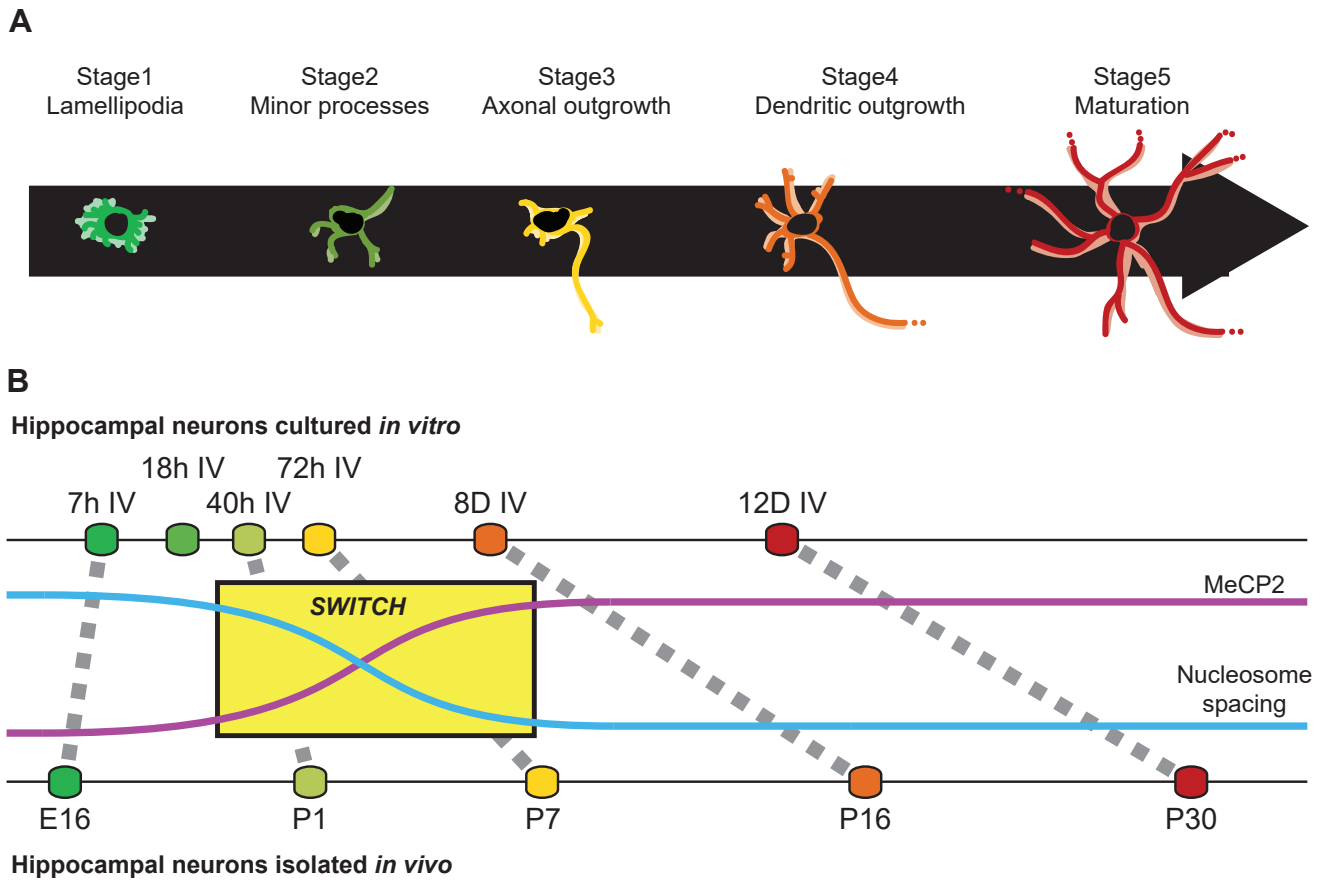


Figure 1 | Hippocampus in vitro culture as a model of neural maturation
(A) Morphologically stages of neural maturation adapted from Dotti et al. 1988 (B) Time line representation of hippocampal maturation. Nucleosome spacing and MeCP2 protein levels are depicted in cyan and purple lines.

Thus, we chose to sample six points between 7 hours and 12 days in culture in order to capture the full process of neuronal maturation. Indeed we observed that briefly after plating, cells have a symmetric appearance extending lamellipodia all around the soma (stage 1, Figure 2A 7h). After 7 hours *in vitro* (h IV) minor neurites are observed (stage 2, Figure 2A 18h). Then, one of the neurites grows without retracting and acquires axonal properties. This axon can be identified in most neurons at 40h IV (stage 3, Figure 2A 40h). The axon continues to grow promptly, whereas the remaining neurites elongate more slowly and become dendrites. At the end of stage 3 (around 72h IV, Figure 2A 72h IV) many axons contact neighboring neurons and the remaining minor neurites acquire

characteristics of dendrites. From 72h to 8D IV (stage 4, Figure 2A 8D IV), a greater number of outgrowing and branching dendrites are observed and the first synapses are formed. Within the next days, an extensive network of synaptic connections will be formed and around 12D IV (stage 5, Figure 2A 12D IV), all cells are fully developed and display a mature neuronal morphology with axons forming a dense mesh on the culture dish[14]. Hence we are able to establish a reliable primary hippocampal neuron culture where all stages of maturation are represented.

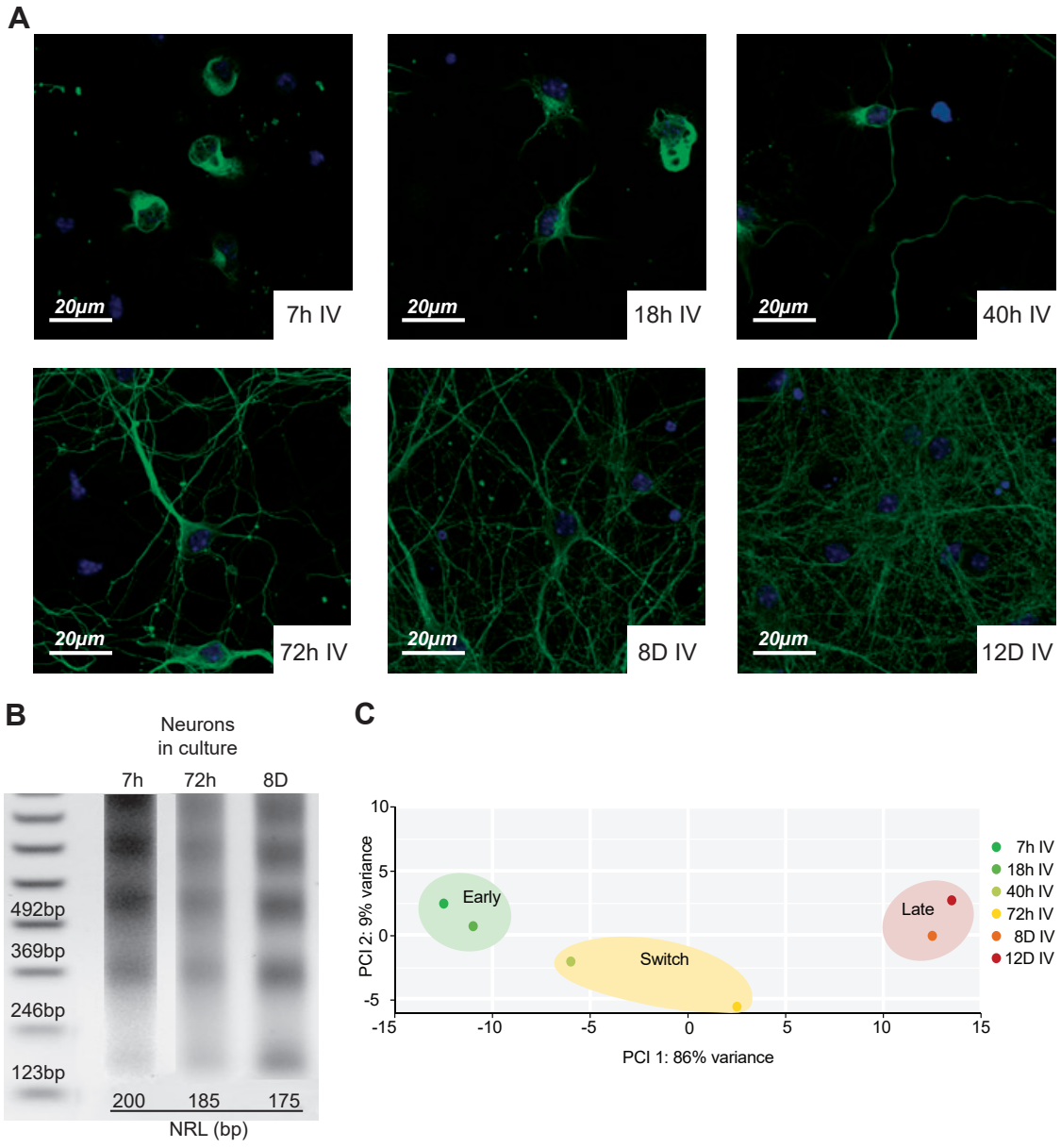


Figure 2 | Confirmation of the epigenetic switch in the hippocampal culture
(A) Immunocytochemistry with Tuj1 (neuron-specific class III beta-tubulin) antibody and dapi staining at different hippocampal neuron cultures time points. (B) Micrococcal nuclease digestion of neuron chromatin isolated from in vitro cultured hippocampal neurons at different times of maturation. (C) Principal component analysis plot of the ChIP-seq data that characterizes the trends exhibited by the different time points. PC1 represents biological variance while PC2 represents experimental variance. Data points are grouped accordingly.

Next we tested if we could observe the switch in NRL, characteristic in healthy adult neurons. By using micrococcal nuclease digestion followed by gel electrophoresis we ascertained that in our hippocampal *in vitro* culture the NLR of neurons started shortening from 72h IV and reached its minimum already at 8D IV (Figure 2B). Thus, our culture system recreates the major chromatin remodelling event during neuronal maturation.

Dynamics of epigenetic active sites during neuronal maturation

We used active chromatin profiling for H3K27ac, a mark of active enhancers and promoters, across the six time points. Principal component analysis of our ChIP-seq data showed a clear distinction of the samples depending on their maturation stage. We decided to merge samples 7h and 18h IV and samples 8D IV and 12D IV into early stage and late stage groups, respectively. Samples 40h and 72h IV represent the maturation transition points and were grouped into a switch category (Figure 2C).

We focused our analysis on regions with a clear dynamic behavior. Hence, we applied several filtering criteria to select regions that at early maturation phase were active and become inactivated at late stages and vice versa, terming those decrease and increase regions, respectively. Additionally, we identified regions which present their maximum in the switch period and we termed those switch regions (Figure 3A). We observed that dynamic regions presented a gradual activation or inactivation with intermediate levels during the switch period. Nonetheless, in addition to the 330 decreased regions and the 1330 increased regions, we could identify 170 regions that were significantly higher during the switch (examples in Figure 3C-E).

As seen from published transcriptomic studies[12–14], genes annotated to decreased regions were associated to cell morphogenesis involved in differentiation. Additionally, we found a significant enrichment of genes associated with alternative splicing and EGF signaling (Figure 3B). Also expected was the association of late increased regions to genes involved in synapse activity and calcium response. However, little information is described on the genetic processes in the transition between these two phases. We observed that regions activated during the switch were enriched in genes involved both in cell morphogenesis, locomotion and adhesion but also synaptic activity. This finding concurs with the hypothesis that synaptic gene expression precedes protein expression of synapse markers and the onset of spiking activity[18]. Taken together, our data represent a map of biological relevant dynamic active regions during neuronal maturation.

Analysis of transcription factor signatures in dynamics regions

We next analyzed the sequences of the different dynamic regions searching for DNA motifs. For each dynamic group several significantly enriched motifs were found (Figure 4A). Among the decreased regions, many of the motifs found belong to transcription factor (TF) families like NFI, Sox and NeuroD, many members of which areinvolved both in neural progenitor regulation and initiation of neurogenesis[19–21]. On the other hand, at late increased regions, we found TFs known to regulate synapse activity such as the MEF and STAT families[22,23]. The Rfx motif has been previously described to be enriched at forebrain tissue although without presenting changes through neuronal

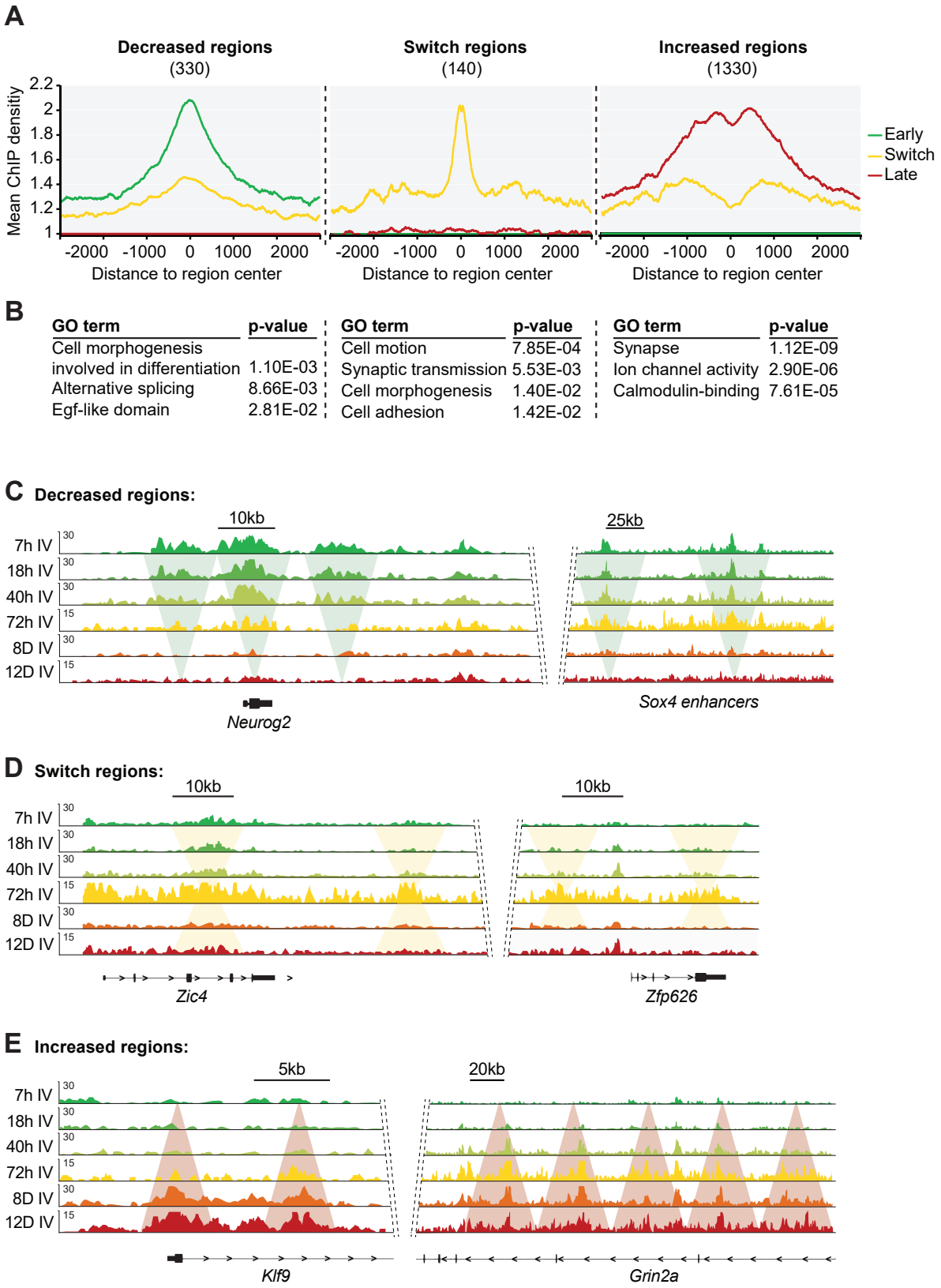


Figure 3 | Genome wide identification of dynamic epigenetic regions

(A) Meta-enhancer profiles of H3K27ac Chip-seq data at the different dynamic subsets. (B) GO analysis on genes annotated to each dynamic subset together with their associated p-value (Benjamini corrected). (C-D) Chip-seq tracks of histone modification H3K27ac across the 6 time points at loci with decreased (A), switch (B) or increased (C) dynamics. Examples of dynamic regions are highlighted.

maturation[24]. Intriguingly, the most significant motifs for the switch regions were for factors not documented to work on neuronal transcription such as Spi-B and the hepatic nuclear factor (HNF) family. In addition, we also found the motif of Pou2f family of transcription factors and for HMG-box transcription factor 1 (Hbp1), which have been associated to neuronal function[25,26].

In order to validate the biological relevance of our findings, we proceeded to evaluate the expression of candidate transcription factors to bind these motifs during *in vivo* brain mouse development (Figure 4B). We retrieved the genome expression data from the Allen developing mouse brain atlas[27] focusing on telencephalic vesicle development and time-points comparable to the *in vitro* development steps in our cultures (Figure 1B). We selected factors binding to the identified DNA motifs that presented clear dynamic expression and grouped them by the dynamic region they were found in.

Regarding decreased region TF candidates, we could confirm the downregulation of NeuroD1,6 and Sox4,5,11 all known to play around in the first steps of neurogenesis. Interestingly, NFI factors, which play a crucial role during the progenitor phase, remain active throughout the whole process of maturation. Strikingly, factors binding the forkhead FOXO motif behave reversely from expected as they have a clear activation towards the end of neuronal maturation. On the other hand, the expression of TF identified in late increased regions followed overall an upregulation at late stages. Not only known synaptic regulators such as MEF and STAT families correlated with the time of expression but also new identified factors such as the Mybl and Tcfap families and Mafk. Finally, we found that most of the switch TF candidates had a transient expression or were peaking halfway through maturation, suggesting a real function for factors with unknown relation to neurons. Hence, we conclude that our list of TF candidates have a potential biological role at different steps of neuronal maturation.

DISCUSSION

Next-generation sequencing approaches provide an exceptional opportunity for understanding brain function and development but also raise unique limitations because of the complexity of the nervous system. One significantly neglected field in neurodevelopment has been the neuronal maturation phase during neurogenesis. Nevertheless, some groups have provided gene expression profiles during the neuronal development either *in vivo* or in *in vitro* culture systems revealing a major switch in neuron behavior since they are born until they integrate to signaling networks[28].

Here, we have complemented those studies trying to decipher which transcription factors govern the epigenetic phases of neuronal maturation. We have chosen an *in vitro* culture of primary hippocampal neurons as it has proven to provide homogenous amounts of pyramidal neurons that re-synchronize upon isolation[15,29]. These two characteristics are fundamental for the study of distinctive cell state events.

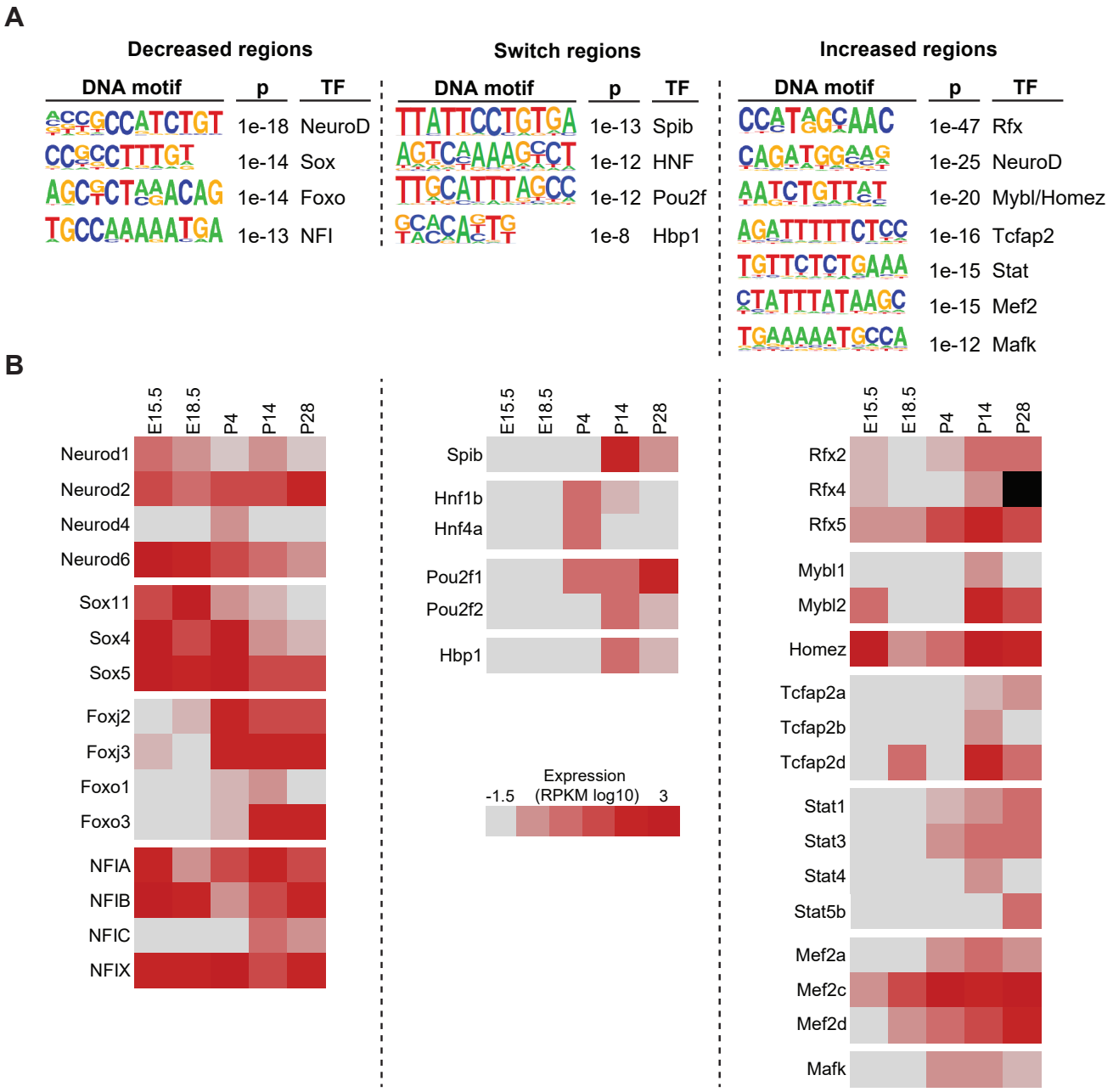


Figure 4 | Identification of candidate transcription factors involved in neural maturation
(A) Significant enriched motifs found at each dynamic subset, associated p-value and transcription factor associated to them. (B) Normalized RNA expression levels of candidate transcription factors though different time points of telencephalon in vivo maturation (Extracted from Allen Developing Mouse Brain Atlas 2013). Missing data is represented by black boxes.

By using ChIP-seq of H3K27ac, our approach provides a catalog of dynamic active regions that change during the course of neuronal maturation. To our knowledge this is the first study identifying open sites throughout the development of neurons. Our data offers specific single-base resolution footprints for TFs candidates to have pivotal roles at different times of neuronal maturation. Combined with *in vivo* expression we found an auspicious subset of transcription factors regulating each time period.

Nevertheless, several limitations in our approach should not be obviated. First, DNA motif identification provides a prediction of described or predicted DNA sequences associated to known factors. Hence, it is possible that unidentified factors could have been missed in our screen. Also, there is the possibility that several factors bind similar motifs but only one is represented. In order to validate our findings we have recurred to *in vivo* expression data with the assumption that the expression levels of candidate TFs would correlate with the time where their motifs are found. However, TF activity can be regulated not only at the transcriptional level but by other processes such as repressor competition or post-translational modification.

Despite these limitations, we have been able to identify TFs following both expression and binding patterns. Further research is needed to confirm the relevance of our findings and to characterize the exact biological role of these candidates. High-throughput screens would provide an excellent tool not only to corroborate our candidate list but also to find other factors involved in maturation with other functions rather than DNA binding. For example, Sharma et al. introduced in 2013 a platform to screen for synaptogenic genes in hippocampus *in vitro* culture evaluating 800 different proteins by shRNA[30]. We propose that the combination of similar screening techniques with genome-wide studies would represent a big step towards an integrated systems biology view of neurodevelopment and would take us closer to understand the vast complexity of the brain.

MATERIAL AND METHODS

Mouse hippocampal primary neuron culture

The morning of identification of vaginal plug was assigned as the first day of gestation, embryonic day E0.5. Fetal E16.5 mouse brains (FVB strain) were dissected in PBS containing 2% glucose and the hippocampi isolated. After trypsin (Invitrogen) and DNAse treatment (Roche Diagnostics), tissue pieces were dissociated, and cells were seeded onto 0.5 mg ml−1 poly-D-lysine (Sigma-Aldrich)-coated coverslips (for immunocytochemistry) in a density of $0.5x10^6$ cells/cm^2 or 100-mm plates (for MNase and ChIP) in a density of $1.5x10^6$ cells/cm^2 in neurobasal medium (Gibco) containing 2 mMglutamax, 120 μg/ml Penicillin, 200 μg/ml Streptomycin and B27 supplement (Invitrogen), and were maintained at 37°C in the presence of 5% CO2.

Immunocytochemistry

Cells were grown on poly-D-lysine–coated cover slips and fixed in 4% formaldehyde for 15min at room temperature. After fixation, they were permeabilized with Triton X-100 in PBS and blocked with blocking buffer (10% FCS, 0.2 M glycine, 0.1% Triton X-100 and 0.05% Deoxycholate in PBS-2% gelatin) for 1 h at room temperature. To label the cells, the following antibodies or dyes were used: anti-βIII Tubulin (Tuj1, Sigma) in blocking buffer for 2 h and with the corresponding secondary antibodies labeled with fluorochromes (Alexafluor 488, Invitrogen). Samples were mounted in MOWIOL (#324590 Sigma-Aldrich) and nuclei were stained with DAPI (Vector laboratories).

Images were captured by confocal microscopy using a Leica SP5 microscope. Image processing was performed using FIJI (ImageJ)[31].

Measurements of chromatin repeat length

We followed Woodcock's protocol as in[32]. Briefly, neural nuclei were digested with a dilution series of micrococcal nuclease (Worthington) at 37°C. Electrophoretic resolution was optimal with digestion times of 120s. Chromatin digests were extracted and the DNA fragments were run in electrophoresis 4%Tris-Glycine agarose gels at pH 7.8 (40 mMTris acetate, 20 mM sodium acetate, 2 mM EDTA) for 2.5 h at a constant current of 100 mA. Visualization of the bands was performed by ethidium bromide and images captured after exposure with Typhoon FLA 9500 scanner (GE Healthcare Life Sciences). Calculation of NRL was performed in Adobe photoshop.

Chromatin immunoprecipitation combined with high-throughput sequencing (ChIP-seq)

For H3K27ac histone modification, we adapted the standard "UPSTATE" protocol. Briefly, around 5-8x10^6 cells were used per condition. Crosslinking was performed "in plate" adding 1% final concentration of formaldehyde. Cells were harvested and 4 times nuclei lysis buffer (1% SDS + 50 mM Tris-HCL(pH8.1) + 10 mM EDTA (pH8.0) was added with a 10min incubation on ice. A Bioruptor Pico sonication device (Diagenode Cat# B01060001) was used for the sonication step. Chromatin concentration was measured, from a 1% of total sample, to equalize DNA content in each condition. After antibody titration (data not shown) 0.75µg H3K27ac antibody (Abcam) was used together with 30µl of Dynabeads Protein G magnetic beads (10003D, Invitrogen). Bound complexes were eluted from the beads in 50 mM Tris-HCl, pH 8.0, 10 mM EDTA and 1% SDS by heating at 65°C for 1 hr with occasional vortexing and crosslinking was reversed by overnight incubation at 65°C. ChIP-seq sample preparation and sequencing was performed as in Chapter 2 of this thesis.

ChIP-seq data analysis

Most of the ChIP-seq analysis, after the first processing steps similar as in Chapter2 of this thesis, were performed using HOMER[33] and BEDTOOLS[34]. Briefly, an input ChIP-seq dataset collected from 18 hours *in vitro* was used as background control to identify significant peaks for each condition. After principal component analysis we decided to group samples in duos (early, switch and late). For every peak identified in any of the 6 conditions, we quantified the coverage of each grouped duos and compare it to the other two. We applied an algorithm to detect regions that were at least 3 fold higher from early to late or vice versa, or 3 fold higher from switch to early and late thus identifying enriched dynamic regions. After each list was completed, we proceed with analysis already described in Chapter 2 such as Gene Ontology cluster annotation, meta-peak profiles and motif recognition (using only known HOMER motifs).

In vivo expression annotation

We retrieved genome expression data from the Allen developing mouse brain atlas[27] focusing on telencephalic vesicle development and time-points comparable to the *in vitro* development steps in our cultures. We selected factors binding to the identified DNA motifs that presented clear dynamic expression and grouped them by the dynamic region they were found in.

REFERENCES

1. Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).

2. Yang, Q., Ke, Y., Luo, J. & Tang, Y. Protocol for culturing low density pure rat hippocampal neurons supported by mature mixed neuron cultures. *J. Neurosci. Methods* **277,** 38–45 (2017).

3. Yan Liu, Huisheng Liu, Conall Sauvey, Lin Yao, Ewa D Zarnowska, and S.-C. Z. Directed differentiation of forebrain GABA interneurons from human pluripotent stem cells. *Nat Protoc* **9,** 1670–1679 (2013).

4. Llorens-Bobadilla, E. *et al.* Single-Cell Transcriptomics Reveals a Population of Dormant Neural Stem Cells that Become Activated upon Brain Injury. *Cell Stem Cell* **17,** 329–340 (2015).

5. Hu, P. *et al.* Dissecting Cell-Type Composition and Activity-Dependent Transcriptional State in Mammalian Brains by Massively Parallel Single-Nucleus RNA-Seq. *Mol. Cell* **68,** 1006–1015.e7 (2017).

6. Yu, D. X., Marchetto, M. C. & Gage, F. H. How to make a hippocampal dentate gyrus granule neuron. *Development* **141,** 2366–2375 (2014).

7. West, A. E. & Greenberg, M. E. Neuronal Activity – Regulated Gene Transcription in Synapse Development and Cognitive Function. 1–21 (2011). doi:10.1101/cshperspect.a005744

8. Malik, A. N. *et al.* Genome-wide identification and characterization of functional neuronal activity–dependent enhancers. *Nat. Neurosci.* **17,** 1330–1339 (2014).

9. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23,** 185–188 (1999).

10. Kishi, N. & Macklis, J. D. MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. *Mol. Cell. Neurosci.* **27,** 306–321 (2004).

11. Shin, J., Ming, G. L. & Song, H. By hook or by crook: Multifaceted DNA-binding properties of MeCP2. *Cell* **152,** 940–942 (2013).

12. Mody, M. *et al.* Genome-wide gene expression profiles of the developing mouse hippocampus. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 8862–7 (2001).

13. Dabrowski, M. *et al.* Gene profiling of hippocampal neuronal culture. *J. Neurochem.* **85,** 1279–1288 (2003).

14. van Spronsen, M. *et al.* Developmental and Activity-Dependent miRNA Expression Profiling in Primary Hippocampal Neuron Cultures. *PLoS One* **8,** (2013).

15. Dotti, C. G., Sullivan, C. a & Banker, G. a. The establishment of polarity by hippocampal neurons in culture. *J. Neurosci.* **8,** 1454–1468 (1988).

16. Wjaeger, A. & Kuenzle, C. C. The chromatin repeat length of brain cortex and cerebellar neurons changes concomitant with terminal differentiation. *EMBO J.* **1,** 811–816 (1982).

17. Ausió, J., de Paz, A. M. artíne. & Esteller, M. MeCP2: the long trip from a chromatin protein to neurological disorders. *Trends Mol. Med.* **20,** 487–498 (2014).

18. Valor, L. M., Charlesworth, P., Humphreys, L., Anderson, C. N. G. & Grant, S. G. N. Network activity-independent coordinated gene expression program for synapse assembly. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 4658–63 (2007).

19. Martynoga, B. *et al.* Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. *Genes Dev.* **27,** 1769–1786 (2013).

20. Pataskar, A. *et al.* NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *EMBO J.* **35,** 24–45 (2016).

21. Baxter, K. K., Uittenbogaard, M., Yoon, J. & Chiaramello, A. The Neurogenic Basic Helix-Loop-Helix Transcription Factor NeuroD6 Concomitantly Increases Mitochondrial mass and Regulates Cytoskeletal Organization in the Early Stages of Neuronal Differentiation. *ASN Neuro* **1,** AN20090036 (2009).

22. Flavell, S. W. *et al.* Genome-Wide Analysis of MEF2 Transcriptional Program Reveals Synaptic Target Genes and Neuronal Activity-Dependent Polyadenylation Site Selection. *Neuron* **60,** 1022–1038 (2008).

23. McGregor, G., Irving, A. J. & Harvey, J. Canonical JAK-STAT signaling is pivotal for long-term depression at adult hippocampal temporoammonic-CA1 synapses. *FASEB J.* **31,** 3449–3466 (2017).

24. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457,** 854–858 (2009).

25. Taguchi, K. *et al.* Identification of hippocampus-related candidate genes for Alzheimer's disease. *Ann. Neurol.* **57,** 585–588 (2005).

26. Watanabe, N., Kageyama, R. & Ohtsuka, T. Hbp1 regulates the timing of neuronal differentiation during cortical development by controlling cell cycle progression. *Development* **142,** 2278–2290 (2015).

27. 2017 Allen Institute for Brain Science. Allen Mouse Developing Brain Atlas. Available from: http://developingmouse.brain-map.org/.

28. Valor, L. M. & Barco, A. Hippocampal gene profiling: Toward a systems biology of the hippocampus. *Hippocampus* **22,** 929–941 (2012).

29. Fletcher, T. L. & Banker, G. A. The establishment of polarity by hippocampal neurons: The relationship between the stage of a cell's development in situ and its subsequent development in culture. *Dev. Biol.* **136,** 446–454 (1989).

30. Sharma, K. *et al.* High-throughput genetic screen for synaptogenic factors: identification of LRP6 as critical for excitatory synapse development. *Cell Rep.* **5,** 1330–41 (2013).

31. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9,** 676–682 (2012).

32. Shim, E. Y., Woodcock, C. & Zaret, K. S. Nucleosome positioning by the winged helix transcription factor HNF3. *Genes Dev.* **12,** 5–10 (1998).

33. Heinz, S. *et al.* Finding Enriched Peaks, Regions, and Transcripts (findPeaks). (2017).

34. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–2 (2010).

Chapter 5

GENERAL DISCUSSION

Study of transcription through the neural lineage

The scope of this thesis has been the study of one of the most fundamental determinants of life, gene expression control. We have studied transcription at several stages of development of the neural lineage. From the early neural induction events described in **Chapter 3**; to the more biochemistry focus studies in neural progenitors in **Chapter 2**; to the epigenetic characterization of maturating neurons in **Chapter 4**; we have covered different sides of transcription of a number of cellular states during neuronal development.

Briefly, **Chapter 2** contributes to the description of the uncharted core transcriptional network that dominates neural stem cells. In **Chapter 3** we propose that Cggbp1 is not only a developmental transcription factor, but that it may act at promoters by a newly described mechanism. Finally, in **Chapter 4**, we have studied the phenomenon of cell maturation, a process that potentially all somatic cells undergo before they become functional, which is very important for neurons but surprisingly perhaps understudied. In neurons this process takes enough time to be able to study step by step reveling epigenetic switches that may relate to other systems. In other words, our endeavor to combine state of the art biochemistry and molecular cell techniques with the study of neural development has provided notable contributions to the general understanding of how transcription is regulated but also new factors involved in neurodevelopment.

Through our studies we have defined several results that we found remarkable to discuss are explore further in this last section of my thesis.

Mediator as a tool for discovery

Since its initial discovery and the first purifications of the complex in the 90s, Mediator has proven to be a central piece not only for RNApol2 regulation but for nearly all aspects of gene transcription in eukaryotes[1]. Hence, the Mediator complex represents an excellent conduit in order to characterize the general transcriptional state of the cell. On one hand, Mediator purifications provide a wide list of chromatin factors acting both at enhancers and promoters. For example, we have identified several cofactors such as Carm1 and Jmjd1c as components of enhancer regulation in neural stem cells. In addition, we have identified Nfia/b, Sox2 and Tcf4/12 as candidates to be the core transcription factor network regulating this cell system. On the other hand, Mediator occupancy reveals active regions, super enhancers among them, which may define the cellular state. For example, we have identified a list of 450 genes candidates to define neural stem cell identity using Mediator ChIP-seq.

We allege that Mediator-based screens could be extended to all eukaryotic model systems in order to identify the key transcriptional components in each situation.

Several innovations could be added to increase the efficiency and clarity of this approach such as the development of models with endogenously tagged versions of Mediator subunits. We envision an example where CRISPR technology[2] could be used to insert FLAG-V5 tag sequences at C- or N-terminal domains of one or several Mediator subunits in mouse embryonic stem cells or directly

injecting blastocysts[3] with the aim to create transgenic Mediator-tagged mice. With this tool, the characterization of any cell type would be achievable implementing efficient protocols of cell isolation or cell culture differentiation. This approach would differ from other methods in delivering functional core chromatin networks instead of full proteomics or transcriptomic lists.

Super enhancers

In **Chapter 2** of this thesis, we have described our work around super enhancers (for brief explanation see **Chapter1 Box1**), mostly due to our focus on new Mediator-based chromatin regulation processes. We have suggested that the Mediator complex is not only a marker but an organizer for assembling super enhancers.

As mentioned, despite the variation in number and genomic assignation that can be explained by different empirical methods used; LCRs, COREs, stretch enhancers and super enhancers fall in the same conceptual and functional category. Nevertheless, despite a redundant description of a chromatin feature already found by other groups, the super enhancer term has reached high citation scores (a metascience topic on itself, Figure 1).

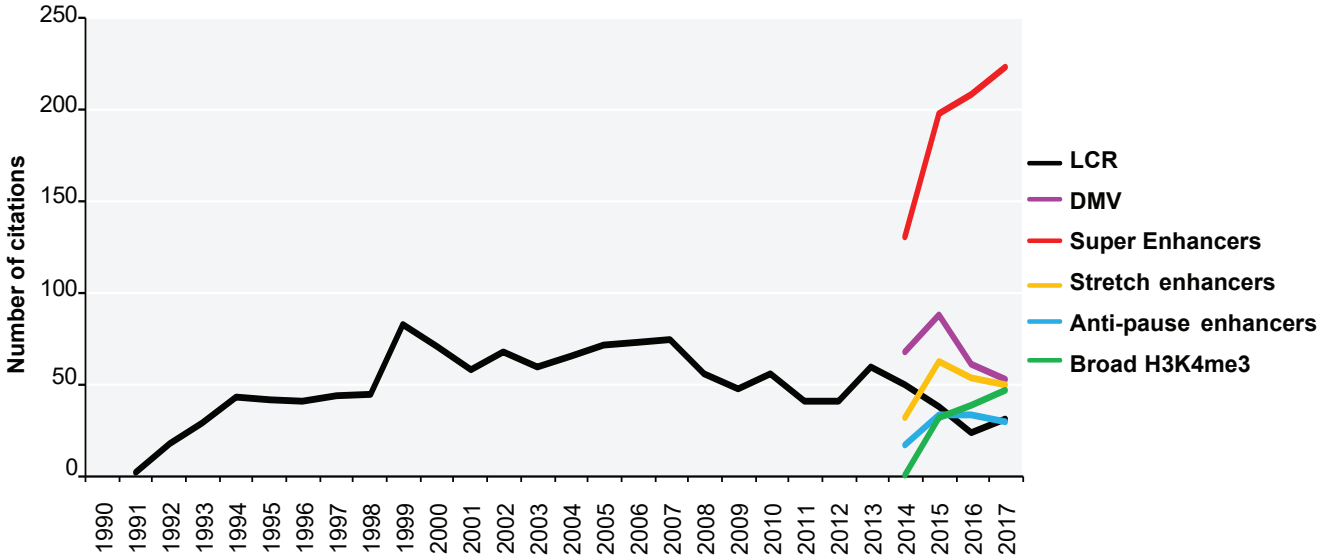


Figure 1 | Citation graph of representative studies in transcription domains

Since the rose of the super enhancer term there has been a debate about if they represent a mechanistically novel regulatory archetype or just a sum of features seen in classical enhancers[4]. Some studies suggested that in fact super enhancers (and synonyms) had specific properties. For example, they present specific enrichments for certain DNA motifs[5]. Another statement was that due to the significant higher accumulation of transcription factors and chromatin remodelers, they are more susceptible to perturbation representing potential therapeutic targets. For example they were claimed to be significantly more affected by inhibitors of Brd4, which are already in use to target super enhancer activated oncogenes such as *Myc* in leukemia[6].

However, a new study from Ameres and Zuber's labs, published during the preparation of this thesis manuscript, rebuts most of these claims. In an elegant approach combining fast degradation/inhibition of chromatin remodelers with newly synthesized mRNA labeling (SLAM-seq), the authors investigated the selective effects of Brd4 depletion. Rather than affecting the chromatin binding of factors such as Mediator or Cdk9, Brd4 inhibition led to a marked stall of RNA Pol II at TSS (including Ser5 phosphorylated forms). The usage of lower, more physiological doses of inhibitor than the used in previous studies revealed a selective inhibition in a subset of targets that correlated poorly with super enhancers. Combing published ChIP-seq data of chromatin binders it was shown that the effects on Brd4 depletion depend on locus specific regulators rather than a unifying chromatin domain[7].

Nevertheless, an accurate evaluation of super enhancer constituents by targeting either specific *trans* or *cis* components will determine their real separation from main enhancer characteristics. A recent study using CRISPR/Cas9-mediated DNA deletions has focused on evaluating super enhancers[8]. The results are still controversial due to high variation in the effects of deleting enhancer clusters. Interestingly, super enhancers did not present an increased regulatory activity compared to enhancers. However, partial redundancy between super enhancer components was observed, suggesting cooperativity between members of the same cluster[8].

About transcriptional factories and phase-separated transcriptional control

In the introductory **Chapter 1** we have seen that chromatin organization is an essential innovation in evolution not only to solve the space limitation that long genomes present but also as a regulatory mechanism to separate different functional domains. Indeed, a highly supported model suggests that nuclear transcription does not occur aimlessly throughout the nucleus but in distinct regions, termed transcription factories[9,10]. These structures would consist in localized domains where multiple active RNA polymerases and other chromatin factors are coordinated together, increasing the temporal and spatial availability of molecules involved in gene expression control[11]. Genes from the same or from different chromosomes associate with the same factory by looping into it[12]. Although representing a promising theory, it is not devoid of controversy, especially because the research of these structures requires innovative approaches to study their composition, assembly and mechanics[13].

Encouragingly, some studies have successfully adopted the concept of transcription factory in order to describe their particular working model. Such is the case of the β-globin locus control region (LCR), where a evolutionary conserved cluster of *cis*- and *trans*-regulatory DNA elements form an active chromatin hub (ACH) involved in the regulation of β-globin genes[14,15].

Although the first studies described transcription factories as fixed scaffolding structures[11], recent advances in our understanding on biomolecular condensates may switch our concept of transcription factories to liquid–like phase separated nuclear compartments driven by macromolecular interactions[16]. Indeed, the application of thermodynamic laws of phase separation systems has been shown to cover many features associated with transcriptional control, from the formation of ACHs or

super enhancers, to the explanation of the transcriptional bursting pattern of enhancers or even the selective vulnerability to drugs[17].

An essential characteristic of biomolecular condensates is the enrichment in multivalent molecules, which provide multiple intra- or inter-molecular interactions[16]. Of course, central components in phase separation transcription would be RNA and DNA molecules harboring multiple *cis*-elements to recruit proteins. However, proteins composed of multiple modular interaction domains represent perfect affinity-scaffolds to promote the formation of large macro-complexes, reducing the solubility and isolating the condensate[16]. Our discoveries in C**hapter 2** of this thesis point the Mediator complex as a central platform with the ability to coordinate a vast number of interactions, thus providing a highly multivalent component that favors the phase separation of the domain.

While it is still not known if all active genes are entangled in transcription factories, it is believed that transcription factories would form *de novo* as the consequence of transcription of some highly active genes with strong active chromatin hubs, and that these factories, once formed, would be able to attract other genes to them[18]. As we have seen in **Chapter 2**, there is a subset of neural stem cell genes that not only are associated to super enhancers but also are marked with broad H3K4me3 signals into their promoters. We showed that these SE+Broad genes present the highest levels of RNApol2, Mediator and cofactors in addition to transcription factor occupancy. These findings, together with the fact that super enhancers are highly enriched among the most interacting TAD triplets[19], and that broad promoters show enhanced DNA looping interactions with super enhancers[20] makes us postulate that SE+Broad domains may represent the strongest transcription factories of the cell.

Interestingly, in our cell model the core transcription network defining cell identity was regulated by SE+Broad domains. It would be interesting to further investigate a possible hierarchy in the formation of secondary factories associated to the targets of this transcriptional network.

Taking into account that Mediator occupancy is ten orders of magnitude higher in SE+Broad compared to typical genes, we suggest that Mediator purifications provide an excel tool to study transcription factory proteomics. With the combination of better techniques for biomolecular condensates isolation[21,22] and chromatin capture innovations[19] further research should make possible the isolation and study of "frozen" factories revealing the native assembly of interactions between proteins and nucleic acids at these domains.

About Mediator subunit composition

An important characteristic of the Mediator complex is that its multi-subunit structure. While a Mediator core is needed to support the basic PIC assembly and transcription initiation, additional subunits can be lost or added affecting the biological function of the complex[23]. A clear example is the association to the Cdk8 kinase module, which changes the behavior of the complex. Moreover, variations in tail subunits influence the capacity of Mediator to bind recruiting transcription factors, thus making the cell "insensitive" to certain transcriptional responses[24]. Furthermore, with our discoveries in **Chapter 2** of this thesis we could also extend this effect to the recruitment of chromatin

modifiers, which could depend on specific Mediator subunits to be recruited to the DNA. Despite few studies mainly focused on Mediator subunit expression alterations in malignancy[25], the variation in Mediator subunit composition has not been characterized.

To bring some light into this matter we purified and compare the Mediator complex subunit composition in ESCs and NSCs (*ESC data from Dr. Debbie van den Berg unpolished data*, Figure 2A). We immunoprecipitated endogenous Med12 instead of using our overexpressing Med15-Flag lines in order to maintain the physiological levels of all subunits. After identifying the peptide counts and emPAI scores by mass spectrometry, we calculated the relative contribution of each subunit to the complex. From the 30 subunits detected, we observe a big conservation in Mediator composition overall. Nevertheless, we identified several subunits with more than 1.5 fold difference in their relative contribution to the complex (Figure 2B). In order to evaluate if our findings were biologically relevant we extracted the gene expression levels of each subunit from a published study[26]. Strikingly, we observed that most of the subunits with a variable contribution to the complex had a correlating change in expression levels between ESCs and NSCs (Figure 2B).

Interestingly, this implies that although the Mediator is an essential complex for transcription; its subunits do not behave as housekeeping genes, presenting significant changes among different cell types. Second, our findings suggest that Mediator complex composition can be explained in large part by the expression level of its subunits.

Accordingly, we analyzed the mRNA levels of almost all Mediator subunits (33 subunits) across a wide palette of human tissues[27] (Figure 2C). Strikingly, Mediator subunit expressions formed clusters between samples of similar tissues. For example, all neural tissues formed a well defined cluster characterized by high levels of CDK19. Interestingly, we observed most of the protein composition variation from ESCs to NSCs to correlate with these findings, although we have to take into account that neural progenitors differ significantly from adult tissues and the mouse-human extrapolation. CDK19 is the paralog of the cyclin-dependent kinase CDK8 and it has been mutated in patients suffering from microcephaly, congenital retinal folds and intellectual disability[28]. Both our protein purifications in **Chapter 2** and the results above show the neural specificity of CDK19 and an explanation for its nervous system phenotype.

Other well defined clusters were digestive system tissues such as intestines, colon and stomach. Interestingly, CDK8 is associated to colon cancer[29,30] but also presented extremely high levels in healthy sigmoid colon. Reproductive system tissues such as uterus, cervix and fallopian tubes also cluster well together. Other interesting findings were the MED23 defining cluster of heart associated tissues and esophagus suggesting a novel role of this protein in this tissues. Finally, the inclusion of transformed malignant cells in the array reveled MED1 as a highly overexpressed outlier in the two cancer samples; a finding that concords with the already described oncogenic potential of this subunit[31].

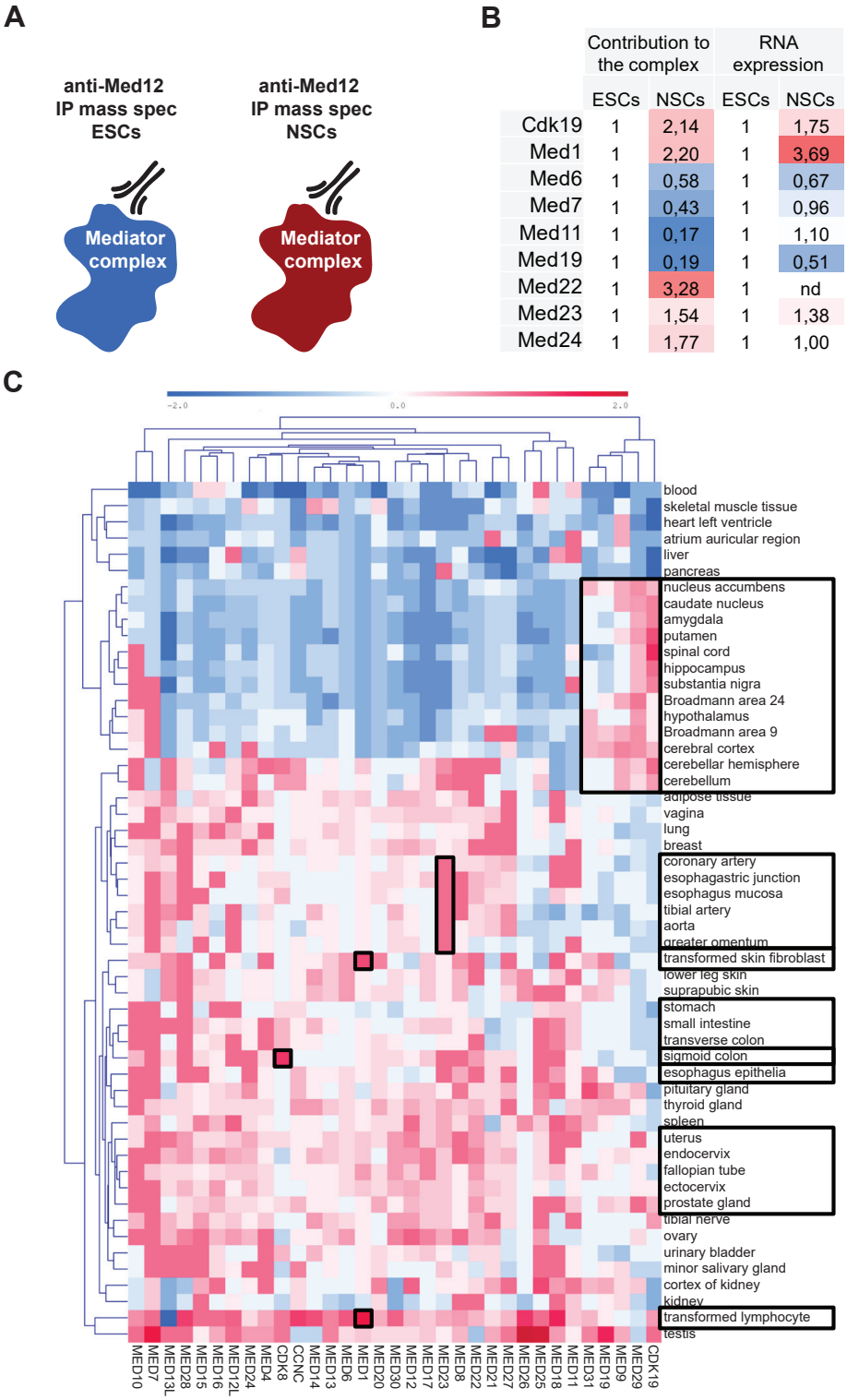


Figure 2 | Identification of Mediator subunit tissue specific variation

A. Schematic representation of Mediator purifications in embryonic stem cells (ESCs) and neural stem cells (NSCs). B. Proteomic subunit contribution to Mediator complex was calculated as the relative emPAI of each subunit normalized to the total emPAI of the complex, calculated as the sum of all subunit emPAIs. Then, ESCs contribution values were used to calculate fold differences against NSCs subunits. Note that this experiment is n=1. Only subunits with more than 1.5 fold difference are presented. RNA expression was extracted from[26]and normalized to ESCs values. C. MeV (Artistic License 2.0) was used to cluster, using Euclidian distance, the normalized Mediator subunits expression profiles along tissues. Expression data extracted from[27].

In summary, our preliminary data suggests that Mediator subunit composition varies significantly across tissues and that part of this variation can be explained by a regulation on the expression levels of several subunits. Nevertheless, little is known about the mechanisms regulating subunit exchange or the determinants of Mediator subunit composition. In relation to the transcriptional regulation of Mediator subunits we observed that not only several subunit of Mediator appear to be poised in ESCs, an indication of a developmental role, but also that both Retinoic acid receptor and Cggbp1, a neural inducer described in **Chapter 3** of this thesis, bind to their promoters representing an example of a direct regulation by a transcription factor (Figure 3).

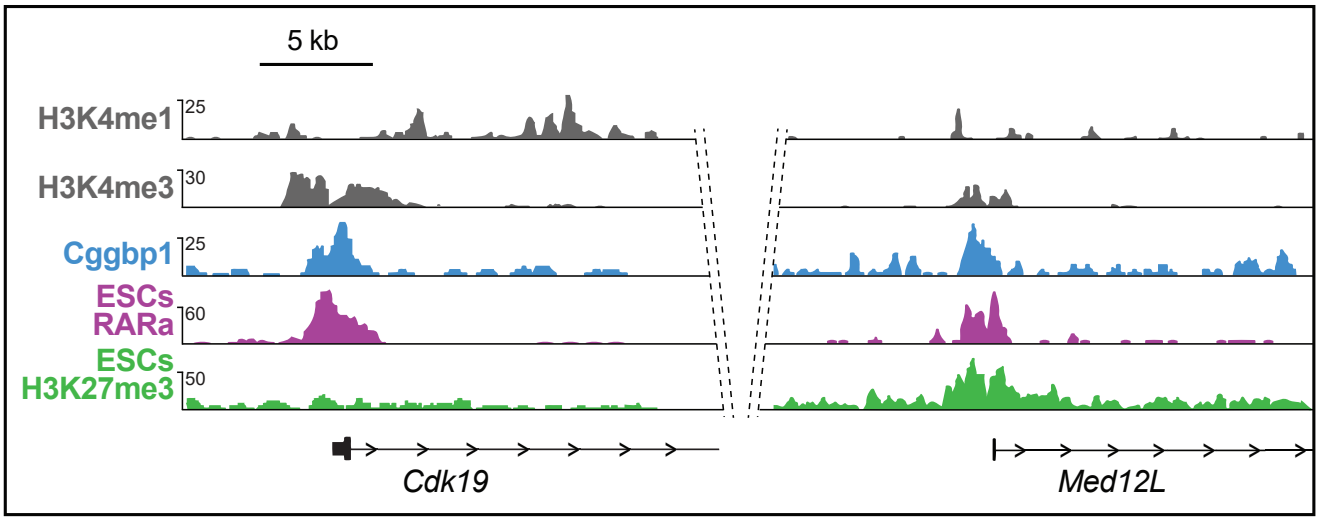


Figure 3 | Cggbp1 binds Mediator subunit promoters
Chip-seq tracks of Cggbp1, histone modifications H3K4me3, H3K4me1 from NSCs and H3K27me3 and RARa from ESCs at Cggbp1 locus. Range of reads per million per base pair is indicated on the y-axis. Scale bar is indicated.

Before ending the discussion on this topic, it is worth mentioning the regulation of Mediator subunit levels post-translationally. As seen in the **Chapter 2** of this thesis, Mediator purifications are rich in protein modifiers. One of the highest hits in Med15-Flag purifications is Trim11, an E3 ubiquitin ligase known to target many neural related proteins such as Humanin and Pax6 for degradation by the proteosome[32]. It is also described that it can target Med15 to the ubiquitin–proteasome degradation system representing a clear example of a Mediator subunit composition regulator[33]. Moreover, deeper analysis of our Med12 protein purifications showed an arginine dimethylation site (data not shown); a post-translation modifications regulated by Carm1/PRMT4, a Mediator interactor identified in **Chapter 2**.

The further characterization of Mediator subunit composition and its regulation represent new layers on top of our expanding knowledge on transcriptional mechanics. Understanding the dynamics and effectors defining Mediator composition will provide new targets in order to alter transcriptional responses defective in disease.

Last but not least, recent advances in the proteomics field allow the study of specific interactions within a complex. For example, chemical cross-linking combined with mass spectrometry is being used to obtain 3-D structural information by detecting peptides that are in close spatial proximity[34]. This approach has been used to characterize Mediator subunit interactions and structure[35]. With advances in the cross-linked peptide-peptide recognizing algorithms, we anticipate an extension of this kind of approaches in order to identify the specific sites within the complex where each interactor is bound, hence providing the exact targets for drug discovery design.

Brd4, the elephant in the room

Bromodomain-containing protein 4 (Brd4) is a member of the eukaryotic BET family that contains two bromodomains (BDI and BDII) and an extraterminal (ET) domain. The bromodomain is a conserved sequence of ~110 amino acids with the ability to bind acetyl-lysine residues in histones and many other proteins[36,37]. Ubiquitously expressed and with an almost 1400 amino acid length and a complex domain structure, this protein constitutes a docking platform for other chromatin regulators to target them to active open regions[38]. For example, Brd4 CTD interacts with the cyclin T1 and Cdk9, subunits of the positive transcription elongation factor b (pTEFb) complex[39] centering the attention on its role in RNApol2 pause-release. Moreover, other interactions such as JMJD6 and NSD3 from Brd4 ET domain have been shown to be important for transcription activation[40].

Recently, its association with acetylated histones has pointed at Brd4 as an enhancer regulator. Brd4 co-occupies with Mediator not only promoters but most enhancers, including super enhancers. Indeed, its relevance has risen exponentially since it has been discovered that the use of inhibitors for Brd4 selectivly disrupted super-enhancers associated to tumor oncogenes[6]. Thus, being part of the same macro regulatory structures in the genome may suggest that some interplay could exist between Brd4 and Mediator. However, the current evidence of a direct association is very controversial. On one hand, several groups have described Brd4-dependent Mediator binding to the DNA and that Brd4 inhibitors displace Mediator from chromatin[6,41]. On the other hand, biochemistry studies focusing on Brd4 purifications have not detected a physical link between these two entities[40] and, as seen in **Chapter 2** of this thesis, our Mediator purifications presented neither Brd4 nor Brd4-specific interactors.

In an attempt to bring some clarity in this controversy, we decided to purify endogenous Brd4 in our NSC system. Our mass spectrometry data showed that we were able to detect almost all Brd4 described interactors (Figure 4) but we detected minimal amounts of Mediator (data not shown). In the need of more experimental replicates we are not able to make a solid statement except that approaches similar to those in **Chapter 2** (flag purification from DNA contamination controlled extracts) would be more suitable. In addition, label-free methods for mass spectrometry quantification such as iBAQ could be applied in order to normalize hit scores in several experiments[42].

We also performed a Mediator ChIP-seq in NSCs in the presence of Brd4 inhibitors, which removed Brd4 at all tested genomic locations. In contrast to results from other labs[43], we found very little effect

on Mediator genome-wide location (data not shown). One conclusion would be that the interaction between Brd4 and Mediator, if existing, is weak and does not appear required to target Mediator to the genome in NSCs. More extensive studies will have to be performed to resolve this puzzle. However, taking into account the phase-separation transcription model we could hypothesize that while not directly interacting with each other, Mediator and Brd4 are major multivalent complexes coordinating many components of the domain, thus interacting in an indirect manner. The removal of one of the two would create a great disturbance in the system leading at some point to the disruption of the other.

Due to the privileged position of both Mediator and Brd4 at (super)enhancers and promoters, we argue that the construction of a combined protein network centered on these two entities would help to understand the mechanics of transcription assembly. In a preliminary approach, we have depicted the most significant chromatin complexes found in our Med12 and Brd4 purifications and the strength of their interaction (Figure 4).

Our results show that while many complexes are shared between the two, other chromatin players appear to be highly specific for either Mediator or Brd4. As discussed above, cleaner approaches would narrow this overlap and provide new mechanistic insights in enhancer-promoter assembly.

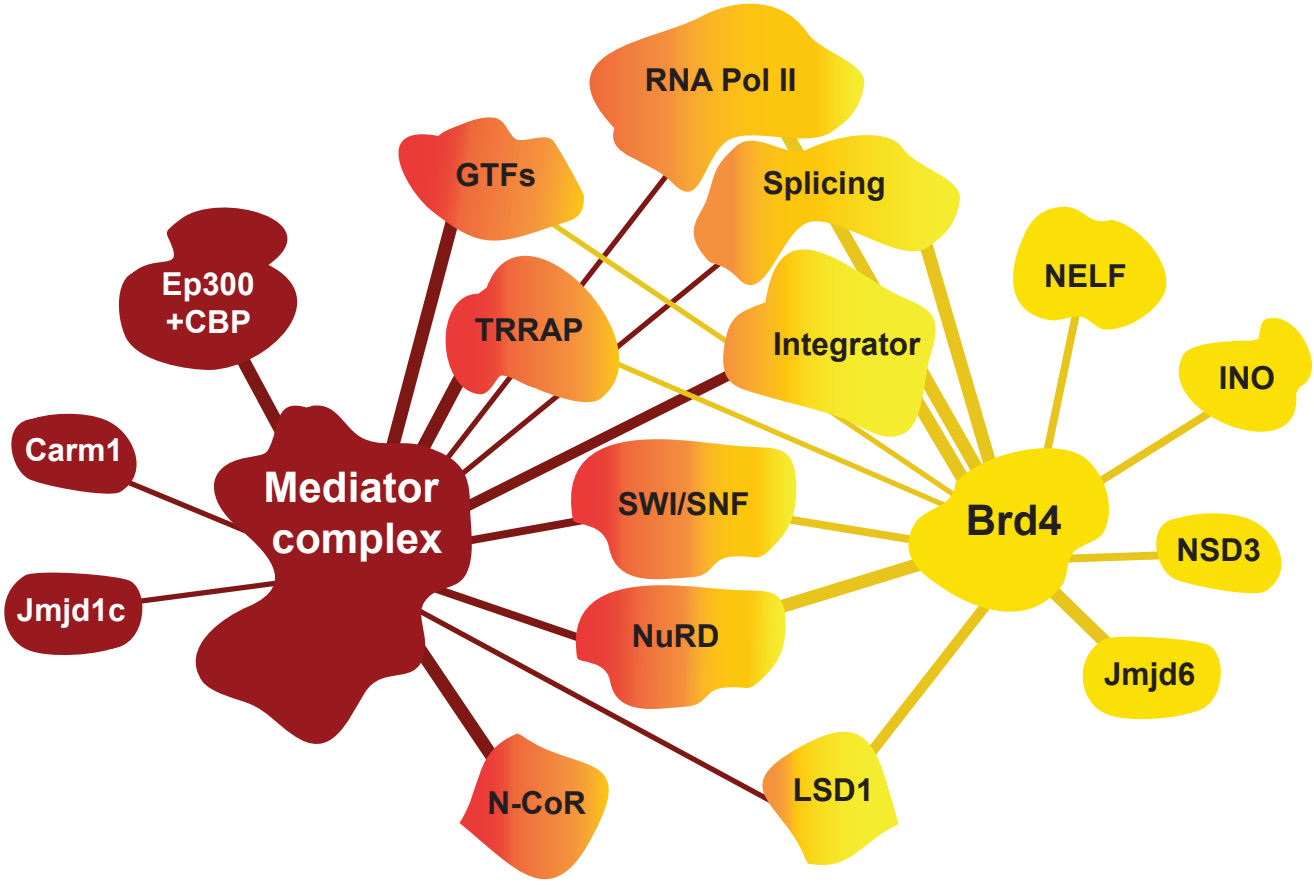


Figure 4 | Summary of Brd4-Mediator proteome in neural stem cells.
Thickness of edges represents abundance (emPAI) corresponding to each purification. Note n=1.

About Cggbp1, an unfinished business

In **Chapter 3** of this thesis we have characterized Cggbp1 as a transcription factor regulating the transition from pluripotency to the neural lineage and neural stem cell homeostasis. As mention above, the identification of Cggbp1 constitutes another example on how Mediator-based screens could provide important transcription regulation candidates to study.

Contemplating our work in progress, several aspects remain to be solved. First of all, we would like to properly characterize the expression of Cggbp1 *in vivo* as the only existing data does not include early developmental time points[44]. Our discovery of a direct retinoic acid regulation make us hypothesize that Cggbp1 could be involved in other RA-dependent developmental processes occurring later in development such as lung or limb formation[45].

In relation to that, a better characterization of the protein levels across tissues would give a better insight in the activity of the protein. Although Cggbp1 is described as ubiquitously expressed, its protein levels and nuclear localization depend on post-translation modifications produced by several signal cascades[46]. In particular, the EGF pathway, which is an essential growth signal to sustain NSCs, has been shown to phosphorylate Cggbp1 at tyrosine Y20 favoring its localization to the nucleus[46]. Hence, Cggbp1 activity could be restricted in different tissues due to a wide variety of regulatory steps (Figure 5A). Nonetheless, our results show a preference for the neural lineage justified in part, by the biased localization of the GCC motif in neural targets.

Although we have provided evidence pointing to an activation role for this transcription factor, we have also found some signs that Cggbp1 may be also a repressor, the clearest example being the drastic mesodermal upregulation after Cggbp1 depletion in ESCs. In addition, as shown in **Chapter 3**, the mesodermal transcription factor Goosecoid (Gsc) appears to be bound by Cggbp1 in neural stem cells, where it is repressed. Further work will address the dual activity of Cggbp1 with special focus on developmental decisions (Figure 5B).

Finally, despite the exploratory nature of our results, we recognize the exciting opportunity that Cggbp1 represents in characterizing Polycomb. The mechanisms of Polycomb regulation and recruitment are still very controversial, especially in mammals[47]. We hypothesize that Cggbp1 activity in development is linked to the Polycomb response. At the DNA sequence level, some studies associate large CpG islands depleted of activating factor motifs to Polycomb occupancy in ESCs[48]. We suggest that the GCC repeat of Cggbp1 may fall into these regions and passed unnoticed as a motif. In addition, *Arabidopsis* screens of Polycomb response elements identified GCC repeats as a recruiting *cis* candidate[49].

On top of that, our results show both an overlap of Cggbp1 occupancy with CpG islands and poised promoters. In parallel, our Cggbp1 interactome identified several members of Polycomb as interaction partners of Cggbp1. Finally, Cggbp1 regulates Polycomb subunits involved in the derepression switch of poised genes (Figure 5B).

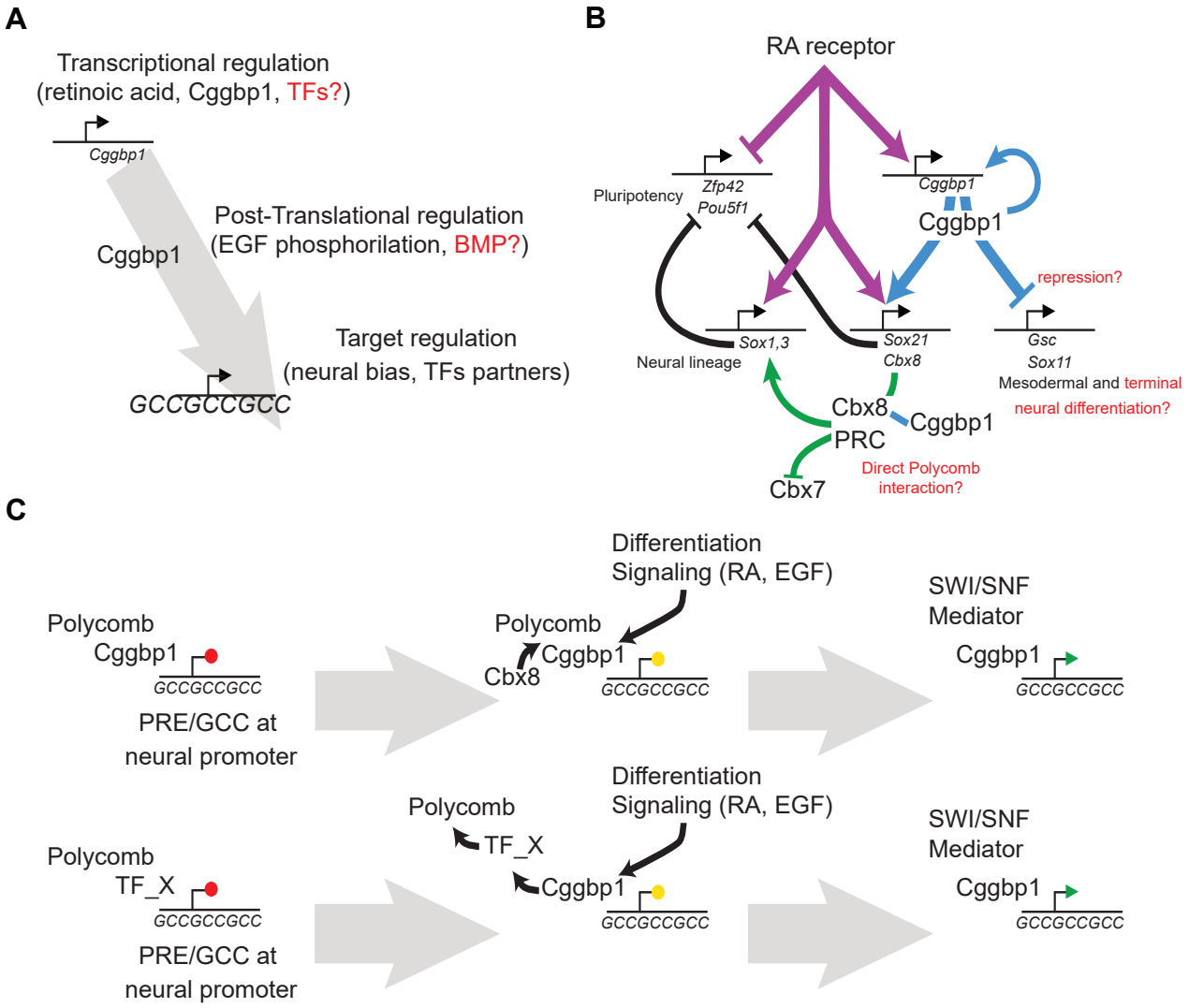


Figure 5 | Cggbp1 and transcription regulation
A. Schematic representation of Cggbp1 regulation. Red indicates speculative ideas. B. Cggbp1 modulation of the retinoic acid (RA) response in ESCs. Red indicates speculative ideas. C. Models of Cggbp1 mediated activation of poised promoters. TF_X refers to a trans-element binding similar sequences as Cggbp1.

Taken together, we proposed two models explaining Cggbp1 regulation of poised promoters of the neural lineage (Figure 5C). The first model takes into account the Cggbp1 presence in ESCs and its Polycomb protein interactions. Signaling could trigger modification of Cggbp1 translating to a switch of interactors from Polycomb to activating complexes. The second model proposes that Cggbp1 activity is only present after differentiation signals and its main role is to displace by competition an unidentified transcription factor from Polycomb response elements. Hence Cggbp1 would indirectly displace Polycomb and recruit the activation machinery.

In a more opportunistic approach, we would like to extend our research of Cggbp1 to chromatin assembly. Despite the vast understanding in transcription factor mediated activation of enhancers, how enhancers recognize specific promoters is still poorly understood. The discovery of new promoter binders such Cggbp1 and the characterization of their mechanism of regulation should provide new insights in transcription assembly.

Neuronal maturation

In addition to the massive success accomplished by the study of individual genes, in the last few decades genome-wide studies have provided new insights into the molecular basis of brain development, neural plasticity, and neurological diseases[50]. Neural maturation is an attractive field of study as discoveries found in one system such as hippocampal development can be easily extrapolated to other neural systems. While early differentiation and late synapse formation appear to be highly extrinsic regulated processes, maturation appears to be more cell-intrinsic[51–53]. Processes such as the nucleosome repeat shortening seem to occur across all neurons independent on their time of birth. We hypothesize that a conserved core transcriptional network must dictate the cell intrinsic response to reach maturity.

In **Chapter 4** of this thesis we have generated a map of the dynamic active chromatin regions during neuronal maturation. We suggest several candidate transcription factors acting in early, switch and late stages. We were particularly interested in the switch phase as the transition between neuronal differentiation transcription factors and synaptic activity-dependent ones had not been explored yet. Interestingly, while many of our candidates are validated by their *in vivo* expression profile, several of them have not been related to nervous system development. However, we have promising hints indicating to relevant physiological role of these candidates.

For example, Spi-B has only been studied in lymphocytes and it has been shown to be a target of Ikaros family zing finger 1 (Ikzf1)[54]. Encouragingly, gene regulation in lymphocytes and neurons share many common players, Tcf4 (E2-2) being one example[55,56]. Indeed, Ikzf1 regulates neural differentiation[57] by activating secondary transcription factors[58]. We hypothesize that Spi-B could be activated by this proneural factor.

Another example is represented by Hepatic nuclear factors (HNF). In contrast to their name, these factors are not restricted to liver development, and new functions continuously appear for members of this transcription family[59,60]. Although we cannot rule out the possibility that other factors share the same motif, both Hnf4a and Hnf1b presented expression profiles correlated to the switch. Interestingly, deletion of the 17q12 locus where Hnf1b localizes causes a syndrome characterized by variable combinations of kidney and urinary tract abnormalities, maturity-onset diabetes and neurodevelopmental or neuropsychiatric disorders[61]. Additionally, the Hnf4a motif has been found significantly enriched at super enhancers of striatal neurons[62] and has been related to depression and brain homeostasis[63]. Further research would consolidate these findings by testing the role in maturation of the identified candidates. For example, in addition to the proposed high-throughput screen (see **Chapter 4**), a more classical approach could be followed by *in vivo* electroporation of mouse embryos with shRNA constructs targeting the candidates and evaluating the neural maturation phenotype.

REFERENCES

1. Ansari, S. A. & Morse, R. H. Mechanisms of Mediator complex action in transcriptional activation. *Cell. Mol. Life Sci.* **70,** 2743–2756 (2013).

2. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (80-.).* **339,** 819–823 (2013).

3. Yang, H. *et al.* One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell* **154,** 1370–1379 (2013).

4. Pott, S. & Lieb, J. D. What are super-enhancers? *Nat. Genet.* **47,** 8–12 (2014).

5. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155,** (2013).

6. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153,** 320–334 (2013).

7. Muhar, M. *et al.* SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. *Science* eaao2793 (2018). doi:10.1126/science.aao2793

8. Moorthy, S. D. *et al.* Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome Res.* **27,** 246–258 (2017).

9. Jackson, D. A., Hassan, A. B., Errington, R. J. & Cook, P. R. Visualization of focal sites of transcription within human nuclei. *EMBO J.* **12,** 1059–65 (1993).

10. Iborra, F. J., Pombo, A., Jackson, D. A. & Cook, P. R. Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei. *J. Cell Sci.* **109,** (1996).

11. Weipoltshammer, K. & Schöfer, C. Morphology of nuclear transcription. *Histochem. Cell Biol.* **145,** 343–58 (2016).

12. Iborra, F. J., Pombo, A., Mcmanus, J., Jackson, D. A. & Cook, P. R. The Topology of Transcription by Immobilized Polymerases. *Exp. Cell Res.* **229,** 167–173 (1996).

13. Sutherland, H. & Bickmore, W. A. Transcription factories: Gene expression in unions? *Nat. Rev. Genet.* **10,** 457–466 (2009).

14. Palstra, R.-J. *et al.* The β-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* **35,** 190–194 (2003).

15. de Laat, W. & Grosveld, F. Spatial organization of gene expression: the active chromatin hub. *Chromosome Res.* **11,** 447–59 (2003).

16. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18,** 285–298 (2017).

17. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169,** 13–23 (2017).

18. Noordermeer, D. *et al.* Transcription and Chromatin Organization of a Housekeeping Gene Cluster Containing an Integrated β-Globin Locus Control Region. *PLoS Genet.* **4,** e1000016 (2008).

19. Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543,** 519–524 (2017).

20. Thibodeau, A., Márquez, E. J., Shin, D.-G., Vera-Licona, P. & Ucar, D. Chromatin interaction networks revealed unique connectivity patterns of broad H3K4me3 domains and super enhancers in 3D chromatin. *Sci. Rep.* **7,** 14466 (2017).

21. Andersen, J. S. *et al.* Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **12,** 1–11 (2002).

22. Ying, Y. *et al.* Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. *Cell* **170,** 312–323.e10 (2017).

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

24. Fondell, J. D., Ge, H. & Roeder, R. G. Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 8329–33 (1996).

25. Syring, I. *et al.* Comprehensive analysis of the transcriptional profile of the Mediator complex across human cancer types. *Oncotarget* **7,** 23043–23055 (2016).

26. Zhu, J. *et al.* Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152,** 642–654 (2013).

27. Abugessaisa, I. *et al.* FANTOM5 transcriptome catalog of cellular states based on Semantic MediaWiki. *Database* **2016,** baw105 (2016).

28. Mukhopadhyay, A. *et al.* CDK19 is disrupted in a female patient with bilateral congenital retinal folds, microcephaly and mild mental retardation. *Hum. Genet.* **128,** 281–291 (2010).

29. Donner, A. J., Szostek, S., Hoover, J. M. & Espinosa, J. M. CDK8 Is a Stimulus-Specific Positive Coregulator of p53 Target Genes. *Mol. Cell* **27,** 121–133 (2007).

30. Firestein, R. *et al.* CDK8 is a colorectal cancer oncogene that regulates β-catenin activity. *Nature* **455,** 547–551 (2008).

31. Schiano, C. *et al.* Involvement of Mediator complex in malignancy. *Biochim. Biophys. Acta - Rev. Cancer* **1845,** 66–83 (2014).

32. Tuoc, T. C. & Stoykova, A. Trim11 modulates the function of neurogenic transcription factor Pax6 through ubiquitin-proteosome system. *Genes Dev.* **22,** 1972–1986 (2008).

33. Ishikawa, H., Tachikawa, H., Miura, Y. & Takahashi, N. TRIM11 binds to and destabilizes a key component of the activator-mediated cofactor complex (ARC105) through the ubiquitin-proteasome system. *FEBS Lett.* **580,** 4784–4792 (2006).

34. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat. Methods* **9,** 904–906 (2012).

35. Cevher, M. A. *et al.* Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nat. Struct. Mol. Biol.* **21,** 1028–1034 (2014).

36. Dey, A., Chitsaz, F., Abbasi, A., Misteli, T. & Ozato, K. The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 8758–63 (2003).

37. Huang, B., Yang, X.-D., Zhou, M.-M., Ozato, K. & Chen, L.-F. Brd4 Coactivates Transcriptional Activation of NF- B via Specific Binding to Acetylated RelA. *Mol. Cell. Biol.* **29,** 1375–1387 (2009).

38. Wang, R., Li, Q., Helfer, C. M., Jiao, J. & You, J. Bromodomain Protein Brd4 Associated with Acetylated Chromatin Is Important for Maintenance of Higher-order Chromatin Structure. *J. Biol. Chem.* **287,** 10738–10752 (2012).

39. Moon, K. J. *et al.* The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell* **19,** 523–534 (2005).

40. Rahman, S. *et al.* The Brd4 Extraterminal Domain Confers Transcription Activation Independent of pTEFb by Recruiting Multiple Proteins, Including NSD3. *Mol. Cell. Biol.* **31,** 2641–2652 (2011).

41. Di Micco, R. *et al.* Control of embryonic stem cell identity by brd4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes. *Cell Rep.* **9,** 234–247 (2014).

42. Arike, L. *et al.* Comparison and applications of label-free absolute proteome quantification methods on Escherichia coli. *J. Proteomics* **75,** 5437–5448 (2012).

43. Bhagwat, A. S. *et al.* BET Bromodomain Inhibition Releases the Mediator Complex from Select cis -Regulatory Elements. *Cell Rep.* **15,** 519–530 (2016).

44. Naumann, F., Remus, R., Schmitz, B. & Doerfler, W. Gene structure and expression of the 5′-(CGG)n-3′-binding protein (CGGBP1). *Genomics* **83,** 106–118 (2004).

45. Maden, M. The role of retinoic acid in embryonic and post-embryonic development. *Proc. Nutr. Soc.* **59,** 65–73 (2000).

46. Agarwal, P. *et al.* Growth signals employ CGGBP1 to suppress transcription of Alu-SINEs. *Cell Cycle* **15,** 1558–1571 (2016).

47. Schuettengruber, B. & Cavalli, G. Recruitment of Polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development* **136,** 3531–3542 (2009).

48. Ku, M. *et al.* Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. *PLoS Genet.* **4,** e1000242 (2008).

49. Xiao, J. *et al.* Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. *Nat. Genet.* (2017). doi:10.1038/ng.3937

50. Valor, L. M. & Barco, A. Hippocampal gene profiling: Toward a systems biology of the hippocampus. *Hippocampus* **22,** 929–941 (2012).

51. Montague, P. R. & Friedlander, M. J. Expression of an intrinsic growth strategy by mammalian retinal neurons. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 7223–7 (1989).

52. Metzger, F. Molecular and cellular control of dendrite maturation during brain development. *Curr. Mol. Pharmacol.* **3,** 1–11 (2010).

53. Sances, S. *et al.* Modeling ALS with motor neurons derived from human induced pluripotent stem cells. *Nat. Neurosci.* **16,** 542–553 (2016).

54. Solomon, L. A., Batista, C. R. & DeKoter, R. P. Lenalidomide modulates gene expression in human ABC-DLBCL cells by regulating IKAROS interaction with an intronic control region of SPIB. *Exp. Hematol.* **56,** 46–57.e1 (2017).

55. Quong, M. W., Romanow, W. J. & Murre, C. E protein function in lymphocyte development. *Annu. Rev. Immunol.* **20,** 301–322 (2002).

56. Sepp, M. *et al.* The Intellectual Disability and Schizophrenia Associated Transcription Factor TCF4 Is Regulated by Neuronal Activity and Protein Kinase A. *J. Neurosci.* **37,** 10516–10527 (2017).

57. Martín-Ibáñez, R. *et al.* Ikaros-1 couples cell cycle arrest of late striatal precursors with neurogenesis of enkephalinergic neurons. *J. Comp. Neurol.* **518,** 329–351 (2010).

58. Martín-Ibáñez, R. *et al.* Helios expression coordinates the development of a subset of striatopallidal medium spiny neurons. *Development* **144,** 1566–1577 (2017).

59. Zhou, H. *et al.* Atypical mouse cerebellar development is caused by ectopic expression of the forkhead box transcription factor HNF-3beta. *Gene Expr.* **9,** 217–36 (2001).

60. Liu, D. *et al.* Targeting the HDAC2/HNF-4A/miR-101b/AMPK Pathway Rescues Tauopathy and Dendritic Abnormalities in Alzheimer's Disease. *Mol. Ther.* **25,** 752–764 (2017).

61. Moreno-De-Luca, D. *et al.* Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *Am. J. Hum. Genet.* **87,** 618–30 (2010).

62. Achour, M. *et al.* Neuronal identity genes regulated by super-enhancers are preferentially down-regulated in the striatum of Huntington's disease mice. *Hum. Mol. Genet.* **24,** 3481–3496 (2015).

63. Yamanishi, K. *et al.* Hepatocyte nuclear factor 4 alpha is a key factor related to depression and physiological homeostasis in the mouse brain. *PLoS One* **10,** e0119021 (2015).

Addendum

SUMMARY

The scope of this thesis has been the study of one of the most fundamental determinants of life, gene expression control by transcription. We have studied transcription at several stages of development of the neuronal lineage.

In **Chapter 1** I have described how the process of transcription regulation is linked to the evolution of life, expanding in complexity as organisms develop more elaborated features. I have reviewed the fundamentals in the transcription process with a special focus on the Mediator complex, a versatile component of the core transcription machinery. In addition I describe the basic transcription regulators involved in mammal brain development representing a model to study chromatin regulation in high eukaryotes.

Chapter 2 describes a transcriptional network that dominates neural stem cells. We have shown that Mediator complex represents a major interaction hub at enhancer-promoter assemblies and that by proteomic approaches it is possible to identify new Mediator interactors involved in transcription regulation at enhancers. Identified transcriptions factors dominate neural stem cell enhancers acting as putative Mediator recruiters. In addition, we have identified neural stem cell genes nearby super enhancers, which are prone to define cell identity, and showed that they are associated with high density of Mediator and its interactors. Combining new concepts such as "broad H3K4me3 promoters" and super enhancers we have identified a core set of highly expressed genes where Mediator may play an essential role coordinating interactions and stabilizing their formation. Our data adds evidence to the concept of phase-separated chromatin domains, where high density of chromatin regulators would form biocondensates in order to efficiently regulate transcription at certain loci.

In **Chapter 3** we explore the functions of Cggbp1, one of the transcription factors identified in our Mediator purifications. We showed that Cggbp1 is a transcription activator acting at promotersof neural stem cell genes. Moreover, combining bioinformatics approaches with published data we suggest a role for Cggbp1 in early neural induction downstream of retinoic acid signaling.

From the neural stem cell model in **Chapter 2** we moved to embryonic stem cells in **Chapter 3** in order to study early neural lineage developmental events. Fittingly, **Chapter 4** of this thesis covers the last steps of neuronal development, neuronal maturation. By assessing active chromatin regions across several points in the development and maturation of ex-vivo hippocampal neurons we map regulatory regions involved in neuronal maturation. We identified candidate transcription factors that may regulate subsequent maturation steps by their binding motifs at active chromatin regions and their expression profiles during brain development.

In summary, this thesis contains a collection of novel insights into the regulation of transcription in the neural lineage. In **Chapter 5** I discuss several implications of the experiments described here and present preliminary work to support new hypothesess in future research.

SAMENVATTING

De scope van dit proefschrift is de studie van één van de meest fundamentele eigenschappen van leven, de controle van genexpressie door transcriptie. We hebben transcriptie bestudeert in verschillende stadia van ontwikkeling van de neuronale lijn.

In **hoofdstuk 1** beschrijf ik hoe het proces van transcriptionele regulatie is verbonden met de evolutie van leven en complexer wordt wanneer organismen meer uitgebreide onwikkelingskenmerken krijgen. Ik bespreek fundamentele aspecten van het transcriptieproces met speciale aandacht voor het Mediator complex, een veelzijdige component van de centrale transcriptiemachinerie. Tevens beschrijf ik de transcriptionele regulators betrokken bij zoogdier-hersenontwikkeling als een model om chromatine regulators te bestuderen in hogere eukaryoten.

Hoofdstuk 2 beschrijft een transcriptioneel netwerk wat neurale stamcellen domineert. We laten zien dat Mediator een significante interactie hub is voor promoter-enhancer complexen en dat het mogelijk is om met een proteomische benadering nieuwe Mediator interactoren te identificerendie betrokken zijn bij transcriptionele regulatie via enhancers.Geïdentificeerde transcriptiefactoren domineren neurale stamcel enhancers en zijn daar mogelijke recruiters van Mediator. Tevens hebben we genen geïdentificeerd nabij super-enhancers, welke waarschijnlijk zijn betrokken bij celidentiteit, en laten zien dat deze genen een hoge dichtheid hebben voor Mediator en zijn interactoren. We vinden dat genen met broad H3K4me3 promoters en super enhancers hoog tot expressie komen in neurale stamcellen en Mediator speelt waarschijnlijk een belangrijke rol bij het coördineren en stabiliseren van hun interacties. Onze data dragen bij aan het concept van fase-gescheiden chromatine-domeinen, waar de hoge dichtheid aan chromatine regulatoren leidt tot biocondensaten om de transcriptie van bepaalde loci efficiënt te kunnen reguleren.

In **hoofdstuk 3** bestuderen we de functies van Cggbp1, een van de transcriptiefactors geïdentificeerd als Mediator interactor. We laten zien dat Cggbp1 neurale stamcelgenen activeert via de promotor. Met een combinatie van bioinformatica en gepubliceerde data suggereren we een role voor Cggbp1 in de vroege neurale inductie via retinezuur signalering.

Van het neurale stamcel model in **hoofdstuk 2** gaan we naar embryonale stam cellen in **hoofdstuk 3** om vroege neurale ontwikkeling te bestuderen. Het is daarom passend dat in **hoofdstuk 4** we de laatste stap van neuronale ontwikkeling behandelen, neuronale rijping. We bepalen actieve chromatine regionen in de ontwikkeling en rijping van ex-vivo hippocampale neuronen en bepalen daarmee de regulatiegebieden voor neuronale rijping. We identificeren kandidaat transcriptiefactoren die mogelijk betrokken zijn bij de achtereenvolgende rijpingsstappen door hun bindingsmotieven en expressieprofielen gedurende hersenontwikkeling.

Samengevat bevat dit proefschrift een aantal nieuwe inzichten in de regulatie van transcriptie in de neurale lijn. In **hoofdstuk 5** bediscussieer ik de verschillende implicaties van de beschreven experimenten en presenteer ik voorbereidend werk voor nieuwe hypotheses in toekomstig onderzoek.

CURRICULUM VITAE

Personal details

Name	Martí Quevedo Calero
Date of birth	11-06-1988
Place of birth	Barcelona, Catalunya (SPAIN)

Education

2013-2018	PhD program at Erasmus Medical Center Department of Cell Biology, Erasmus Medical Center, Rotterdam, The Netherlands
2011-2013	Master's degree, Biomedical Research at The Universitat Pompeu Fabra (UPF), Barcelona, Catalunya (Spain)
2006-2011	BS, Human Biology at The Universitat Pompeu Fabra (UPF), Barcelona, Catalunya (Spain)
2000-2006	High school, IES La LLauna, Badalona, Catalunya (SPAIN)

Research

2018-present	Post-doctoral researcher at Umeå Plant Science Center (UPSC), Sweden PI: Prof. Åsa Strand "The role of the Mediator complex in histone and DNA methylation"
2013-2017	PhD student researcher at Erasmus MC, Rotterdam, The Netherlands PI: Dr. Raymond Poot "Transcriptional regulation in the neural lineage "
2011-2013	Graduate Research Assistant at Institute for Research in Biomedicine (IRB), Barcelona, Catalunya, Spain PI: Prof. Eduardo Soriano "Function of the mitochondrial Armcx cluster in neuronal development"
2010-2011	Student intern at the Faculty of Medicine of the Universitat de Barcelona, Barcelona, Catalunya, Spain PI: Dr. J.M. Canals "The role of Ikaros transcription factor in the neuronal development of the striatum"
2009	Student intern at Barcelona's Biomedicine Research Park (PRBB), Barcelona, Catalunya, Spain PI: Prof. J.M Muñoz "Oxidative stress and Alzheimer's disease"

PUBLICATIONS

(Under revision)
Martí Quevedo, Mike R. Dekker, Dick H.W. Dekkers, Johannes H. Brandsma, Debbie L.C. van den Berg, Zeliha Ozgür, Wilfred F.J. van IJcken, Jeroen Demmers, Maarten Fornerod, Raymond A. Poot. Mediator complex interaction partners organize the transcriptional network that defines neural stem cells.

(Published)
Estruch SB, Graham SA, **Quevedo M**, Vino A, Dekkers DHW, Deriziotis P, Sollis E, Demmers J, Poot RA, Fisher SE. Proteomic Analysis of FOXP Proteins Reveals Interactions between Cortical Transcription Factors Associated with Neurodevelopmental Disorders. Hum Mol Genet. 2018 Jan 222018.

Moen MJ, Adams HH, Brandsma JH, Dekkers DH, Akinci U, Karkampouna S, **Quevedo M**, Kockx CE, Ozgür Z, van IJcken WF, Demmers J, Poot RA. An interaction network of mental disorder proteins in neural stem cells" Transl Psychiatry. 2017 Apr 4;7(4):e1082.

Barbosa DJ, Serrat R, Mirra S, **Quevedo M**, de Barreda EG, Àvila J, Ferreira LM, Branco PS, Fernandes E, Lourdes BastosMd, Capela JP, Soriano E, Carvalho F.The mixture of "ecstasy" and its metabolites impairs mitochondrial fusion/fission equilibrium and trafficking in hippocampal neurons, at in vivo relevant concentrations.

Barbosa DJ, Serrat R, Mirra S, **Quevedo M**, Gómez de Barreda E, Avila J, Fernandes E, BastosMde L, Capela JP, Carvalho F, Soriano E. MDMA impairs mitochondrial neuronal trafficking in a Tau- and Mitofusin2/Drp1-dependent manner. Arch Toxicol. 2014 Aug;88(8):1561-72.

Serrat R, Mirra S, Figueiro-Silva J, Navas-Pérez E, **Quevedo M**, López-Doménech G, Podlesniy P, Ulloa F, Garcia-Fernàndez J, Trullas R, Soriano E.The Armc10/SVH gene: genome context, regulation of mitochondrial dynamics and protection against Aβ-induced mitochondrial fragmentation.Cell Death Dis. 2014 Apr 10;5:e1163.

Serrat R, López-Doménech G, Mirra S, **Quevedo M**, Garcia-Fernàndez J, Ulloa F, Burgaya F, Soriano E.The non-canonical Wnt/PKC pathway regulates mitochondrial dynamics through degradation of the arm-like domain-containing protein Alex3.PLoS One. 2013 Jul 2;8(7):e67773.

PHD PORTFOLIO

Name PhD student	Marti QuevedoCalero
Erasmus MC Department	Cell Biology
Research school	Graduate School MGC
PhD period	2013 - 2018
Promoters	Danny Huylebroeck, Frank G. Grosveld
co-promoter	Raymond A. Poot

PhD training

Courses

2013	Genetics (Rotterdam)
2013	Safely working in the laboratory (Leiden)
2013	Biochemistry and Biophysics (Rotterdam)
2013	Genetics (Rotterdam)
2013	Cell and Developmental Biology (Rotterdam)
2013	Biostatistical Methods (Rotterdam)
2014	Course in R (Rotterdam)
2014	Article 9. Animal handling (Rotterdam)

Workshops and Conferences

2013-2017	Departmental morning meetings and PhD seminars
2013	20th MGC PhD workshop (Luxembourg)
2013	23rd MGC Symposium (The Netherlands)
2014	TRR81 retreat (Austria)
2014	Organizer of 21st MGC PhD workshop (Germany)
2014	24th MGC Symposium (The Netherlands)
2015	TRR81 retreat (Austria)
2015	25th MGC Symposium (The Netherlands)
2015	TRR81 symposium (Germany)
2016	TRR81 retreat (Austria)
2016	14th Dutch chromatin meeting, speaker (The Netherlands)
2016	26th MGC Symposium, speaker (The Netherlands)
2017	Keystone symposia (US), Transcriptional and Epigenetic Control in Stem Cells + Neurogenesis during Development and in the Adult Brain (poster presentation)
2017	TRR81 retreat (Austria)
2017	Dev-Repair Symposia, speaker (Belgium)
2017	TRR81 symposium, poster (Netherlands)

Teaching activities

2015-2016	Supervisor of master student (I. Gordaliza)
2016-2017	Supervisor of master student (A. Lleches)

ACKNOWLEDGEMENTS

The defense date approaches and the only step left to do before sending this book to print are these last pages. While I started writing these lines I became a father (and kiss goodbye my good sleep) so my excuses beforehand if I forget somebody (thankszzz to all). What is also true is that somehow, I feel that after placing these final words of gratitude my time as a PhD student will finally come to an end and it makes their writing even heavier.

Let me start by emphasizing how fortunate I feel in having two excellent professors as promoters, Prof. Grosveld and Prof. Huylebroeck. **Frank**, I still remember the first question you asked me during the application to my PhD fellowship. Since then, all your comments and questions have been sharp and straight to the bone. You are clearly one of the greatest scientific influences I had during my PhD and a true inspiration to pursue my career in science in the field of transcriptional regulation.

With one promoter expert on the field of transcription then came another promoter expert in neurodevelopment. **Danny**, I want to thank you for all the help and support you have provided me since the very beginning of your arrival. I am most obliged for the interest you have shown on me and your efforts to keep me in the department. Believe that I feel confronted on the new challenging path I am taking. For certain I will miss your encyclopedial answer-questions and the new breeze you are bringing to the Cell Biology department.

But the real accountable for me being able to write these lines is my supervisor, Dr. Raymond Poot. **Raymond**, I can't thank you enough for putting your trust in me. Since my first days in the lab, when I would rush to your office with my "small technical questions" to the present day, while discussing paper revisions and thesis submissions, I have kept learning from you. As your student I have not only absorbed your biochemistry expertise, but I have also developed a scientific mind of my own. Your enthusiasm for research but also your directness in our discussions have always encourage me into do better science. In fact, you have flipped around my preconceptions of lab work (as you would often remind me that 1 good experiment is better than 5 mediocre ones). You have given me a lot of space to develop my own ideas while being very objective and incisive with my results. At the same time, you have been concerned for my life out of the lab, which has been critical to endure 5 years abroad separated from my wife and family. I hope you the best, both in research and with your family.

I am looking forward to the defense date as an excellent committee will assemble. My thanks to **Dr. Creighton** from the Hubrecht Institute; your presence honors me. Your exciting research in the field of epigenetics and evolution captivates my interest and makes me wonder about exploring new chromatin evolutionary adaptations. I extend my sincere gratitude also to Prof. Philipsen. Thank you **Sjaak** for all your support during these 5 years and especially for your corrections on this manuscript, they were much needed. Also, I would like to acknowledge your role as coordinator of the TRR81. The winter retreats at Kleinwasertal must have been the best yearly activity of my PhD, dancing to Jimi Hendrix included. It is also very special to me the presence in my committee of Dr. Galjart. **Niels**, without your efforts I would not have joined the PhD program at Erasmus. You were a fundamental piece to convince me on graduating in this department. I am happy you can witness also the end of this cycle. I would like also to thank **Prof. Gribnau** for being in my inner committee and **Prof. Baarends** to evaluate my defense. Although not completely in the same research field both of you

have been following my progress at several meetings and symposia and we have had a lot of fun at KWT.

Talking about KWT, I had the amazing opportunity to get in touch with outstanding scientists. **Dr. Wendt**, I am thankful for your support and interest in my projects. It was wonderful being able to discuss chromatin "stuff" with you while drinking a good Weiße. Thanks to **Dr. Mermoud** for the nice discussions about science politics and mountain hikes while sitting on the sky lifts. And also I send my gratitude to **Prof. Uta-Maria Bauer**, which her enthusiast and interest for my research were stimulating.

Many of the conclusions shown on this thesis wouldn't have been possible to reach without the collaborative effort of many scientists that I want to acknowledge. My thanks to **Dr. van den Berg**; the true initiator of the Mediator research in Dr. Poot's group. Your support at my embryonic days in the lab was much appreciated. I wish you tons of success in your come back to the department. Assistance provided by the biomics and proteomics facilities was greatly appreciated. I want to mention the outstanding support provided by **Dick Dekkers** from the proteomics center, for his affable access and efficient results. I would also like to thank **Dr. Fornerod**, despite our short time collaboration his inputs arrived in a critical time for my thesis.

Equally important has been the work of **Marike van Geest**, our department secretary. I want to express my sincere gratitude to her commitment and professionalism. I am amazed on her skills to manage a billion issues, including my defense, so much in detail.

I would also like to acknowledge the time spent with some collaborators. In particular, our work with **Prof. Fisher** at the Max Planck Institute for Psycholinguistics was a perfect fit both at the personal level (Sara wait for your acknowledge later!) but also in the research topic. I am happy we could materialize some of our efforts in a form of a published study and hope that the bigger piece of work on Tbr1 crystalizes in the future.

Before addressing my appreciation to friends and family, I would like to recognize the inspiration and mentorship I received during my whole education besides the PhD. These people constitute small pieces that together have built the pillars sustaining my love for the science of Biology. Starting from high school, I would like to thank **Silvia Lope** for her innovative ways of teaching science and her advice on university careers. Once in the university, **Prof. F.J Muñoz** showed me the joy on continuously asking questions, hinting me a career in research. I am also grateful for the few months I was on his group, my first contact with a real lab. I would like to thank **Dr. Martin-Ibáñez** for her kind support and cheerfulness during my long stay at the laboratory of Dr. JM Canals. I would have been close to lose hope in research without her side advice. In addition, I would like to thank **Dr. Serrat** and **Dr. Mirra** for their supervision during my master. Their efforts on teaching me while finishing their PhDs were amazing and I had a lot of fun as a visiting member of the ALEX team. Finally, I would like to acknowledge **Prof. Soriano** for his implication during my master, his will to keep me in his lab and his honesty during the deepest period of scientific economic crisis in Spain.

Indeed, it is irrefutable that the economic crisis in Spain and the continuously neglected investment in research and development from the Spanish government accounts for more than I wished for regarding my departure. Despite the denial state that is officially broadcasted from our central

institutions, there is a clear brain drain happening in Spain. I thank all the support that the scientific community has provided to scientists in exile like me.

Nevertheless, I was lucky to land in a wonderful place to not only work and learn but to have fun doing so. I am sincerely grateful for the warm welcome I received from all the members of back in the day lab 706.

Maaike, I really enjoyed all the great times we had outside the lab, from the improvised BBQs (I have your gift here in Sweden!) to our excursions to Hoek van Holland (what an amazing day!). I wished we could have had some of our stress-release cappuccino-breaks during my last phase of PhD (now I understand way better!). It was a pleasure to learn and work alongside you. I hope we can see each other when my little one can stand some hiking and invite your boys to explore some natural parks here in the north and to make a massive BBQ in the wild.

Johan, thanks for your cheerful presence and our talks in the office. Your efforts in analyzing my initial data and teaching me bioinformatics while finishing your PhD, having a baby and buying a house were at superhero level. I had tons of fun with you specially developing the infamous tradition of the flaming Blutwurz shots at KWT or around a pair of Kasteel Donkers. I have recently started my own garden here and sometimes I remember your dissertations on the progress of your garden with a smile. Congratulations with the new one at home and a big hug to Lídia.

Mike, I would like to thank you for all the support you have given me. During these 5 years we have both evolved together at the work place and fit into an efficient working team. I envy your attention to detail in protocols and lists. I hope you can continue your efforts to improve the lab and I wish you all the best outside of it (with your partner, new house and your cutie doggo!).

Erik, although we shared a small fraction of time in the lab I really appreciated your invitations to have some afternoon beers and your experienced advice on how to deal with supervisors. I wish you all the best in Germany.

Ernie, you were always there to calm my inexperienced hands. Your patience and cheer were essential to balance the energy in the lab. You possess an extraordinary aura of vitality that inspired me every day. I hope that with your retirement you get to travel even more and to continue hiking and rowing.

I also enjoyed to share the working space with the people from Harbour Antibodies; **Dubi**, **Alex**, **Rien** and **Michael** but specially **Rick**. You had always the will to stand any party we threw at you and I could always count on you for having a cocktail at Tiki's.

Working in academia has the bitter characteristic of meeting people always in transition. PhDs and Post-docs specially, we are a particular tribe of nomads in the 21st century. I met excellent people in the first years of my PhD that helped settle my tent in Rotterdam such as **Maria**, **Ileana** and **Aristea** to whom I send my sincere gratitude.

Many people from floors 6th, 7th, 9th and 10th have contributed to the friendly environment. From the 10th I would like to mention **Silvia**, which we have shared the same path in the PhD program and

Sreya, Valentina, Evelien and **Mieke** to keep the beer sessions going. From the 9th, I would like to specially thank Teresa not only as a bit of Catalan brought me home but also for the good times we shared. Also cheers to **Ruben**, my constant roommate at KWT, xd and **Hegias** and **Cristina** that would always have time to chat and as good Spaniards gossip a bit.

I had the extraordinary luck of being introduced to the Italian clan of expats. I was charmed by your closeness, **Andrea** and your culinary expertise (I miss the Christmas dinners!). Thanks also for your experienced advice and to start the climbing movement in the department! Congratulations on your wedding and good luck with your future research. **Enrico**, thanks for being my paranimf. I wished we could have joined the PhD at the same time to spend a longer period together. I really enjoyed our guitar jams with the band and I am really happy you are getting into climbing (I spy the whatsup group, xd). Forza for the last sprint on your thesis! Finally, **Luca**; my partner in crime during the whole PhD. I am so glad we got the same fellowship and started so close together in time. I could always count on you to talk about my worries or to cheer with some beers. I found salvation in our sky runaways to Austria or our music jams with the band. I am really going to miss you for my defense date but I am not sad as I am certain we will meet either in Barcelona (you lucky bastard!) or in Sweden (we have pending a X-country expedition).

My sincere gratitude to my other two paranimfs. **Lize**, I have seen you grow as a PhD student since you arrived and I am very happy how you are managing all the challenges. I am sure you will succeed in the path that lies ahead; it is less than 5c for sure! **Judith**, your energy and directness have been always a quality I loved. We are still missing a proper saxo-guitar jam! Thanks to both of you for the way you shape the wave of "new" students at the department. Your motivation for both research and partying has carried me on the last months where I was a social zombie.

Talking about the next wave, thanks to **Irene**, **Pablo**, **Rodrigo** and **Lukas** for the climbing moments. I had always wished for a climbing group and you were a perfect fit! Also, thanks to **Ilias** and **Jente** for the good times partying.

Besides climbing, what has kept me going in Rotterdam is music. In my mind I have assigned people to this category as we have met or befriended because of music. First of all, I would like to give special thanks to my guitar teachers in the Erasmusic initiative. Both **Olmo Marín** and **Alvaro Rovira** have brought me new flavors to add to my style and gave me hard homework to deviate my mind from the lab. **Lennart**, the remaining member of our rock band "The microtubules". Thanks for brining the groove and being a "sharp dress man", xd. **Cristina**, we met talking about Brazilian jazz at the cell culture room and since then I have enjoyed our music exchanges and your invites to DJ sessions. **Yasemin**, we shared the same fellowship and the same interests in music so it was inevitable to collide at some point. Luckily it was to make some nice music performances and to organize our own music jams where I met such as talented musicians such as **Danny**, **Isabella** and **Carmen**. I hope you keep the riffs going!

Before I talked about mentorship, and during my PhD I had the opportunity to supervise two wonderful students. **Adela**; thanks for your time in the lab, I learnt a lot from your stay and I hope to have transmitted a bit as well, at least in the climbing side! You have a raw potential that I hope soon you will exploit. **Isa**, I could not have asked for a better student to be my first supervision duty. You

shined the lab with your good mood and crazy laughs. I hope you both all the best in your future work and PhD and hope to meet you both in Barcelona!

I have been living in Umeå for almost 8 months now and I could not dismiss the opportunity to express my gratitude to several people that have welcomed me in the next step of my adventure. First of all, I would like to express my sincere gratitude to **Prof. Strand** for taking me into account and value my skills beyond background. I am looking forward to combine the best of our fields. Also I would like to thank your patience during the first phase of the project. I am extremely happy with the colleagues and friends I have met in my new group; **Tamara**, **Xu**, **Tim**, **Nora**, **Yan**, **Qi**, **Nico** and **Jannek**. In addition, I would like to specially thank **Simon** and **Lucy** for making me feel home even before I moved here. You guys are amazing! And also my thanks to the Spanish family we gather with **Ruben** and **Sonia** (and little **Einar**), **Juan**, and **Enara** and **Jaime** (and little **June**).

The remaining lines will be in Catalan, to the most inner circle of my soul.

Escric aquestes línies amb en Blai arraulit a sobre així que pot ser que passin dues coses; o em poso massa sentimental, o no m'expresso prou clarament; al cap i a la fi, espero expressar la meva gratitud en persona. El més important, però, és que sapigueu que tots vosaltres sou el veritable sostén de la meva felicitat.

Sara, no em puc creure que haguem arribat tan lluny. Qui diria que els dos ens faríem doctors quan començàvem a descobrir la ciència amb la poliploïdia de les maduixes. Les nostres òrbites sempre han anat sincròniques, des de les nostres primers feines al parc científic, a l'experiència holandesa fins al punt de graduar-nos amb una setmana de diferència! Fora de la feina sempre he tingut una connexió especial amb tu. No tinc germans de sang però crec que puc saber el que se sent tenint-te a tu. Espero poder assistir a la teva tesi i transmetre't aquests sentiments directament. **Joan**, des de que vam conèixer hem compartit les nostres ganes de gaudir del que la natura ens ofereix. M'encanta la teva energia i espero poder organitzar moltes més aventures amb tu. Encara recordo quan els dos em veu portar a navegar per dir-me que venia l´**Ona**. Em moro de ganes de trobar-nos tots sis (tots sis!!, que fort).

I ja que estem posats a organitzar viatges, que menys que incloure a l'equació a la **Laia** i la **Núria**, les nostres companyes de viatge. No sabeu el que hem feu riure cada una a la vostre manera. Compartir amb vosaltres els moments de relax és la millor medicina per l'estrès. Una forta abraçada, ens veiem a París?

Potser una de les persones que més me n'orgulleix de tenir com amic és l´**Álvaro**. Des de que ens vam conèixer m'has tractat no com la parella de la Marta sinó com un amic i amb el temps com família. La teva dedicació, bon humor i empenta m'inspiren cada dia. Per molt que la vida giri (i déu ni do com gira al voltant teu!) sempre tindràs un lloc amb nosaltres que pots anomenar casa. **Sara**, no saps que content estic de que hagis trobat feina a Hèlsinki. M'encanta veure com el vincle entre tots plegats no deixar de fer-se més fort. Us estimo!

Del mateix grupet d'umpa-lumpas estic contentíssim d'haver-me trobat amb l'**Albert**. Tot i els mals principis, els quals em conclòs que es un fenotip crònic teu xd, m´encanta haver-me apropat i establir una gran amistat amb tu. Em commou la teva fluïdesa d'interacció i les ganes de menjar-te la vida.

Estic impressionat amb la teva aventura californiana, et desitjo el més gran dels èxits! Una forta abraçada també per la **Carla**, espero que trobis en Blai tant "niño teletienda" com em vas trobar a mi. Segueix amb la teva força i il·lusió tant característiques. **Jordi**, tinc moltíssimes ganes de presentar-te en Blai i veure el teu jo pedagògic interior que encara no he pogut veure en acció! Si el tractes la meitat de bé que a mi serà increïble. No sabeu com espero amb ànsia veure's a tots junts.

Com he dit, no tinc germans però la meva família no sanguínia és extraordinària. **Oriol**, sempre et porto en els meus pensaments! Et trobo moltíssim a faltar, però em tranquil·litzo amb la certesa que aviat podrem començar a organitzar unes bones excursions amb els dos "pimpollos" a l'esquena, a poder ser als Alps prop de casa teva. Ets la proba de que la distància no importa quan el vincle és prou fort. Una abraçada per les teves xurris, la **Laura** i la petita **Emma**.

Parlant de família! Ara sí que podem dir que ja sou família oficial (amb boda i nen en menys de 15 dies!) tot i sempre m'hi heu fet sentir en aquests 10 anys. **Conxita**, **Joan**, **Laura** i **Helena**; moltes gràcies pel vostre amor i tota l'ajuda que ens doneu contínuament.

Papa i mama, pare i mare, **Manel** i **Marta**, (i aviat abu i iaia). Resumir en paraules escrites tots el sentiments que floreixen quan penso en vosaltres és una gran simplificació. Deixeu-me dir però, que sempre heu estat i sereu el meu més gran referent de la meva vida. No sabeu l'afortunat que em sento de tenir una relació tant directa i honesta amb vosaltres. Gràcies per tota paciència, sacrificis, lliçons i sobretot tot l'amor que m'heu donat per poder esdevenir la persona que sóc ara. I què bé que ens ho hem passat arribant fins aquí! Ben endins porto un gran sac d'experiències magnífiques que de tant en tant obro per il·luminar els dies grisos en països estrangers. Em moro de ganes de seguir-lo omplint amb noves aventures, ara amb un participant més. Us estimo tant que fa mal!

Blai, ara mateix dorms plàcidament en el caliu dels meus braços. El rítmic tecleig de l'ordinador actua de soroll blanc mentre el sol de primavera t'acaricia els peus. Esperar la teva arribada ha sigut el millor refugi on la meva ment trobava aixopluc durant els tempestuosos mesos finals de tesi. Cada dia amb tu és un milió d'experiments esplèndids.

I parlant d'experiments! Portem més de 10 anys rebutjant la hipòtesi nul·la amb tu **Marta**, i cada cop ens dona més significatiu. En altres paraules, cada cop t'estimo més i segueixo enamorant-me cada dia del teu somriure i de la teva mirada. Sembla estrany donar-te gràcies ara, quan en tot el temps que portem junts no has parat de transmetre'm el teu amor i la teva força. Quan tot s'enfosqueix i la nit m'envolta sempre trobo el camí buscant la llum de la teva estrella. Endavant ens queden infinitat de camins per explorar junts amb el nostre xipironet.

I si la nit no es torna clara
em tindràs al teu costat
per creuar el camp de batalla.

And if the night doesn't turn bright
you will have me on your side
to cross the battlefield.

Camp de batalla (Txarango)