

From The Department of Cell and Molecular Biology
Karolinska Institutet, Stockholm, Sweden

TRANSCRIPTIONAL REGULATION OF DEVELOPMENT IN TIME AND SPACE

Daniel W. Hagey



**Karolinska
Institutet**

Stockholm 2017

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by AJ E-print AB

© Daniel W. Hagey, 2017

ISBN 978-91-7676-819-8

On the cover: Embryonic day 18 mouse cortex labelled with antibodies against POU3F2/3 in dark blue, BCL11B in purple, HMGA2 in dark cyan and SOX3 in maroon.

Transcriptional Regulation of Development in Time and Space
THESIS FOR DOCTORAL DEGREE (Ph.D.)

November 3, 2017 at 9:30am

Samuelssonsalen, Berzelius väg 1

Daniel W. Hagey

Principal Supervisor:

Professor Jonas Muhr
Karolinska Institutet
Department of Cell and Molecular Biology

Co-supervisor(s):

Assistant Professor Ola Hermanson
Karolinska Institutet
Department of Neuroscience

Professor Carlos Ibanez
Karolinska Institutet
Department of Neuroscience

Opponent:

Professor Francois Guillemot
The Francis Crick Institute

Examination Board:

Professor Stefan Thor
Linköpings Universitet
Department of Clinical and Experimental
Medicine

Assistant Professor Arne Lindqvist
Karolinska Institutet
Department of Cell and Molecular Biology

Professor Per Uhlén
Karolinska Institutet
Department of Molecular Biochemistry and
Biophysics

To my family - past, present and future

“There is no scientific study more vital to man than the study of his own brain. Our entire view of the universe depends on it.”

— Francis Crick

ABSTRACT

Human development requires the generation of trillions of cells with myriad functions from a single cell. This requires that restriction of stem cell fate competence and proliferation are precisely temporally and spatially patterned as the embryo grows. To accomplish this, the chromatin landscape of individual stem cells progressively constrains gene expression in a context specific manner in order to guide cell behavior. In turn, this context is provided by the cellular environment and intrinsic determinants via the activity of transcription factors.

In paper I, we utilize ChIP-sequencing to study the overlapping and specific activities of the transcription factor sex determining region Y-box 2 (SOX2) in the developing cortex, spinal cord, stomach and lungs. We show that cell type specific binding is associated with tissue specific gene expression, while commonly bound cis-regulatory modules neighbor genes involved in the core processes of stem cell maintenance and proliferation.

In paper II, we use DNase- and ChIP-sequencing to demonstrate that, though the accessible chromatin landscape in the spinal cord and cortex are highly overlapping, SOX2 binding is primarily specific to one region. We find that this is due to an association with the specifically expressed partner transcription factors HOXA9 in spinal cord and LHX2 in cortex, which are capable of respecifying gene expression when misexpressed.

In paper III, we exploit single cell RNA-sequencing to establish that the stem cell population of the early cortex expresses high levels of *Sox2*, exhibits features of multipotency, and is enriched for genes involved in mitosis, such as *Ccnb1/2*. In contrast, the committed progenitor pool expresses high levels of the G1/S-phase genes, including *Ccnd1*, which is capable of inducing differentiation when overexpressed.

In paper IV, we find that *Sox2* acts in a dose-dependent fashion to control proliferation in the developing cortex by directly repressing *Ccnd1*. We show that this is accomplished via the binding of off-consensus sites in the *Ccnd1* promoter, and an association with Wnt signal transducing, TCF/LEF, transcription factors and their established co-repressor, TLE1.

LIST OF SCIENTIFIC PAPERS

- I. Hagey DW, Zouter C, Bergsland M, Klum S, Kurtsdotter I, Andersson O, and Muhr, J. SOX2 regulates common and specific stem cell features in the CNS and endoderm derived organs. *Manuscript*.
- II. Hagey DW*, Zaouter C*, Combeau G, Lendahl MA, Andersson O, Huss M# and Muhr J#. Distinct transcription factor complexes act on a permissive chromatin landscape to establish regionalized gene expression in CNS stem cells. *Genome Research* 2016 Jul; 26(7):908-17.
- III. Hagey DW*, Topcic D*, Kee N, Perlmann T# and Muhr J#. Cell cycle dependent regulation of cortical progenitor multipotency. *Manuscript*.
- IV. Hagey DW and Muhr J. Sox2 acts in a dose-dependent fashion to regulate proliferation of cortical progenitors. *Cell Reports* 2014 Dec 11; 9(5):1908-20

* These authors contributed equally to this work

Co-corresponding authors

CONTENTS

1	Introduction	1
1.1	Embryogenesis	1
1.2	the Chromatin Landscape.....	3
1.3	Transcription factor activity	5
1.4	Signaling pathways	7
1.5	Proliferation.....	9
1.6	Neurogenesis	11
1.7	Concluding remarks	13
2	Methods.....	15
2.1	Massively parallel sequencing	15
2.2	Genome wide transcription factor and chromatin profiling	15
2.3	RNA and protein expression profiling.....	16
2.4	Functional assesment of protein function	17
2.5	Functional assesment of CRM activity	18
3	Aims, Results and Discussion.....	19
3.1	Papers I and II: Spatial Patterning of gene expression in the Endoderm and Neural Tube	19
3.1.1	Aims	19
3.1.2	Results	19
3.1.3	Discussion	20
3.2	Papers III and IV: Regulation of stem cell proliferation is essential to temporal cell fate commitment	21
3.2.1	Aims	21
3.2.2	Results	21
3.2.3	Discussion	22
4	Acknowledgements.....	27
5	References	29

LIST OF ABBREVIATIONS

bHLH	Basic helix-loop-helix
BrdU	Bromodeoxyuridine
ChIP	Chromatin immunoprecipitation
CNS	Central nervous system
CRM	Cis-regulatory module
DHS	DNase hypersensitive site
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EMSA	Electrophoretic mobility shift assay
ESC	Embryonic stem cell
FACS	Fluorescence activated cell sorting
GABA	Gamma-aminobutyric acid
G1	Gap 1 phase
G2	Gap 2 phase
GFP	Green fluorescent protein
HMG	High mobility group
mRNA	Messenger ribonucleic acid
M	Mitosis phase
PH3	Phosphohistone 3
PNS	Peripheral nervous system
SE	Super enhancer
S	DNA synthesis phase
SVZ	Subventricular zone
TAD	Topologically associated domain
tSNE	t-distributed neighbor embedding
RGC	Radial glia cell
RNA	Ribonucleic acid

1 INTRODUCTION

1.1 EMBRYOGENESIS

The generation of an adult human from a single cell is amongst the most remarkable of phenomenon in our known universe. This feat requires that the blueprints for all of our attributes are contained within and decoded by the activities of twenty-three pairs of linear molecules of DNA, which meet for the first time in the fertilized egg. Following conception, a carefully orchestrated and amazingly reproducible series of chemical reactions unfold to give rise to trillions of individual cells (1), with myriad different functions, in an organism capable of understanding the process from which it arose. Although our actions are carried out by differentiated cells with specific roles, the construction and maintenance of our body relies on stem cells.

During development, stem cells become progressively fate restricted from the totipotency of the fertilized egg to the multipotency or unipotency of tissue resident stem cells, which are responsible for maintaining individual adult organs (2). Recent years have brought incredible advances to our understanding of the extent to which our own organs turn over from stem cells during our life time, such that some of our largest organs are regrown multiple times every year (3-6). However, it is as embryogenesis proceeds that individual stem cells become fated to distinct lineages, and it is these decisions that my thesis will focus upon.

After fertilization, individual cells at the four cell stage begin to show molecular signatures revealing their bias towards trophoectoderm, which will become the placenta and extraembryonic support cells in utero, or inner cell mass fates (7-8). The inner cell mass will progressively give rise to the hypoblast, which will form the yolk sac, and a single cell layer of epiblast cells, which will form every part of the embryo from a single layer of cells (9). Once it has expanded, an organizing center is established on the epiblast surface and orchestrates localized cell invagination in a process called gastrulation (from latin stomach – formation) (10). While the invaginated cells will form the endoderm and mesoderm, the surface of the epiblast will form the ectoderm (11). These three germ layers will each form well defined organs within the developing embryo, and comprise the first major restriction in embryonic stem cell (ESC) fate (12). Thus, in what was a uniform cell layer, gastrulation simultaneously restricts cell fate and imbues dorsal-ventral (back-belly), anterior-posterior (head-foot) and right-left axes based on the orientation of organizer migration.

While the organizer is still migrating posteriorly, further inductive events continue to unfold anteriorly. At the midline of the mesodermal cell layer, a newly formed structure called the notochord signals to the overlying ectoderm to take on a neuroectodermal fate and form the neural plate, which will fold into the neural tube in a process called neurulation (13). This tube will form every cell of the central nervous system (CNS), while a migratory derivative of it called the neural crest, will form numerous cell types, including the peripheral nervous

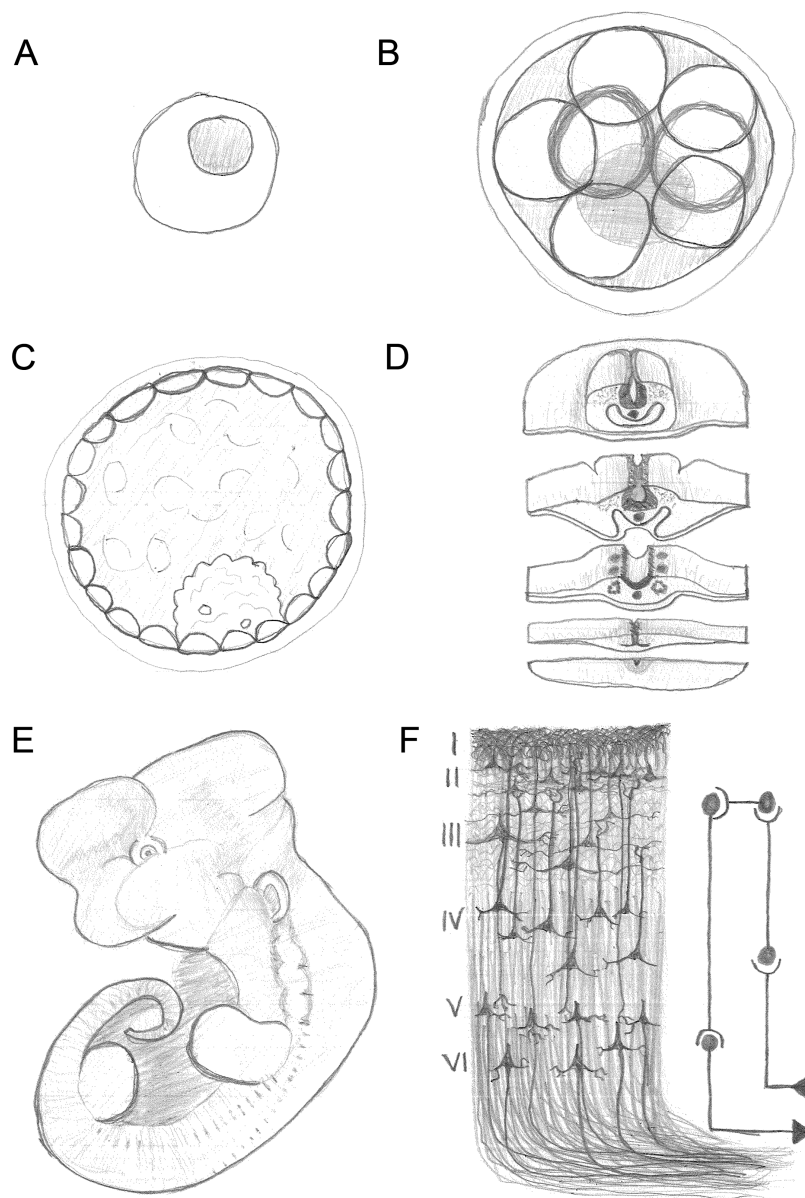


Figure 1: Embryogenesis
 A) A single totipotent cell results from fertilization.
 B) The eight cells of the pre-implantation embryo begin to undergo specification.
 C) The blastocyst consists of the inner cell mass and supporting trophoblast.
 D) Gastrulation proceeds as the organizer induces cellular invagination and patterns the three germ layers. Subsequent neurulation proceeds via the folding of the neural plate to form the neural tube.
 E) The early embryo appears similar in all vertebrate species, and has specific structures that will each give rise to individual organ in the adult.
 F) The fully formed cortex consists of columns of six neuronal layers, whereby inputs arrive to layer IV, the signals are processed by layers II and III, and outputs are projected to subcortical targets from layers V and VI.

system (PNS) (14). The non-neural ectodermal cell layer will give rise to the outer most barriers of the body, such as the epidermis, hair and nails (15). Similarly, the mesoderm will give rise to many defined internal structures, such as muscle, fat, bone and cartilage, as well as the urinary and immune systems (16). Finally, the endodermal cell layer will also form a tube, which is the precursor to our respiratory and digestive systems (17). Through these various lineage trajectories, all the cells of the human body are each narrowly fate restricted to their specific roles.

The mechanisms that segregate lineages at various steps can fall into various categories. As an example, the mammalian cortex is anatomically described to consist of six neuronal layers, which all utilize the neurotransmitter molecule glutamate, but have different axonal targets, such that upper layer neurons project to other neurons in the cortex, while deep layer neurons project to other regions of the brain (18-19). However, in order to give rise to neurons with these different attributes at the correct place and time, cortical stem cells must be instructed in

a process called patterning. If we follow the example of the lineage leading to the generation of a cortical neuron from neural tube precursors, the first consideration we must take into account is the location where the neuron is born within this stem cell population. Termed spatial patterning, this means that there are specific signals present where cortical neurons will be born, from the dorsal half of the neural tube, at its anterior tip – the forebrain or telencephalon (20). In this case, the location of cortical stem cells imbues them with both the ability to generate the structure of the cortex, as opposed to that of the spinal cord for example, as well as the predisposition to give rise to neurons that utilize glutamate as a neurotransmitter, in comparison to those born ventral to the cortex, which use gamma-aminobutyric acid (GABA) as a signaling molecule (21). Following cortical fate restriction, the different layers of neurons are all born from the same stem cell pool in a sequential fashion, such that the deep layers arise first and are migrated past by later born upper layer neurons (22). Thus, once spatially patterned, cortical stem cells must then shift their competence as neurogenesis proceeds, in a process termed temporal patterning. By repeatedly utilizing these complimentary strategies to make progressively more specific lineage decisions, a single stem cell can give rise to an independent organism with a brain capable of understanding and studying these processes.

1.2 THE CHROMATIN LANDSCAPE

The human genome project has made clear that, although the blueprints for human development are contained within every cell of our body, decoding the information into messenger RNA (mRNA) and protein expression profiles occurs in a cell type and context specific manner (23). This is because the genome never exists as isolated molecules of DNA, but instead continuously interacts with transcription factors, histones and architectural proteins, which imbue cells with their specific transcriptional profiles (24). Since cis-regulatory module (CRM)-promoter interactions are essential for gene expression, these factors regulate which regions of the genome are accessible to the transcriptional machinery (25) and limit the CRMs that can interact with a gene's promoter (26).

In general, as stem cells become more specified during development, and then differentiate into distinct cell types, the genome becomes progressively less accessible, though this feature has been shown to be paralleled by the selective formation of de novo open chromatin at specific CRMs (27). This suggests that certain transcription factors with pioneering function, the ability to bind and open heterochromatin, may play a unique role in controlling stem cell fate competence during development (28). However, it is also important to note that the degree of chromatin accessibility can vary greatly between open and closed. Histone octamer's tails can be modified in many ways that affect their association with DNA, and can even form higher order heterochromatin structures involving linker histones and associated proteins (29-30). Thus, although essential to cell fate decisions, chromatin accessibility is highly dynamic, and the process of somatic cell reprogramming clearly demonstrates that chromatin accessibility can be reset. Moreover, the finding of reprogramming factors in

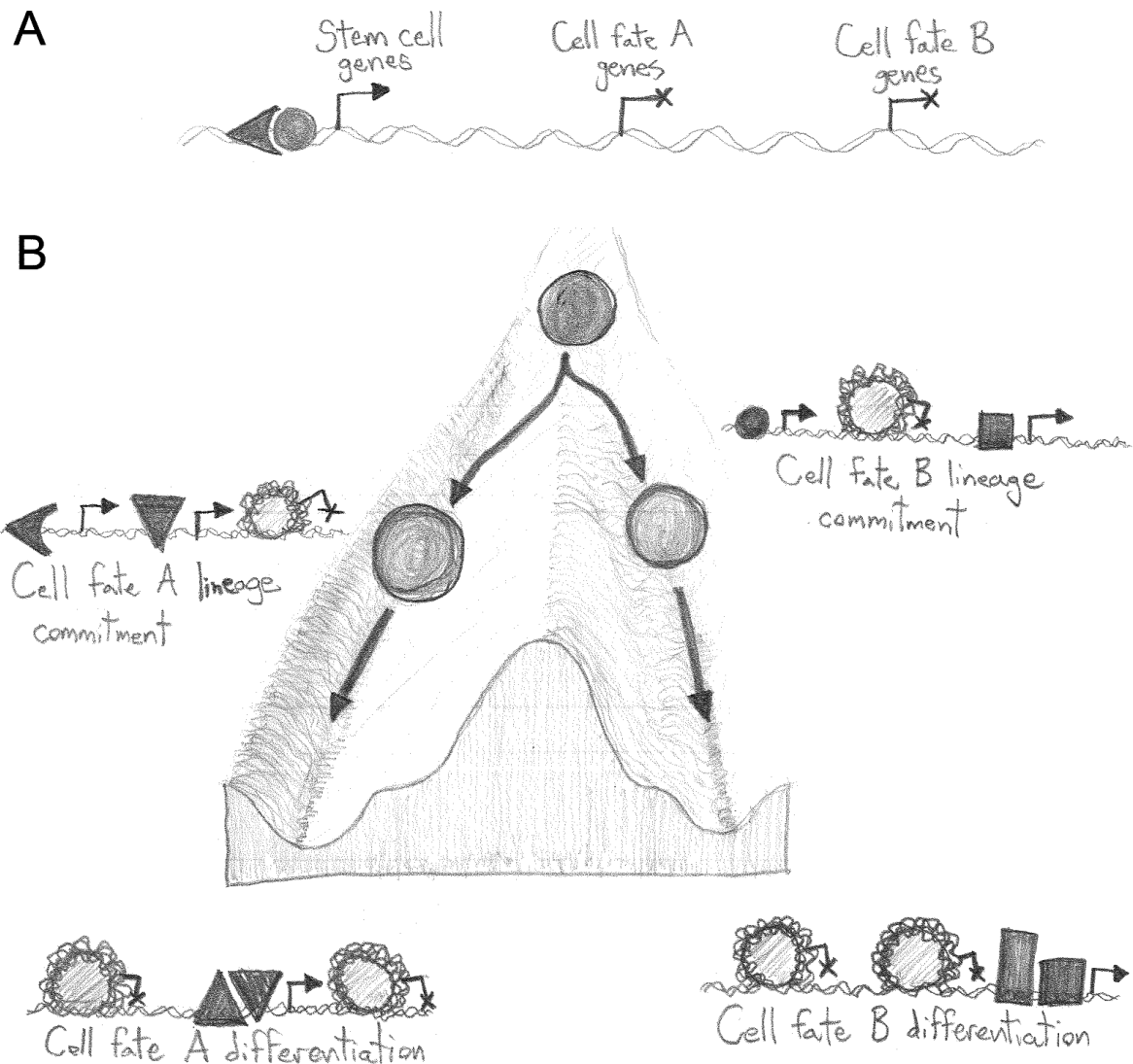


Figure 2: Chromatin and transcription factor dynamics during cell lineage differentiation. A) Depiction of three classes of genes, one expressed by cycling stem cells, one expressed by lineage A, and one expressed by lineage B. In multipotent stem cells, the stem cell transcriptional network is maintained by specific transcription factors that activate stem cell genes. Lineage specific genes are accessible, but are not robustly activated due to a lack of targeting transcription factors. B) Waddington model of cell specification shows lineage commitment linked to the upregulation of specific transcriptional determinants that activate each lineage's specific stem cell transcriptional network and lineage genes, as well as the compaction of chromatin around opposing lineage genes. Cell differentiation then proceeds by the reinforcement of lineage specific gene expression by other specific transcription factors, as well as the shutting down of the stem cell transcriptional network.

cancer cells illustrates the importance of maintaining tight regulation of chromatin accessibility (31).

Despite its necessity, DNA accessibility is not sufficient to specify transcriptional programs. Due to the size of the genome, the number of potential CRM-promoter interactions must be limited in order to ensure accurate gene expression, and this is achieved by insulating different regions from interacting with one another. Upon fertilization, the genome is

unorganized, with very little compacted heterochromatin or nuclear architecture (32). As transcription begins, architectural proteins, such as CCCTC binding factor (CTCF) and the Cohesin and Mediator complexes, bind to essential regulatory domains termed topologically associated domain (TAD) boundaries. These organize both large, megabase sized TADs, which are less variable between cell types, and smaller, kilobase sized loops, which are highly dynamic and cell type specific (33). Essentially, CRM-promoter interactions can only occur within a TAD loop, and if a TAD boundary is lost it can result in ectopic regulatory interactions and deregulated gene expression. Importantly, several transcription factors shown to be key determinants of cell state have been shown to be enriched at TAD boundaries, and thus have direct input on the cell type specific architecture of the genome (34).

Although complex, the biology of transcription functions just as any chemical reaction, whereby modifications and interactions occur in a concentration dependent manner. By targeting relevant CRMs, transcription factors bring the transcriptional machinery into proximity with specific genes and promote their expression (33). In certain cases, around key cell fate genes, termed super enhancers (SEs), many such CRM associations come together to form densely packed areas of protein aggregation. These are so dense, and involve so many interactions, that they exclude water from their cores and form almost pure protein droplets within the otherwise aqueous environment. This produces highly efficient transcription factories, which drive high expression levels of nearby genes (35). In addition, the large number of interdependent interactions involved make SEs highly sensitive to the concentrations of their constituent proteins (36). Notably, it is precisely this mode of sensitivity to transcription factor levels that is required for the dynamic regulation of cell fate decisions during embryogenesis.

Thus, by limiting what genes and CRMs are accessible to the transcriptional machinery, stem cells become progressively fate restricted during development. However, it is by biasing the probability of transcription from specific genes that the chromatin landscape generates precise transcriptional profiles from the stochastic activities of molecules in the nucleus.

1.3 TRANSCRIPTION FACTOR ACTIVITY

Because the chromatin landscape is primarily shaped by proteins with little respect for which segment of the genome they are interacting with, the specificity required for distinct transcriptional states must be mediated by transcription factors, which bind to precise genomic regions based on their DNA sequences. The human genome encodes over a thousand transcription factors, which can be grouped based on their DNA binding domains, as these each bind to characteristic DNA sequence motifs (37). Despite the number of factors, their recognition sequences are relatively short, and thus the size of the human genome implies that the specificity provided by individual factors is still not sufficient to direct the processes necessary for development. To overcome this, transcription factors often bind in complexes, which not only increases their affinity for longer and more distinct DNA

sequences, but also provides specificity to the co-factors they will recruit for gene regulation in different contexts (38-39).

As stem cells differentiate from a totipotent state, the competition and cooperation between transcription factors is essential to directing individual cells towards specific fates. Amongst the most studied transcriptional complexes in biology is that formed by the core ESC transcription factors SOX2 and POU5F1, which maintain the ESC transcriptional program despite the relaxed chromatin environment of ESCs (40). However, when SOX17 becomes expressed in presumptive endodermal cells, it competes with SOX2 to partner with POU5F1, and causes a shift in binding to a slightly compressed DNA motif (41). This difference is sufficient to disrupt the ESC state and begin activating the endodermal transcriptional program. This is only one of many examples where the switching of partner factors re-specifies the binding of a core transcription factor.

Many transcription factors can act to both enhance and repress transcription of different target genes within the same cell at the same time (42). In order to perform these opposing actions simultaneously, individual transcription factors act in a context specific manner. This context can be provided by the CRM sequences surrounding different target genes, via attraction of distinct partner transcription factors and co-factors, or by other components within the cellular or nuclear environment (42). Transcription factors acting downstream of signaling pathways almost always function in this manner, such that the active signaling pathway permits nuclear localization of a co-activator that displaces a co-repressor and thereby induces target gene expression (43). Additionally, dependent on the cellular environment, transcription factors can be covalently modified and thus alter their interacting co-factors (44). Finally, variations in the splicing patterns of transcription factor genes can generate proteins with highly divergent functions (45).

Transcription factor activity is central to cell fate determination and must be tightly regulated in a context specific manner. In order to simultaneously regulate the global chromatin landscape and distinct activities of specific CRMs throughout development, transcription factors must utilize several strategies. First, master regulators of cell fate identity bind to TAD boundaries and SEs around key genes to assemble large protein complexes and regulate the architecture of the nucleus (34). Second, by utilizing different partner transcription factors, an individual factor can target divergent sets of CRMs and genes in diverse cell types (38). Lastly, the co-activators and co-repressors recruited by an individual factor can vary greatly dependent on the partner factors utilized around specific genes, and the modifying enzymes present to regulate their interactions (42). Thus, transcription factors provide the foundation required for cell type specific gene regulation via their context dependent activities.

1.4 SIGNALING PATHWAYS

The serial restriction of stem cell fate necessary for embryogenesis is executed within cell populations, which must give rise to appropriate cell types in a defined order and pattern. These processes require fine-tuned coordination between cells at a distance from one another. This has been shown to be mediated by only eleven signaling pathways, which act by binding to receptors on target cells and affecting intracellular events that influence transcriptional regulation (43). While the Notch and Hippo pathways act only on neighboring cells, by cell surface bound ligands, the fibroblast growth factor (FGF), epidermal growth factor (EGF), Wingless/Integration-1 (Wnt), Transforming Growth Factor Beta (TGF β), Hedgehog (Hh), cytokine tyrosine kinase, Jun kinase, nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) and retinoic acid receptor (RAR) pathways act at distance by secreting ligands into the extracellular space (43). Here, I will briefly summarize the soluble morphogen pathways relevant to my thesis, and how these act on target cells in a concentration dependent manner to pattern embryonic structures.

To begin with the earliest patterning events, organizer induction has been shown to be executed by the simultaneous activation of the Wnt and TGF β pathways in presumptive organizer cells within the epiblast (46). *Wnt* genes code for proteins, which are secreted into the extracellular space from source cells following serial post-translational modifications (47). Upon diffusion to target cells, WNT proteins bind to Frizzled (FZD) and lipoprotein receptor-related (LRP) receptors at the cell surface, which allows them to recruit Disheveled (DSH) and AXIN proteins to the inner surface of the cell membrane (48). AXIN is a key component of the destruction complex, which also consists of glycogen synthase kinase 3 (GSK3), adenomatous polyposis coli (APC) and casein kinase 1a (CK1a). One of the key targets of the destruction complex is the cell-cell adhesion molecule β -catenin, which it ubiquitinates and targets for degradation (49). When AXIN is recruited to the cell membrane by active Wnt signaling, soluble β -catenin begins to accumulate in the cytoplasm and translocates to the nucleus (50). In canonical Wnt signaling, β -catenin binds to members of the T-cell factor/lymphoid enhancer factor (TCF/LEF) family of transcription factors upon translocation to the nucleus (51), where it displaces transducin like enhancer of split (TLE1) co-repressors and acts as a potent co-activator of their target genes (52).

The TGF β superfamily of ligands includes bone morphogenic proteins (BMPs), growth differentiation factors (GDFs), Activins and TGF β s, which are each involved in various cellular patterning events (53). Despite the diversity of the family, they all act by bringing together different type I and type II receptors (Alk genes), which induces the phosphorylation of the type I receptor (54). Activated type I receptors then phosphorylate either Smad1, Smad5 or Smad8 in the BMP signaling pathway, or Smad2 or Smad3 in TGF β , Activin and most GDF pathways (53). These phosphorylated Smads then have a great affinity for Smad4, which mediates their translocation to the nucleus and specific transcriptional activities (55).

Once the telencephalon has been established, morphogen signaling is essential for the regionalization of the brain. First, the cortex must be distinguished from the ventral forebrain,

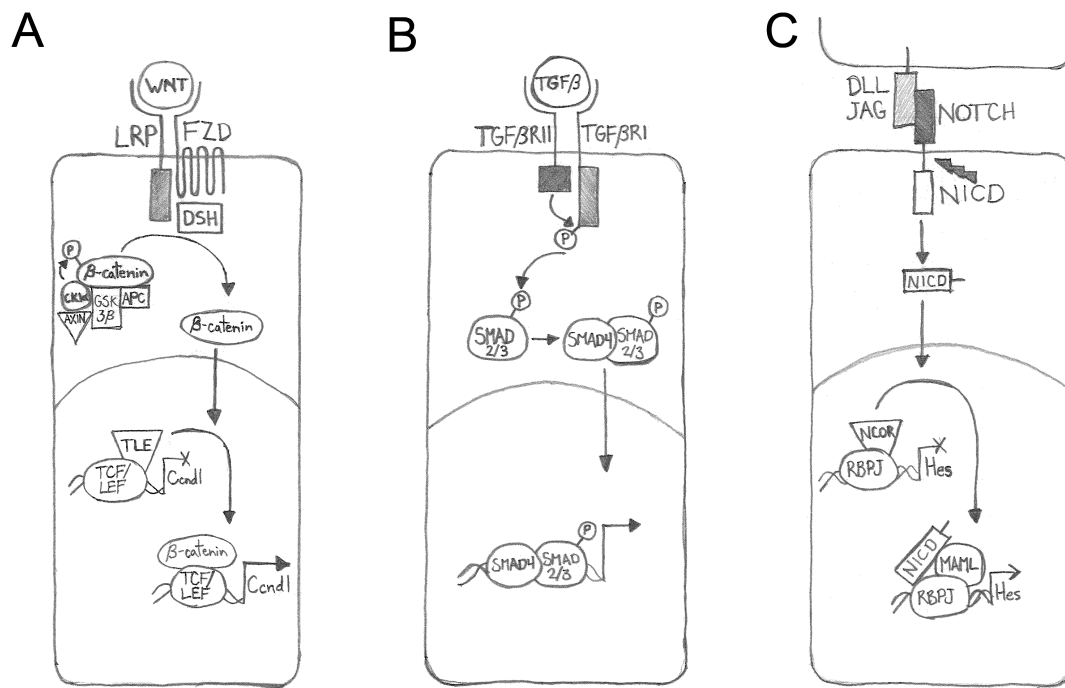


Figure 3: Extracellular signaling pathway transmission to the nucleus.

A) The Wnt signaling pathway is activated when WNT ligands bind to LRP and FZD receptors at the cell surface. This induces the recruitment of DSH protein to the plasma membrane, and the inactivation of the β -catenin destruction complex. As soluble β -catenin accumulates, it translocates to the nucleus, where it displaces TLE co-repressors from TCF/LEF family transcription factors and activates target genes such as *Ccnd1*.

B) The TGF β signaling pathway begins with TGF β binding inducing the dimerization of TGF β RI and TGF β RII receptors. This allows TGF β RII to phosphorylate TGF β RI, which in turn phosphorylates SMAD2/3. Phosphorylated SMAD2/3 has a high affinity for binding to SMAD4, which translocates this active complex to the nucleus where it activates target genes.

C) The Notch signaling pathway is induced when a neighbouring cell commits to differentiation and upregulates DLL/JAG ligands at its surface. When these ligands bind to NOTCH receptors, they induce the intracellular cleavage and release of the NOTCH intracellular domain (NICD). When NICD translocates to the nucleus, it displaces RBPJ co-repressors, such as NCOR, and activates target genes, such as *Hes*, in complex with MAML co-activators.

which will give rise to the ganglionic eminence (56). Subsequently, it must be patterned along its surface in order to generate all of the specific areas dealing with everything from sensory perception to speech (57). The molecules most characterized in these events are Sonic Hh (SHH), which broadly patterns the dorsal-ventral axis of the neural tube (58-59), and FGF8/17, which have been shown to be essential for cortical surface arealization (60). In Shh signaling, the cholesterol bound SHH ligand is excreted from signaling center cells (61), notably those of the notochord and floor plate, and binds to Patched (PTCH) receptors on target cells (62). This binding relieves PTCH receptor inhibition of the transmembrane protein Smoothed (SMO), which in turn leads to stabilization of GLI activator transcription factor function (62). This creates a gradient of GLI activator to GLI repressor activity, which is highest near ventral SHH sources and most repressive in the cortex, and thus acts as a potent regulator of telencephalon patterning (63).

Upon establishment of a cortical fate within the telencephalon, FGF8/17 begin to be expressed from the anterior midline of the cortex (57). FGF8/17 bind to the extracellular regions of FGF receptors (FGFR) 1-4, and mediates their dimerization. This results in the trans-autophosphorylation of their intracellular kinase domains and activation for downstream signaling (64). As with other growth factor receptors, activated FGFRs can phosphorylate proteins involved in the major intracellular kinase pathways: mitogen activated protein kinase (MAPK), phosphatidylinositide-3-kinase (PI3K)-Akt, phosphoinositide phospholipase C (PLC) and signal transducer and activator of transcription (STAT) (65). Each of these have highly specific functions in diverse cellular processes and lead to various transcriptional outputs. However, in the context of cortical patterning, loss of FGF8 leads to a drastic reduction in anterior motor cortex areas at the expense of an expansion in the more posterior somatosensory and visual areas (66-67).

Along with the other signaling pathways mentioned above, the Wnt, TGF β , Shh and FGF pathways play crucial roles at various stages of embryonic development. By consecutively restricting stem cell fate, and then reapplying the same pathways on the new landscape of cell competence, the embryo rapidly produces finely patterned cellular diversity from a small, uniform population of cells.

1.5 PROLIFERATION

The cell cycle is a core stem cell processes, and its components have also been demonstrated to play important roles in differentiation and cell fate determination (68-69). In order to generate the massive cell numbers required in the construction and maintenance of an adult human, stem cells divide rapidly during embryogenesis, and then slow their proliferation as we age (70). As DNA damage and mutations can only become fixed in the stem cell genome when those errors are replicated and inherited by daughter cells, it is important for the process of cell division to be continuously checked to allow for repair (71). Thus, in order to sustain their longevity, stem cells utilize several strategies and mechanisms to ensure the fidelity of their reproduction.

One important strategy, widely applied to prevent replicative errors and sustain stem cell endurance, is to reduce their overall number of divisions. This is achieved by a stem cell dividing to produce one stem cell, and one transit amplifying cell. Transit amplifying cells are able to divide a defined number of times before all of their progeny differentiate into defined cell types dependent upon the tissue in which they were generated (72). Since an individual differentiated cell has a limited impact on an individual's health and committed transit amplifying cells produce a limited number of them, they replicate their DNA more rapidly than stem cells, and thus take on the mutagenic replicative burden (73-74). This hierarchy of proliferation is utilized extensively throughout the developing embryo and in adult tissue maintenance, and is tightly integrated with the process of differentiation (75).

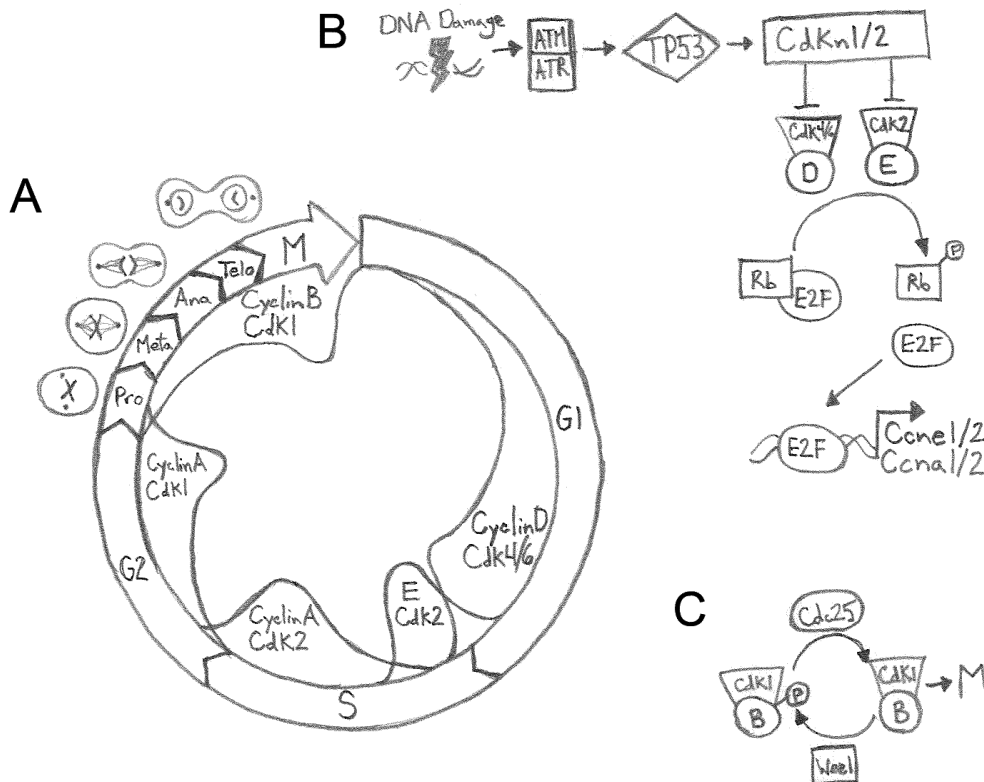


Figure 4: Regulatory mechanisms of cell cycle progression. A) Schematic depiction of cell cycle progression through G1, S, G2 and M-phases. M-phase is broken down into pro-, meta-, ana- and telophases, with chromosome and centrosome movements depicted. Inside the circle, activity levels of the various Cyclin/Cdk complexes are depicted throughout the different cell cycle phases. B) The Cyclin/Rb/E2F-pathway promotes G1/S phase progression, and is negatively regulated by an ATM/ATR/TP53 mediated DNA-damage response pathway. C) CyclinB/Cdk1 activity promotes M-phase progression, but is negatively regulated by WEE1 phosphorylation and positively regulated by CDC25 phosphatase activity.

Despite variations in kinetics, the cell cycle proceeds as a repeated order of events, regardless of the cell type. As a newborn stem cell separates from its sister, it re-enters a new cell cycle that begins in gap 1 (G1) phase, and progresses through DNA synthesis (S), gap 2 (G2) and concludes with mitosis (M) (76). During G1, various inputs lead to the accumulation of D-type cyclins (CCND1-3), which partner with cyclin dependent kinases (CDK) 4 and 6 upon reaching sufficient levels (77). Active CDK4/6-CyclinD complexes stimulate cell cycle progression by phosphorylating retinoblastoma protein (RB) and inhibiting its binding to E2 factor (E2F) family transcription factors (78). Upon release, E2F factors bind to and activate their downstream target genes, which include the E-type (CCNE1/2) cyclins and A-type (CCNA1/2) cyclins (79). CyclinEs accumulate rapidly at the end of G1 and partner with CDK2 to further phosphorylate RB and other targets involved in the activation of DNA-synthesis (80). CyclinAs accumulate slowly throughout DNA-synthesis and G2, but when partnered with CDK1/2 they mediate the phosphorylation of targets involved in DNA-synthesis progression and entry into G2 (81). Upon completion of DNA-synthesis, B-type cyclins (CCNB1-3) accumulate and ultimately license the entry into and progression of M-

phase (82-83). M-phase unfolds as pro-, meta-, ana- and telophases, and begins as centrosomes migrate to opposing sides of the cell and extend spindles between them, the nuclear envelope dissolves and individual chromosomes condense into discrete chromatids, with sisters joined by a centromere. Once the chromatids have aligned along the center of the cell and the mitotic spindles have attached to each centromere, this arrangement is resolved when the centromeres split and sister chromatids migrate to opposite poles of the cell. The whole process is completed in cytokinesis, when the cell membrane pinches in, divides the daughter cell's cytoplasm and permits cells to reconstruct the nuclear envelope and re-enter G1 (84).

Although the ordered series of events that unfold during cell division are licensed by Cdk/Cyclin complexes, the activity of these is heavily regulated in order to ensure the fidelity of stem cell division. For instance, at all stages of the cell cycle, DNA breaks can be detected by ataxia telangiectasia mutated (ATM)/ataxia telangiectasia and Rad3-related (ATR) kinases, which directly phosphorylate tumor protein 53 (TP53) (85). TP53 is widely regarded as one of the most essential tumor suppressors in cell biology, as it directly activates cyclin dependent kinase inhibitor p21 (CDKN1A), which results in cell cycle arrest, and bcl2-like protein 4 (BAX), which pushes cells towards apoptosis (86). During normal cell cycle progression, Cdkn1a and other cyclin dependent kinase inhibitors, CDKN1B (p27), CDKN1C (p57), CDKN2A (p16), CDKN2B (p15), CDKN2C (p18) and CDKN2D act to inhibit CDK2 and/or CDK4/6 containing complex activity (87). Finally, CDK1 activity is competitively repressed by WEE1 G2 checkpoint kinase phosphorylation and activated by cell division cycle 25 (CDC25) phosphatase activity (88). Thus, many layers of cell cycle licensing proteins and their inhibitors act during specific cell cycle phases in order to integrate cellular inputs and ensure orderly cell cycle progression.

1.6 NEUROGENESIS

Ultimately, a neural stem cell's role is to generate the differentiated progeny that will play functional roles in the activity of the adult brain. How to balance the maintenance of a stem cell population with commitment to differentiation is a question, which has important implications for understanding how our brain is formed during development. In addition, these processes continue to be active as adults, with important implications to mental health, and tumor formation. Thus, I will summarize the key pathways regulating neurogenesis with a focus on the cortex.

As mentioned above, the Notch pathway acts only between cells in contact with one another and executes the lateral inhibition of differentiation in cells neighboring those that have committed to neurogenesis. Upon commitment, progenitors upregulate Delta-like (DLL1-4) and Jagged (JAG1/2) ligands, which they present on their cell surface (89). These bind to NOTCH receptors on neighboring cells, and induce their cleavage, which results in the release of the NOTCH intracellular domain (NICD) in those cells (89). The NICD is then

transported to the nucleus, where it displaces co-repressors, such as nuclear co-repressor 2 (NCOR2), from recombination binding protein suppressor of hairless (RBPJ) and activates its target genes in cooperation with Mastermind-like transcriptional coactivator (MAML1) (90). Key target genes activated by NICD/RBPJ include members of the Hairy/enhancer of split (HES1-7) family of basic helix-loop-helix transcription factors (bHLH), which directly repress expression of proneural bHLH transcription factors such as Neurogenins (NEUROG1-3) and Achaete-scute like (ACL1-5) (91). A threshold of NEUROG or ASCL protein activity is a powerful pathway committing cells to differentiation, as it triggers a cascade of other bHLH proteins (92-94). Sufficient concentrations are essential, as all proneural bHLH proteins must bind an E protein (TCF3, 4 or 12) partner in order to bind DNA. This means that proneurals must compete with E-protein homodimers and sequestration by the inhibitor of DNA-binding (ID1-4) protein family (95).

In parallel to the environmental regulation of differentiation imposed via Notch signaling, intrinsic determinants also play powerful roles in influencing commitment decisions. The sex-determining region Y-box (SOX) family of transcription factors all share a high mobility group box DNA binding domain, and play important roles at all stages of neurogenesis. The SOXB1 (Sox1-3) group of proteins have been shown to strongly repress differentiation, even in the absence of active Notch signaling (96-97). SOX2 expression levels have been shown to be inversely correlated with the degree of commitment to neurogenesis (98), and it is one of the core factors used to reprogram somatic cells back to a pluripotent ESC state (99). In contrast, the SOXB2 group protein SOX21 is strongly upregulated by SOXB1s and has been demonstrated to force cell cycle exit without committing cells to neurogenesis, likely via a tumor suppressor pathway shared with SOX5 and SOX6 (100-101). Once stem cells have committed to neurogenesis, the SOXC group (Sox4,11 and 12) are upregulated by bHLH proteins, whereby they directly activate neuronal genes, such as Beta-III-tubulin (TUBB3) (102). These examples suggest that the application of single families of transcription factors to a linear differentiation pathway has particular potency in regulating the progression of neurogenesis at least partly due to the competition for, and differential regulation of, their shared target genes.

In the cortex, neurogenesis begins around E10.5 as radial glia cell (RGC) nuclei migrate between the apical ventricular surface, where mitotic cell division occurs, and the subventricular zone (SVZ), where DNA-synthesis occurs. All RGCs retain both an apical foot and a basal process that extends to the surface of the cortex, as this allows them to receive soluble signals from the ventricular zone and provides a scaffold for newborn neurons to migrate along (103). Upon commitment to differentiation by NEUROG1-3, progenitors upregulate B-cell translocation gene 2 (BTG2) and lose their apical foot. The vast majority will then enter an EOMES expressing, transit amplifying state that occurs in the SVZ and outer SVZ and can last over ten cell divisions before terminal neurogenesis in humans (104). This system permits the generation of over a thousand neurons from a single RGC division, and is one mechanism that has allowed the massive expansion of our brains from ancient molecular mechanisms.

1.7 CONCLUDING REMARKS

For a single species, humans have put an unprecedented mark on the planet earth. Many of our most sacred documents suggest that this is our right, but our origins and actions arise from the same physical and chemical reactions as any process in the universe. Moreover, the similarities between early human embryos and those of other animals provides a humbling counter perspective to any belief in human exceptionalism (105). Ultimately, the path leading to humanity's ascension on earth is the culmination of billions of years of reproductive chemistry, and a respect for that history is essential to guiding our future.

Our modern disconnect from nature may make it difficult to believe, but the attributes that separate us from other species are not as grand as often purposed. This has been powerfully demonstrated by the incredible similarity between the sequence of our genome and those of other animals (106). For example, SOXB1 proteins have existed for over a billion years, since the innovation of multicellularity. However, genomic data has also given us the first solid leads to understanding what it is that has allowed us to be so successful. Unsurprisingly, this information has pointed strongly to the structure of our brains as the target of the majority of recent mutations in our genome (107).

The cortex makes up over three quarters of the human brain's volume, and has shown the most rapid relative expansion of any part of the brain when compared to rodents (108). Interestingly, regardless of cortical area, radial units that extend from the ventricle to the cortical surface form integrated circuits, which receive widespread inputs, project to other cortical areas from the upper layers and transmit signals to subcortical areas from the deeper layers (19). Thus, increasing the processing power of our brains during evolution has at least partly been achieved in a similar way to that which we have used to develop computers: simply adding more transistors.

By adding slowly to an existing foundation, humanity has achieved feats that would be unimaginable to our ancestors. The central driver of this has been our ability to understand previous accomplishments, disseminate our advances and build upon their work. Although this cycle is mediated in human societies by our unique capability for writing and language, the process fundamentally resembles what life has achieved by utilizing the molecular properties of nucleic acids. Thus, it is important to remember that in the history of life on earth, there have been many false starts and dead ends. Similarly, there have been many human ideas that have gained large followings, but that have ultimately turned out to be unproductive. In order to avoid our own species becoming another truncated branch on the tree of life, it will be essential that we utilize the skills that got us to this point in order to find a sustainable way forward.

2 METHODS

2.1 MASSIVELY PARALLEL SEQUENCING

Modern sequencing techniques generate huge quantities of data by sequencing many short pieces of DNA in parallel. The Illumina platform begins with DNA fragments between 200 and 1000 base pairs being run over a specialized flow cell and binding to the short DNA molecules embedded within. By serially binding both ends of the fragments to be sequenced, priming their duplication and releasing one end, bridge amplification can then multiply individual molecules within a defined area to produce a spot (109).

A mix of primed individual nucleotides (dNTPs), where each nucleotide is labelled with a different fluorescent dye, is then spread over the flow cell. When free nucleotides are washed away, this leaves each spot fluorescing in a single color dependent upon what the next base in the sequence of the spotted fragments was. After a camera has taken a picture of each of the millions of spots, the fluorescent molecules are then quenched and the process is repeated to build up the sequence of each spot. When these sequences are aligned to a corresponding genome assembly, the resultant reads cluster together at the loci that gave rise to the sequencing sample, whether in gene bodies for RNA-sequencing or at CRMs for DNase-sequencing or ChIP-sequencing (109).

The novelty of this approach is not only in the amount of data generated and its cost efficiency, but in that by performing the sequencing reactions in parallel using the same primers, it has allowed the unselected sequencing of samples of unknown contents. To illustrate its effect on biological research, simply consider the near exponential growth in sequences deposited in public databases (110).

2.2 GENOME WIDE TRANSCRIPTION FACTOR AND CHROMATIN PROFILING

Our ability to understand the state of the chromatin landscape and transcription factor's binding profiles in vivo has taken massive leaps in recent years due to advances in sequencing technology. By performing classical experiments used to analyze the accessibility of DNA and occupancy of transcription factors, and using them as input to high throughput sequencing, we can now analyze their results across the entire genome.

Deoxyribonuclease I (DNase) is an enzyme that cleaves DNA sequences in a sequence unspecific manner. However, in order to do so, it requires unhindered access to both sides of the helix, and thus its activity is blocked when DNA is bound by transcription factors or histones. Thus, by applying DNase to a cell type of interest, and processing the resultant fragments for sequencing by ligating adapters to their ends, we can produce a sequencing library of DNA fragments that were DNase hypersensitive (111). Sequencing such fragments (DNase-sequencing) not only produces the least biased possible picture of the chromatin

landscape, but can also imply the binding of specific transcription factors within hypersensitive regions based on the slightly altered cleavage profile around their target sequences (112).

However, the only way to assign the binding of an individual transcription factor to a distinct target site is by using an antibody of known specificity to pull down the factor along with the target site it is bound to. Known as chromatin immunoprecipitation (ChIP), this technique requires cells to be fixed and the DNA sheared into fragments of sequencable length before the antibody pull down. By ligating adapters to the ends of these fragments and performing massively parallel sequencing (ChIP-sequencing), we can capture a picture of the regions bound by an individual transcription factor genome wide within a population of cells (113).

2.3 RNA AND PROTEIN EXPRESSION PROFILING

The activities of individual cells are defined by the genes and proteins that they express. The ability to simultaneously capture the relative expression of every gene expressed in a cell has led to a much deeper understanding of the networks involved in regulating cell states and how specific actions are accomplished. Due to the ability to amplify nucleic acid sequences, RNA expression profiling has taken great strides as sequencing technology has improved.

However, it is proteins that perform the actions of cells, and thus any information gleaned from RNA profiling must be confirmed on the protein level.

RNA-sequencing technology has become so sensitive that it is estimated we can currently capture and sequence up to 40% of transcripts present in a single cell (114). Most of these techniques rely on the poly-adenylation of protein coding RNAs, which allows for them to all be simultaneously converted to DNA in a single reaction, though the analysis of non-coding, short RNAs has also recently reached single cell sensitivity (115). One of the most important innovations in reaching such sensitive levels of detection has been the engineering of the transposase enzyme. Since this enzyme has the intrinsic ability to cut DNA and simultaneously ligate an associated, foreign DNA fragment to it, this allows for the efficient generation of a library of sequencable fragments in a single step (116). Thus, the ability to capture all of the genes expressed in a population, or in single cells, has given us a global view of gene regulation and transcriptional networks.

The ultimate output of RNA expression is on the protein level, where genes gain their ability to act within the cell. Although possible, it is much more difficult to gain a complete and unbiased picture of a cell's proteome than its transcriptome (117). However, there are many layers of gene regulation between RNA and protein, and thus it is essential that any conclusions made from RNA expression data be confirmed on the protein level (118). The most common way to analyze protein expression pattern within a tissue is by using immunohistochemistry. The ability to conjugate fluorophores that are excited at different frequencies in the visible part of the electromagnetic spectrum to different antibodies simultaneously on the same sample has allowed us to look at the co-expression, and even co-

localization of proteins *in vivo* (119). This technique is complemented by western blotting, whereby proteins are separated based on size using gel electrophoresis and visualized individually by antibodies to see their relative levels in different samples. This method is indispensable for assessing the enrichment of a protein following functional assays (120).

2.4 FUNCTIONAL ASSESSMENT OF PROTEIN FUNCTION

The activities of individual cells are defined by the genes and proteins that they express. The ability to simultaneously capture the relative expression of every gene expressed in a cell has led to a much deeper understanding of the networks involved in regulating cell states and how specific actions are accomplished. Due to the ability to amplify nucleic acid sequences, RNA expression profiling has taken great strides as sequencing technology has improved. However, it is proteins that perform the actions of cells, and thus any information gleaned from RNA profiling must be confirmed on the protein level.

RNA-sequencing technology has become so sensitive that it is estimated we can currently capture and sequence up to 40% of transcripts present in a single cell (114). Most of these techniques rely on the poly-adenylation of protein coding RNAs, which allows for them to all be simultaneously converted to DNA in a single reaction, though the analysis of non-coding, short RNAs has also recently reached single cell sensitivity (115). One of the most important innovations in reaching such sensitive levels of detection has been the engineering of the transposase enzyme. Since this enzyme has the intrinsic ability to cut DNA and simultaneously ligate associated, foreign sequences to it, this allows for the efficient generation of a library of sequencable fragments in a single step (116). Thus, the ability to capture all of the genes expressed in a population, or in single cells, has given us a global view of gene regulation and transcriptional networks.

The ultimate output of RNA expression is on the protein level, where genes gain their ability to act within the cell. Although possible, it is much more difficult to gain a complete and unbiased picture of a cell's proteome than its transcriptome (117). However, there are many layers of gene regulation between RNA and protein, and thus it is essential that any conclusions made from RNA expression data be confirmed on the protein level (118). The most common way to analyze protein expression pattern within a tissue is by using immunohistochemistry. The ability to conjugate fluorophores that are excited at different frequencies in the visible part of the electromagnetic spectrum to different antibodies simultaneously on the same sample has allowed us to look at the co-expression, and even co-localization of proteins *in vivo* (119). This technique is complemented by western blotting, whereby proteins are separated based on size using gel electrophoresis and visualized individually by antibodies to see their relative levels in different samples. This method is indispensable for assessing the enrichment of a protein following functional assays (120).

2.5 FUNCTIONAL ASSESMENT OF CRM ACTIVITY

In order to understand how specific CRMs affect their target genes, there are several techniques utilized to infer where in the embryo they are active, the affinity of specific transcription factors for them and how those transcription factors affect their activities. Although general functions can be ascribed to different factors, it is only in the specific context of individual enhancers that conclusive statements about their roles in biological processes can be made.

One of the first questions that genome wide transcription factor binding or chromatin accessibility data leads to is whether or not signals in these assays correspond to CRMs driving activation of their neighboring genes. Although electroporation can address where in the limited electroporated area a CRM is active (123), in order to assess activity throughout an embryo the manipulation must occur at an early developmental stage and be retained in every cell throughout development. Since fish eggs are large and develop externally, it is possible to inject special plasmids containing transposase recognition sites along with transpose mRNA directly into the single cell embryo. If the plasmid contains a CRM in proximity to a promoter and green fluorescent protein (GFP) reporter gene then, as the fish develops and the enhancer becomes active in specific cell types, distinct regions of the fish will fluoresce and reveal what tissues may utilize the enhancer (123).

ChIP-sequencing is readily capable of revealing the CRMs to which a transcription factor is bound *in vivo*. However, the resolution of this technique is not capable of revealing the specific target motifs within a sequence that the transcription factor binds, nor the relative affinity that the transcription factor binds to different motifs. In order to reveal such deep mechanistic detail, DNA oligonucleotides can be radioactively labelled, mixed with purified transcription factor protein and incubated. When these mixes are run on a gel, the migration of oligonucleotides bound by protein will be retarded and form a band running with the protein. The relative intensity of such bands reveals the association between the protein of interest and each oligonucleotide sequence (125).

However, even when binding to an individual motif has been established, an individual transcription factor's effect on neighboring target gene activity is still an open question. Particularly in the presence of specific partner factors, transcription factors can activate or repress gene transcription to varying degrees. In order to pick apart these interactions and their functional effects on gene transcription, CRM regions are often cloned next to a promoter and the reporter gene luciferase. When these plasmids are transfected into cell lines or *in vivo*, the amount of luciferase protein produced can be assayed by an enzymatic reaction generating light. Thus, when expression plasmids coding for transcription factors that target the CRM are co-transfected, it is possible to assess their functional effects on basal CRM activity (126).

3 AIMS, RESULTS AND DISCUSSION

3.1 PAPERS I AND II: SPATIAL PATTERNING OF GENE EXPRESSION IN THE ENDODERM AND NEURAL TUBE

3.1.1 Aims

Transcription factors are often broadly expressed in distal parts of the developing embryo, where they perform distinct and overlapping functions in each region where they are expressed. However, it has remained unclear to what extent the chromatin landscape and differential expression of partner factors specify gene expression through the activities of ubiquitously expressed transcription factors. In Papers I and II, we wished to understand how specific and overlapping gene expression patterns arise from differences in the activities of ubiquitous and specifically expressed transcription factors, and the chromatin landscapes of different cell types.

3.1.2 Results

In paper I, we performed ChIP-sequencing and RNA-sequencing on cortex, spinal cord, stomach and lung from E11.5 embryos. We find that the vast majority of SOX2 binding is cell type specific and associated with transcription factor binding motifs corresponding to factors specifically expressed in each organ. Moreover, while specific SOX2 binding reflects the distinct gene expression patterns of each region where it is detected, common SOX2 bound CRMs are found around genes involved in core stem cell processes such as proliferation and the suppression differentiation. We go on to show that CRMs bound by SOX2 specifically in the endoderm or neural tube drive expression specifically in those organs, while CRMs bound in both germ layers are expressed in both. However, we do observe occurrences of ectopic reporter expression in the neural tube of CRMs bound exclusively in the endoderm, and find that these correspond to regions of inaccessible chromatin in the neural tube. Finally, we show that SOX2 represses proliferation in both the spinal cord and stomach, likely due to the binding of shared CRMs involved in proliferation.

In paper II, we performed ChIP-sequencing, Prom1 sorted DNase-sequencing and Prom1 sorted RNA-sequencing on the spinal cord and cortex. We find that DNase hypersensitive sites (DHSs) are largely overlapping in these two regions, but that the relative accessibility of DHSs is predictive of differential expression in neighboring genes. We go on to show that SOX2 binding is mostly cell type specific, but that many specifically bound CRMs are within commonly accessible DHSs. Importantly, CRMs specifically bound by SOX2 in the spinal cord are enriched for HOXA9 binding motifs, while those in the cortex are enriched for LHX2 motifs, both of which can be found as footprints neighboring SOX-motifs in corresponding DHSs. Moving forward, we cloned CRMs commonly or specifically bound in cortex or spinal cord into GFP reporter vectors, and found them to drive expression only in the same region of the zebrafish embryo that SOX2 binding was detected. Moreover, CRM

reporter expression was found to require intact SOX2, LHX2 and HOXA9 motifs, and to even be repatterned to the inappropriate region of the neural tube when HOXA9 and LHX2 motifs were swapped. The co-factor dependence of gene expression was further confirmed by sequencing cortical progenitors following electroporation of HOXB6, as spinal cord genes were upregulated under these conditions, and found to be associated with HOXA9 motif containing CRMs bound by SOX2 in spinal cord. Finally, this led us to a model whereby the presence of specifying transcription factor motifs guide distinct SOX2 binding profiles, which in turn leads to differential CRM accessibility – the single most predictive factor in divergent gene expression patterns.

3.1.3 Discussion

The question of how specific gene expression is achieved by transcription factor interactions within a defined chromatin landscape has important implications to both development and health, as deregulation of gene expression is an underlying cause of many diseases (126-127). In paper I, we find that the chromatin landscape plays an essential role in specifying CRM activity, as several endoderm specific CRMs showed ectopic activity in the neural tube when removed from their endogenous chromatin environment. This contrasts with our findings in paper II, whereby the chromatin landscape was found to be much more similar than the SOX2 binding profile in spinal cord and cortex. This suggests that the more closely related two tissues are, the more similar their chromatin landscapes (27) and thus the more essential the proper patterning of specific transcription factors becomes to proper genes expression. This conclusion is supported by the finding that switching the specific transcription factor motifs in CRM reporters is capable of respecifying their activities, and that the misexpression of HOXB6 in the cortex upregulated spinal cord genes neighboring HOXA9 motif containing CRMs bound by SOX2 in the spinal cord. Thus, we arrive at a model whereby specific transcription factors guide ubiquitous ones to distinct CRMs, where these complexes increase CRM accessibility and the expression of neighboring genes. However, following lineage commitment, as between the endoderm and neural tube, CRM accessibility becomes sufficiently restricted as to override the activities of transcription factors.

Although we found that SOX2 binding is highly divergent in the different organs that we study in paper I, it has been repeatedly shown that SOX2 has shared stem cell activities in many of the systems it has been studied (128-130). The ability of SOX2 to suppress differentiation and maintain a stem cell state regardless of context suggests that it might do this via the same target genes bound in many cell types. In line with the role of SOX2 as a master regulator of the stem cell state (40, 96), we find that, despite binding only 232 genes in both the endoderm and neural tube, these genes include key regulators of the Wnt, Shh, Hippo and Notch pathways. Moreover, we find that many of these genes are directly involved in cell cycle regulation and that, by manipulating its levels using electroporation, SOX2 represses cell proliferation in the spinal cord and stomach. Thus, SOX2 maintains its shared functions in different stem cell populations by binding common genes involved in key stem cell processes.

3.2 PAPERS III AND IV: REGULATION OF STEM CELL PROLIFERATION IS ESSENTIAL TO TEMPORAL CELL FATE COMMITMENT

3.2.1 Aims

The mammalian cortex is a powerful system for studying cell fate commitment due to the impeccably timed sequential differentiation of neuronal subtypes and glia from a stem cell population that is maintained throughout life. However, since the cortical ventricular zone is a mixed population of differentiating cells, the mechanism of stem cell competence progression has remained controversial (131-134). As a core stem cell process, cell cycle regulation has a key role in stem cell maintenance and competence progression, though the mechanics that underpin this relationship have remained elusive. In papers III and IV, we wished to understand the mechanisms connecting stem cell maintenance to cell cycle regulation and temporal cell fate commitment.

3.2.2 Results

In paper III, we utilized single cell RNA-sequencing to analyze cells from E9.5, E11.5, E13.5, E15.5 and E18.5 in order to study stem cell competence progression from deep to upper layer neuronal fate production. Using a t-distributed neighbor embedding (tSNE) based adjacency matrix to cluster our cells, we charted neurogenesis through our single cells along two streams, corresponding to deep and upper layer neuronal fates. This structure allowed us to perform differential expression analysis between cell clusters of similar neurogenic maturation stage in order to find lineage specific genes, and to group these genes based on their shared expression patterns within our data set. This analysis revealed not only a multipotent group of E11.5 stem cells that clustered together with upper layer stem cells, but also that the defining characteristic of this relationship was their shared expression of genes involved in mitosis. In contrast, other E11.5 cells that fell further along the stream towards deep layer neuron commitment expressed higher levels of markers for the G1 and S-phases of the cell cycle.

In order to characterize the cell fate competence of these different groups of progenitors, we identified cell surface markers of each in our RNA-sequence data, fluorescence activated cell sorted (FACS) E11.5 progenitors based on their expression of these markers and differentiated them in vitro. Indeed, the cells that expressed markers of the multipotent group (*Hmnr*, *Gpc6*, *Ednrb* or high *Sox2* levels) were much more likely to be found in M-phase and to differentiate into upper layer neurons when compared to those expressing markers of the committed clusters (*Slc1a5*, *EfnA5* or low *Sox2* levels). This raised the question of whether cell cycle phase might be instructive to temporal cell fate decisions. To address this possibility, we overexpressed and knocked down *CCNB1/2* and *CCND1/2*, which license M- and G1-S phase progression, respectively. We found that the overexpression of *CCNB1/2* or knockdown of *CCND1* made progenitors assume an upper layer neuronal fate, while the overexpression of *CCND1* or knock down of *CCNB1/2* was capable of biasing neurogenesis

towards deep layer fates. Finally, bulk RNA-sequencing of our sorted populations and electroporated cells revealed the TGF β and Notch signaling pathways to be commonly upregulated in early progenitors that go on to produce upper layer neurons.

In paper IV, we used a subset of our cortex single cell RNA-sequence data set and immunohistochemistry to show that markers of active proliferation, including BrdU incorporation and phosphohistone 3 (PH3) staining, are negatively correlated with *Sox2*/SOX2 expression levels. Moreover, by manipulating SOX2 levels using in utero electroporation, we observed that its overexpression reduced proliferation, while knock down increased proliferation. Importantly, we also saw that *Sox2*/SOX2 levels negatively correlated with markers of differentiation, and that forcing differentiation by knocking down SOX2 or overexpressing NEUROG2 induced a transient increase in proliferation. In order to understand the mechanisms by which SOX2 suppressed proliferation and differentiation, we compared cortex SOX2 ChIP-sequencing profiles to genes differentially regulated following the overexpression or knockdown of SOX2. Genes negatively regulated by SOX2 were found to be enriched for cell cycle genes, and though several of these were able to increase BrdU incorporation upon overexpression, only CCND1 completely retained this capacity when co-electroporated with SOX2. As a potential downstream target, we found that the *Ccnd1* promoter was one of the most robustly occupied regions in the genome, contained several SOX motifs of varying affinity for SOX2 binding, and was repressed in a dose dependent fashion by SOX2 in reporter assays. Interestingly, we found this repressive capacity to be dependent on the off consensus SOX motifs in the promoter region, which were most weakly bound in EMSA assays, and that these were characteristic of CRMs around genes repressed by SOX2 genome wide.

Importantly, the *Ccnd1* promoter is a known target of the Wnt pathway and, while overexpression of stabilized β -catenin activated the promoter and cell proliferation, co-overexpression of SOX2 completely blocked this capacity. We found that it was this SOX2 function that required the on consensus sites in the *Ccnd1* promoter and that this relied on an interaction with, Wnt signal transducing, TCF/LEF transcription factors. Moreover, we saw that TCF/LEF motifs were associated with genes repressed by SOX2 genome wide when spaced between five and nine bases from a consensus SOX motif. Notably, we could show that SOX2 bound to the TCF/LEF co-repressor TLE1, and that the presence of SOX2 increased the association between LEF1 and TLE1 in a dose dependent fashion. Conversely, we found that not only did TLE1 repress the *Ccnd1* promoter and proliferation, but that this capacity required SOX2, and that a LEF1 protein lacking the ability to bind TLE1 could block SOX2's effects on proliferation and *Ccnd1* promoter activity.

3.2.3 Discussion

Temporal cell fate competence progression can only occur within uncommitted stem cell populations, many of which will eventually reside in adult tissues throughout life. Thus, the first major issue facing the analysis of this process is to identify the stem cells within the complex population of progenitors at varying levels of commitment. By using single cell

RNA-sequencing in paper III, we have taken one of the first genome wide pictures of this complexity and used it to identify mechanisms involved in stem cell fate progression. We find that stem cell maintenance and temporal competence progression are profoundly tied together in the core processes involved in cell cycle progression, and in paper IV we work out a direct mechanistic link between these.

There has been a debate in the field as to whether CUX2⁺ cells in the early cortical ventricular zone are fate restricted to generate neurons of the upper cortical layers (131-134). This would suggest that cortical stem cells do not temporally progress in their cell fate competence over time, but instead become committed to a specific fate very early in development and are then activated in a defined temporal order (131, 133). However, our data in paper III does not support this model, as even though we do find a population of early stem cells that shows a close molecular link to upper layer progenitors, these cells are still highly related to the other early progenitor populations, and express markers of deep layer lineages. Moreover, although these cells do show an increased competence for upper layer neurogenesis when we FACS sort them from the overall population and differentiate them, we find that they also produce many deep layer neurons as well. Thus, we prefer a model whereby this population is best characterized as the least differentiated stem cells in the ventricular zone, which retain a multipotent capacity for differentiation into multiple cortical lineages. This contrasts with other cell populations in the early cortical ventricular zone, which are unipotent and found at various stages of commitment to deep layer neurogenesis.

It was surprising to us to find that cell cycle phase specific genes were so prominent in segregating multipotent stem cells from unipotent progenitors in our single cell RNA-seq data in paper III. However, it is interesting to note that a link between CCNB1 and pluripotency maintenance in ESCs was also recently made using an unbiased screening approach (68), while in the same system, CCND1-3 activity has been shown to be involved in germ layer specification and differentiation by directly activating the genes responsible (69). Moreover, other cell cycle regulators, such as Rb and Cdc25, have also been implicated in stem cell maintenance and cell fate decisions (135, 136). These precedents suggest that similar mechanisms may be utilized during cortical development, though how they are then applied to this specific systems is an open question. Interestingly, when we sequenced cortical cells electroporated with CCNB1, we found that both Notch and TGF β pathway components were upregulated. This fits with roles in the simultaneous inhibition of differentiation and progression of cell fate competence, as TGF β was also identified to maintain pluripotency in the aforementioned study (68) and has been shown to perform this function in another neural stem cell system (137). Furthermore, these roles contrast with those of CCND1, which has been previously shown to induce commitment to differentiation in both the cortex and spinal cord (75, 138). Our data supports this conclusion, and suggests it may perform this function via Myc and chemokine pathways, both of which have been shown to induce differentiation in different neurogenic systems (139-140). Although these indirect mechanisms surely play a role in cyclin dependent regulation of cell fate commitment, the finding that CCND1 binds directly to specific target genes in ESCs suggests that there may also be mechanisms that

allow these factors to directly inhibit or execute context specific differentiation programs (69).

The ability of different cyclins to so powerfully influence stem cell maintenance suggests that regulating the expression of these factors must be incorporated into the transcriptional networks essential to these processes. SOX2 is amongst the best characterized stem cell factors, as its levels have been shown to be inversely correlated with differentiation in the ventral forebrain (98), it is a core component of the ESC pluripotency network (40) and its overexpression was key to the discovery of somatic cell reprogramming (99). Thus, it is surprising that the mechanisms by which SOX2 actually maintains stemness in different contexts remain largely uncharacterized. In paper IV, we find that, at high levels, SOX2 maintains stem cells in a slowly cycling state, while differentiation involves a transient burst of proliferation. The mechanistic insight that SOX2 binds off-consensus motifs to perform this function has precedent in the drosophila embryo, and likely has widespread implications in many cellular processes (141). Interestingly, quiescence is a well characterized property of adult and cancer stem cell populations, but a slow cell cycle has only recently been ascribed to stem cells in the embryo (142-143). As cancer stem cells have also been shown to upregulate pluripotency factors (144), the direct mechanistic link that we have found between these traits represents a potential pathway for identifying and targeting them within tumors. These findings also highlight the intrinsic contradiction in tackling cancer, as the cells that are most dangerous and difficult to kill are often a small population of slowly cycling cells that are largely unaffected by traditional treatments (145).

When considering the role of stem cells in cancer, it is important to look for commonalities within this diverse disease. Two of the main drivers of many cancers are mutations that hyperactivate the Wnt pathway (146) or inactivate the TP53 pathway (85). As discussed above, the signaling networks downstream of these pathways are intrinsically linked with cell cycle regulation and stem cell maintenance. Although many SOX proteins have been shown to influence the Wnt pathway, most of these studies have lacked mechanistic insight into how SOX proteins achieve their effects in a context specific manner (147). By demonstrating the HMG-domain dependent binding of SOX2 and TCF/LEF transcription factors, as well as the functional effect that SOX2 has on the transcriptional output of this complex, we uncover a DNA-motif that is likely to play important roles in diverse contexts, such as sex determination (146), cancer (128) and organ maintenance in diverse tissues (147, 149). Moreover, although we do not address TP53 directly in this work, it is interesting to note that the upstream ATM/ATR and CHEK1/2 kinases were also identified as maintaining pluripotency in the aforementioned screen (68). This network is linked to SOX2 via SOX5, SOX6 and SOX21, which are upregulated upon SOX2 overexpression, and have recently been shown to play an essential role in the activation of the P53 pathway under oncogenic stress (101). Although only a small component of a stem cell transcriptional network, SOX protein cross regulation is yet another example of their key role in balancing stemness and differentiation in multiple contexts. Thus, the Sox family of transcription factors is integrally linked with the core pathways of stem cell maintenance via multiple mechanisms, and fully

understanding their activities will be key to gaining insight into stem cell linked diseases such as cancer.

4 ACKNOWLEDGEMENTS

Ten years is a long time to learn about a subject, get to know people and for life to happen. My time at Ludwig Institute and Karolinska Institutet has been incredibly fulfilling, and I will not soon forget it. So many people have contributed to this process that it is difficult to thank them all here, but I will try, and I hope that everyone knows I have appreciated their help.

As scientific mentors go, I have been very lucky to have my supervisor, Jonas Muhr, throughout this process. He gave me a solid grounding in how to plan and build a scientific study, but also allowed me incredible freedom to explore new techniques and hypotheses. However, I have also learned a great deal from the other group leaders at Ludwig, Johan Holmberg, Rickard Sandberg, Thomas Perlmann, Qiaolin Deng, Jan Stenman and Susanne Schlisio, as watching so many gifted scientists start their independent careers and maintain their groups has been incredibly valuable.

Of course, I could not have accomplished much without the help of my collaborators past and present in the Muhr lab; Danijal Topcic, Maria Bergsland, Idha Kurtzotter, Susanne Klum, Magnus Sandberg, Magdalena Källström, Alexandra Karlen, Cecile Zaouter, Gaele Combeau, Bhumica Singla and Florie Reynaud. I have also been fortunate enough to get help from and work with many other people over the years, including Nigel Kee, Stuart Fell, Jens Magnusson, Maria Papathanou, Michael Ratz, Vilma Rraklli, Jose Dias, Chika Yokota, Mikael Huss, Olov Andersson, Diogo Castro, Omid Faridani, Michal Malewicz, Åsa Björklund, Daniel Ramsköld and the indispensable support staff at Ludwig and CMB. I have enjoyed spending time with many people during my PhD, and it is these friendships that will stay with me long after the degree is finished.

This unique opportunity to reflect on where I have arrived invokes deep consideration of how I got here and where I'm going. For one, I never even would have considered the possibility of living so far from home if it were not for my always outward looking parents, Patrick Hagey and Donna Wood, and the center I received from my friends and family in Canada. Moreover, the motivation to build a life in Stockholm and keep working towards my goals could not have been sustained without tireless support from the love of my life, Tove, and the kindness of my friends and family in Sweden. However, now that this chapter is closing and I look to new challenges closer to human health, it is my children, Edwin and Leah, that give me cause to work for a better future. Arriving at this point, I would like to thank all who have followed this mad pursuit and provided shoulders to stand on.

5 REFERENCES

1. Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F, Vitale L, Pelleri MC, Tassani S, Piva F, Perez-Amodio S, Strippoli P, Canaider S. An estimation of the number of cells in the human body. *Ann Hum Biol.* 2013 Nov-Dec;40(6):463-71.
2. Jaenisch R, Young R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell.* 2008 Feb 22;132(4):567-82.
3. Bergmann O, Zdunek S, Felker A, Salehpour M, Alkass K, Bernard S, Sjöström SL, Szewczykowska M, Jackowska T, Dos Remedios C, Malm T, Andrä M, Jashari R, Nyengaard JR, Possnert G, Jovinge S, Druid H, Frisén J. Dynamics of Cell Generation and Turnover in the Human Heart. *Cell.* 2015 Jun 18;161(7):1566-75.
4. Darwich AS, Aslam U, Ashcroft DM, Rostami-Hodjegan A. Meta-analysis of the turnover of intestinal epithelia in preclinical animal species and humans. *Drug Metab Dispos.* 2014 Dec;42(12):2016-22.
5. Spalding KL, Bergmann O, Alkass K, Bernard S, Salehpour M, Huttner HB, Boström E, Westerlund I, Vial C, Buchholz BA, Possnert G, Mash DC, Druid H, Frisén J. Dynamics of hippocampal neurogenesis in adult humans. *Cell.* 2013 Jun 6;153(6):1219-1227.
6. Iizuka H. Epidermal turnover time. *J Dermatol Sci.* 1994 Dec;8(3):215-7.
7. Goolam M, Scialdone A, Graham SJL, Macaulay IC, Jedrusik A, Hupalowska A, Voet T, Marioni JC, Zernicka-Goetz M. Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos. *Cell.* 2016 Mar 24;165(1):61-74.
8. Petropoulos S, Edsgård D, Reinius B, Deng Q, Panula SP, Codeluppi S, Reyes AP, Linnarsson S, Sandberg R, Lanner F. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell.* 2016 Sep 22;167(1):285.
9. Niakan KK, Han J, Pedersen RA, Simon C, Pera RA. Human pre-implantation embryo development. *Development.* 2012 Mar;139(5):829-41.
10. Garcia-Fernández J, D'Aniello S, Escrivà H. Organizing chordates with an organizer. *Bioessays.* 2007 Jul;29(7):619-24.
11. Solnica-Krezel L, Sepich DS. Gastrulation: making and shaping germ layers. *Annu Rev Cell Dev Biol.* 2012;28:687-717.
12. Murry CE, Keller G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell.* 2008 Feb 22;132(4):661-80.
13. Stemple DL. Structure and function of the notochord: an essential organ for chordate development. *Development.* 2005 Jun;132(11):2503-12.

14. Butler SJ, Bronner ME. From classical to current: analyzing peripheral nervous system and spinal cord lineage and fate. *Dev Biol.* 2015 Feb 15;398(2):135-46.
15. Fuchs E. Scratching the surface of skin development. *Nature.* 2007 Feb 22;445(7130):834-42.
16. Evseenko D, Zhu Y, Schenke-Layland K, Kuo J, Latour B, Ge S, Scholes J, Dravid G, Li X, MacLellan WR, Crooks GM. Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. *Proc Natl Acad Sci U S A.* 2010 Aug 3;107(31):13742-7.
17. D'Amour KA, Agulnick AD, Eliazar S, Kelly OG, Kroon E, Baetge EE. Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat Biotechnol.* 2005 Dec;23(12):1534-41.
18. Jones EG. Neurotransmitters in the cerebral cortex. *J Neurosurg.* 1986 Aug;65(2):135-53.
19. Custo Greig LF, Woodworth MB, Galazo MJ, Padmanabhan H, Macklis JD. Molecular logic of neocortical projection neuron specification, development and diversity. *Nat Rev Neurosci.* 2013 Nov;14(11):755-69.
20. Wilson SW, Houart C. Early steps in the development of the forebrain. *Dev Cell.* 2004 Feb;6(2):167-81.
21. Beccari L, Marco-Ferreres R, Bovolenta P. The logic of gene regulatory networks in early vertebrate forebrain patterning. *Mech Dev.* 2013 Feb;130(2-3):95-111.
22. Molyneaux BJ, Arlotta P, Menezes JR, Macklis JD. Neuronal subtype specification in the cerebral cortex. *Nat Rev Neurosci.* 2007 Jun;8(6):427-37.
23. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature.* 2001 Feb 15;409(6822):860-921.
24. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science.* 2004 Oct 22;306(5696):636-40.
25. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutuyavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE, Stamatoyannopoulos JA. The accessible chromatin landscape of the human genome. *Nature.* 2012 Sep 6;489(7414):75-82.

26. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012 Apr 11;485(7398):376-80.
27. Stergachis AB, Neph S, Reynolds A, Humbert R, Miller B, Paige SL, Vernot B, Cheng JB, Thurman RE, Sandstrom R, Haugen E, Heimfeld S, Murry CE, Akey JM, Stamatoyannopoulos JA. Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell*. 2013 Aug 15;154(4):888-903.
28. Chen T, Dent SY. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet*. 2014 Feb;15(2):93-106.
29. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res*. 2011 Mar;21(3):381-95.
30. Woodcock CL, Ghosh RP. Chromatin higher-order structure and dynamics. *Cold Spring Harb Perspect Biol*. 2010 May;2(5):a000596.
31. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. *Genes Dev*. 2014 Dec 15;28(24):2679-92.
32. Flyamer IM, Gassler J, Imakaev M, Brandão HB, Ulianov SV, Abdennur N, Razin SV, Mirny LA, Tachibana-Konwalski K. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*. 2017 Apr 6;544(7648):110-114.
33. Bonev B, Cavalli G. Organization and function of the 3D genome. *Nat Rev Genet*. 2016 Oct 14;17(11):661-678.
34. de Wit E, Bouwman BA, Zhu Y, Klous P, Splinter E, Verstegen MJ, Krijger PH, Festuccia N, Nora EP, Welling M, Heard E, Geijsen N, Poot RA, Chambers I, de Laat W. The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature*. 2013 Sep 12;501(7466):227-31.
35. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. A Phase Separation Model for Transcriptional Control. *Cell*. 2017 Mar 23;169(1):13-23. doi: 10.1016/j.cell.2017.02.007.
36. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013 Apr 11;153(2):307-19.
37. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet*. 2009 Apr;10(4):252-63.
38. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*. 2012 Sep;13(9):613-26.

39. Jolma A, Yin Y, Nitta KR, Dave K, Popov A, Taipale M, Enge M, Kivioja T, Morgunova E, Taipale J. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature*. 2015 Nov 19;527(7578):384-8.
40. Rizzino A. Concise review: The Sox2-Oct4 connection: critical players in a much larger interdependent network integrated at multiple levels. *Stem Cells*. 2013 Jun;31(6):1033-9.
41. Aksoy I, Jauch R, Chen J, Dyla M, Divakar U, Bogu GK, Teo R, Leng Ng CK, Herath W, Lili S, Hutchins AP, Robson P, Kolatkar PR, Stanton LW. Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO J*. 2013 Apr 3;32(7):938-53.
42. Fry CJ, Farnham PJ. Context-dependent transcriptional regulation. *J Biol Chem*. 1999 Oct 15;274(42):29583-6.
43. Perrimon N, Pitsouli C, Shilo BZ. Signaling mechanisms controlling cell fate and embryonic patterning. *Cold Spring Harb Perspect Biol*. 2012 Aug 1;4(8):a005975.
44. Filtz TM, Vogel WK, Leid M. Regulation of transcription factor activity by interconnected post-translational modifications. *Trends Pharmacol Sci*. 2014 Feb;35(2):76-85.
45. Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol*. 2013 Mar;14(3):153-65.
46. Crease DJ, Dyson S, Gurdon JB. Cooperation between the activin and Wnt pathways in the spatial control of organizer gene expression. *Proc Natl Acad Sci U S A*. 1998 Apr 14;95(8):4398-403.
47. Hausmann G, Bänziger C, Basler K. Helping Wingless take flight: how WNT proteins are secreted. *Nat Rev Mol Cell Biol*. 2007 Apr;8(4):331-6.
48. Gordon MD, Nusse R. Wnt signaling: multiple pathways, multiple receptors, and multiple transcription factors. *J Biol Chem*. 2006 Aug 11;281(32):22429-33.
49. MacDonald BT, Tamai K, He X. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev Cell*. 2009 Jul;17(1):9-26.
50. Henderson BR, Fagotto F. The ins and outs of APC and beta-catenin nuclear transport. *EMBO Rep*. 2002 Sep;3(9):834-9.
51. Clevers H, van de Wetering M. TCF/LEF factor earn their wings. *Trends Genet*. 1997 Dec;13(12):485-9.

52. Chodaparambil JV, Pate KT, Hepler MR, Tsai BP, Muthurajan UM, Luger K, Waterman ML, Weis WI. Molecular functions of the TLE tetramerization domain in Wnt target gene repression. *EMBO J.* 2014 Apr 1;33(7):719-31.
53. Weiss A, Attisano L. The TGFbeta superfamily signaling pathway. *Wiley Interdiscip Rev Dev Biol.* 2013 Jan-Feb;2(1):47-63.
54. Wrana JL, Attisano L, Cárcamo J, Zentella A, Doody J, Laiho M, Wang XF, Massagué J. TGF beta signals through a heteromeric protein kinase receptor complex. *Cell.* 1992 Dec 11;71(6):1003-14.
55. Macias MJ, Martin-Malpartida P, Massagué J. Structural determinants of Smad function in TGF- β signaling. *Trends Biochem Sci.* 2015 Jun;40(6):296-308.
56. Fishell G. Regionalization in the mammalian telencephalon. *Curr Opin Neurobiol.* 1997 Feb;7(1):62-9.
57. O'Leary DD, Chou SJ, Sahara S. Area patterning of the mammalian cortex. *Neuron.* 2007 Oct 25;56(2):252-69.
58. Chiang C, Litingtung Y, Lee E, Young KE, Corden JL, Westphal H, Beachy PA. Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature.* 1996 Oct 3;383(6599):407-13.
59. Ericson J, Muhr J, Jessell TM, Edlund T. Sonic hedgehog: a common signal for ventral patterning along the rostrocaudal axis of the neural tube. *Int J Dev Biol.* 1995 Oct;39(5):809-16.
60. Storm EE, Garel S, Borello U, Hebert JM, Martinez S, McConnell SK, Martin GR, Rubenstein JL. Dose-dependent functions of Fgf8 in regulating telencephalic patterning centers. *Development.* 2006 May;133(9):1831-44.
61. Li Y, Zhang H, Litingtung Y, Chiang C. Cholesterol modification restricts the spread of Shh gradient in the limb bud. *Proc Natl Acad Sci U S A.* 2006 Apr 25;103(17):6548-53.
62. Martí E, Bovolenta P. Sonic hedgehog in CNS development: one signal, multiple outputs. *Trends Neurosci.* 2002 Feb;25(2):89-96.
63. Stamatakis D, Ulloa F, Tsoni SV, Mynett A, Briscoe J. A gradient of Gli activity mediates graded Sonic Hedgehog signaling in the neural tube. *Genes Dev.* 2005 Mar 1;19(5):626-41.
64. Turner N, Grose R. Fibroblast growth factor signalling: from development to cancer. *Nat Rev Cancer.* 2010 Feb;10(2):116-29.
65. Oda K, Matsuoka Y, Funahashi A, Kitano H. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol Syst Biol.* 2005;1:2005.0010.

66. Fukuchi-Shimogori T, Grove EA. Neocortex patterning by the secreted signaling molecule FGF8. *Science*. 2001 Nov 2;294(5544):1071-4.
67. Garel S, Huffman KJ, Rubenstein JL. Molecular regionalization of the neocortex is disrupted in Fgf8 hypomorphic mutants. *Development*. 2003 May;130(9):1903-14.
68. Gonzales KA, Liang H, Lim YS, Chan YS, Yeo JC, Tan CP, Gao B, Le B, Tan ZY, Low KY, Liou YC, Bard F, Ng HH. Deterministic Restriction on Pluripotent State Dissolution by Cell-Cycle Pathways. *Cell*. 2015 Jul 30;162(3):564-79.
69. Pauklin S, Madrigal P, Bertero A, Vallier L. Initiation of stem cell differentiation involves cell cycle-dependent regulation of developmental genes by Cyclin D. *Genes Dev*. 2016 Feb 15;30(4):421-33.
70. Orford KW, Scadden DT. Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. *Nat Rev Genet*. 2008 Feb;9(2):115-28.
71. Bielas JH, Heddle JA. Proliferation is necessary for both repair and mutation in transgenic mouse cells. *Proc Natl Acad Sci U S A*. 2000 Oct 10;97(21):11391-6.
72. Rangel-Huerta E, Maldonado E. Transit-Amplifying Cells in the Fast Lane from Stem Cells towards Differentiation. *Stem Cells Int*. 2017;2017:7602951.
73. Turrero García M, Chang Y, Arai Y, Huttner WB. S-phase duration is the main target of cell cycle regulation in neural progenitors of developing ferret neocortex. *J Comp Neurol*. 2016 Feb 15;524(3):456-70.
74. Arai Y, Pulvers JN, Haffner C, Schilling B, Nüsslein I, Calegari F, Huttner WB. Neural stem and progenitor cells shorten S-phase on commitment to neuron production. *Nat Commun*. 2011 Jan 11;2:154.
75. Lange C, Huttner WB, Calegari F. Cdk4/cyclinD1 overexpression in neural stem cells shortens G1, delays neurogenesis, and promotes the generation and expansion of basal progenitors. *Cell Stem Cell*. 2009 Sep 4;5(3):320-31.
76. Harashima H, Dissmeyer N, Schnittger A. Cell cycle control across the eukaryotic kingdom. *Trends Cell Biol*. 2013 Jul;23(7):345-56.
77. Donjerkovic D, Scott DW. Regulation of the G1 phase of the mammalian cell cycle. *Cell Res*. 2000 Mar;10(1):1-16.
78. Harbour JW, Dean DC. The Rb/E2F pathway: expanding roles and emerging paradigms. *Genes Dev*. 2000 Oct 1;14(19):2393-409.
79. Dimova DK, Dyson NJ. The E2F transcriptional network: old acquaintances with new faces. *Oncogene*. 2005 Apr 18;24(17):2810-26.

80. Siu KT, Rosner MR, Minella AC. An integrated view of cyclin E function and regulation. *Cell Cycle*. 2012 Jan 1;11(1):57-64.
81. Yam CH, Fung TK, Poon RY. Cyclin A in cell cycle control and cancer. *Cell Mol Life Sci*. 2002 Aug;59(8):1317-26.
82. Gavet O, Pines J. Progressive activation of CyclinB1-Cdk1 coordinates entry to mitosis. *Dev Cell*. 2010 Apr 20;18(4):533-43.
83. Lindqvist A, van Zon W, Karlsson Rosenthal C, Wolthuis RM. Cyclin B1-Cdk1 activation continues after centrosome separation to control mitotic progression. *PLoS Biol*. 2007 May;5(5):e123.
84. Blow JJ, Tanaka TU. The chromosome cycle: coordinating replication and segregation. Second in the cycles review series. *EMBO Rep*. 2005 Nov;6(11):1028-34.
85. Maréchal A, Zou L. DNA damage sensing by the ATM and ATR kinases. *Cold Spring Harb Perspect Biol*. 2013 Sep 1;5(9).
86. Meek DW. Tumour suppression by p53: a role for the DNA damage response? *Nat Rev Cancer*. 2009 Oct;9(10):714-23.
87. Graña X, Reddy EP. Cell cycle control in mammalian cells: role of cyclins, cyclin dependent kinases (CDKs), growth suppressor genes and cyclin-dependent kinase inhibitors (CKIs). *Oncogene*. 1995 Jul 20;11(2):211-9.
88. Perry JA, Kornbluth S. Cdc25 and Wee1: analogous opposites? *Cell Div*. 2007 May 4;2:12.
89. Kopan R, Ilagan MX. The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell*. 2009 Apr 17;137(2):216-33.
90. Borggrefe T, Liefke R. Fine-tuning of the intracellular canonical Notch signaling pathway. *Cell Cycle*. 2012 Jan 15;11(2):264-76.
91. Kageyama R, Ohtsuka T. The Notch-Hes pathway in mammalian neural development. *Cell Res*. 1999 Sep;9(3):179-88.
92. Roztocil T, Matter-Sadzinski L, Alliod C, Ballivet M, Matter JM. NeuroM, a neural helix-loop-helix transcription factor, defines a new transition stage in neurogenesis. *Development*. 1997 Sep;124(17):3263-72.
93. Castro DS, Guillemot F. Old and new functions of proneural factors revealed by the genome-wide characterization of their transcriptional targets. *Cell Cycle*. 2011 Dec 1;10(23):4026-31.

94. Mattar P, Britz O, Johannes C, Nieto M, Ma L, Rebeyka A, Klenin N, Polleux F, Guillemot F, Schuurmans C. A screen for downstream effectors of Neurogenin2 in the embryonic neocortex. *Dev Biol.* 2004 Sep 15;273(2):373-89.
95. Bertrand N, Castro DS, Guillemot F. Proneural genes and the specification of neural cell types. *Nat Rev Neurosci.* 2002 Jul;3(7):517-30.
96. Holmberg J, Hansson E, Malewicz M, Sandberg M, Perlmann T, Lendahl U, Muhr J. SoxB1 transcription factors and Notch signaling use distinct mechanisms to regulate proneural gene function and neural progenitor differentiation. *Development.* 2008 May;135(10):1843-51.
97. Bylund M, Andersson E, Novitsch BG, Muhr J. Vertebrate neurogenesis is counteracted by Sox1-3 activity. *Nat Neurosci.* 2003 Nov;6(11):1162-8.
98. Hutton SR, Pevny LH. SOX2 expression levels distinguish between neural progenitor populations of the developing dorsal telencephalon. *Dev Biol.* 2011 Apr 1;352(1):40-7.
99. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.* 2006 Aug 25;126(4):663-76.
100. Sandberg M, Källström M, Muhr J. Sox21 promotes the progression of vertebrate neurogenesis. *Nat Neurosci.* 2005 Aug;8(8):995-1001.
101. Kurtzdotter I, Topcic D, Karlén A, Singla B, Hagey DW, Bergsland M, Siesjö P, Nistér M, Carlson JW, Lefebvre V, Persson O, Holmberg J, Muhr J. SOX5/6/21 Prevent Oncogene-Driven Transformation of Brain Stem Cells. *Cancer Res.* 2017 Sep 15;77(18):4985-4997.
102. Bergsland M, Werme M, Malewicz M, Perlmann T, Muhr J. The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes Dev.* 2006 Dec 15;20(24):3475-86.
103. Götz M, Huttner WB. The cell biology of neurogenesis. *Nat Rev Mol Cell Biol.* 2005 Oct;6(10):777-88.
104. Johnson MB, Wang PP, Atabay KD, Murphy EA, Doan RN, Hecht JL, Walsh CA. Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nat Neurosci.* 2015 May;18(5):637-46.
105. Kuratani S. Modularity, comparative embryology and evo-devo: developmental dissection of evolving body plans. *Dev Biol.* 2009 Aug 1;332(1):61-9.
106. Boffelli D, Nobrega MA, Rubin EM. Comparative genomics at the vertebrate extremes. *Nat Rev Genet.* 2004 Jun;5(6):456-65.
107. Jones B. Genome evolution: Functional antagonism and human brain evolution. *Nat Rev Genet.* 2012 Jul;13(7):452.

108. Swanson LW. Mapping the human brain: past, present, and future. *Trends Neurosci.* 1995 Nov;18(11):471-4.
109. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet.* 2010 Jan;11(1):31-46.
110. Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, Iyer R, Schatz MC, Sinha S, Robinson GE. Big Data: Astronomical or Genomical? *PLoS Biol.* 2015 Jul 7;13(7):e1002195.
111. Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc.* 2010 Feb;2010(2):pdb.prot5384.
112. Vierstra J, Reik A, Chang KH, Stehling-Sun S, Zhou Y, Hinkley SJ, Paschon DE, Zhang L, Psatha N, Bendana YR, O'Neil CM, Song AH, Mich AK, Liu PQ, Lee G, Bauer DE, Holmes MC, Orkin SH, Papayannopoulou T, Stamatoyannopoulos G, Rebar EJ, Gregory PD, Urnov FD, Stamatoyannopoulos JA. Functional footprinting of regulatory DNA. *Nat Methods.* 2015 Oct;12(10):927-30.
113. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet.* 2009 Oct;10(10):669-80.
114. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods.* 2013 Nov;10(11):1096-8.
115. Faridani OR, Abdullayev I, Hagemann-Jensen M, Schell JP, Lanner F, Sandberg R. Single-cell sequencing of the small-RNA transcriptome. *Nat Biotechnol.* 2016 Dec;34(12):1264-1266.
116. Chao MC, Abel S, Davis BM, Waldor MK. The design and analysis of transposon insertion sequencing experiments. *Nat Rev Microbiol.* 2016 Feb;14(2):119-28.
117. Larance M, Lamond AI. Multidimensional proteomics for cell biology. *Nat Rev Mol Cell Biol.* 2015 May;16(5):269-80.
118. Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* 2008 Jun 18;582(14):1977-86.
119. Stack EC, Wang C, Roman KA, Hoyt CC. Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis. *Methods.* 2014 Nov;70(1):46-58.
120. Mahmood T, Yang PC. Western blot: technique, theory, and trouble shooting. *N Am J Med Sci.* 2012 Sep;4(9):429-34.

121. LoTurco J, Manent JB, Sidiqi F. New and improved tools for in utero electroporation studies of developing cerebral cortex. *Cereb Cortex*. 2009 Jul;19 Suppl 1:i120-5.
122. Kaboord B, Perr M. Isolation of proteins and protein complexes by immunoprecipitation. *Methods Mol Biol*. 2008;424:349-64.
123. Uchikawa M. Enhancer analysis by chicken embryo electroporation with aid of genome comparison. *Dev Growth Differ*. 2008 Aug;50(6):467-74.
124. Birnbaum RY, Everman DB, Murphy KK, Gurrieri F, Schwartz CE, Ahituv N. Functional characterization of tissue-specific enhancers in the DLX5/6 locus. *Hum Mol Genet*. 2012 Nov 15;21(22):4930-8.
125. Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc*. 2007;2(8):1849-61.
126. Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, Kraft K, Kempfer R, Jerković I, Chan WL, Spielmann M, Timmermann B, Wittler L, Kurth I, Cambiaso P, Zuffardi O, Houge G, Lambie L, Brancati F, Pombo A, Vingron M, Spitz F, Mundlos S. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*. 2016 Oct 13;538(7624):265-269.
127. Lybæk H, de Bruijn D, den Engelsman-van Dijk AH, Vanichkina D, Nepal C, Brendehaug A, Houge G. RevSex duplication-induced and sex-related differences in the SOX9 regulatory region chromatin landscape in human fibroblasts. *Epigenetics*. 2014 Mar;9(3):416-27.
128. Sarkar A, Huebner AJ, Sulahian R, Anselmo A, Xu X, Flattery K, Desai N, Sebastian C, Yram MA, Arnold K, Rivera M, Mostoslavsky R, Bronson R, Bass AJ, Sadreyev R, Shivdasani RA, Hochedlinger K. Sox2 Suppresses Gastric Tumorigenesis in Mice. *Cell Rep*. 2016 Aug 16;16(7):1929-41.
129. Xie W, Lynch TJ, Liu X, Tyler SR, Yu S, Zhou X, Luo M, Kusner DM, Sun X, Yi Y, Zhang Y, Goodheart MJ, Parekh KR, Wells JM, Xue HH, Pevny LH, Engelhardt JF. Sox2 modulates Lef-1 expression during airway submucosal gland development. *Am J Physiol Lung Cell Mol Physiol*. 2014 Apr 1;306(7):L645-60.
130. Hagey DW, Muhr J. Sox2 acts in a dose-dependent fashion to regulate proliferation of cortical progenitors. *Cell Rep*. 2014 Dec 11;9(5):1908-20.
131. Gil-Sanz C, Espinosa A, Fregoso SP, Bluske KK, Cunningham CL, Martinez-Garay I, Zeng H, Franco SJ, Müller U. Lineage Tracing Using Cux2-Cre and Cux2-CreERT2 Mice. *Neuron*. 2015 May 20;86(4):1091-1099.
132. Guo C, Eckler MJ, McKenna WL, McKinsey GL, Rubenstein JL, Chen B. Fezf2 expression identifies a multipotent progenitor for neocortical projection neurons, astrocytes, and oligodendrocytes. *Neuron*. 2013 Dec 4;80(5):1167-74.

133. Franco SJ, Gil-Sanz C, Martinez-Garay I, Espinosa A, Harkins-Perry SR, Ramos C, Müller U. Fate-restricted neural progenitors in the mammalian cerebral cortex. *Science*. 2012 Aug 10;337(6095):746-9.
134. Eckler MJ, Nguyen TD, McKenna WL, Fastow BL, Guo C, Rubenstein JLR, Chen B. Cux2-positive radial glial cells generate diverse subtypes of neocortical projection neurons and macroglia. *Neuron*. 2015 May 20;86(4):1100-1108.
135. Inaba M, Yuan H, Yamashita YM. String (Cdc25) regulates stem cell maintenance, proliferation and aging in *Drosophila* testis. *Development*. 2011 Dec;138(23):5079-86.
136. Julian LM, Blais A. Transcriptional control of stem cell fate by E2Fs and pocket proteins. *Front Genet*. 2015 Apr 28;6:161.
137. Dias JM, Alekseenko Z, Applequist JM, Ericson J. Tgfb signaling regulates temporal neurogenesis and potency of neural stem cells in the CNS. *Neuron*. 2014 Dec 3;84(5):927-39.
138. Lukaszewicz AI, Anderson DJ. Cyclin D1 promotes neurogenesis in the developing spinal cord in a cell cycle-independent manner. *Proc Natl Acad Sci U S A*. 2011 Jul 12;108(28):11632-7.
139. Zinin N, Adameyko I, Wilhelm M, Fritz N, Uhlén P, Ernfors P, Henriksson MA. MYC proteins promote neuronal differentiation by controlling the mode of progenitor cell division. *EMBO Rep*. 2014 Apr;15(4):383-91.
140. Edman LC, Mira H, Erices A, Malmersjö S, Andersson E, Uhlén P, Arenas E. Alpha-chemokines regulate proliferation, neurogenesis, and dopaminergic differentiation of ventral midbrain precursors and neurospheres. *Stem Cells*. 2008 Jul;26(7):1891-900.
141. Ephrussi A, St Johnston D. Seeing is believing: the bicoid morphogen gradient matures. *Cell*. 2004 Jan 23;116(2):143-52.
142. Martynoga B, Mateo JL, Zhou B, Andersen J, Achimastou A, Urbán N, van den Berg D, Georgopoulou D, Hadjur S, Wittbrodt J, Ettwiller L, Piper M, Gronostajski RM, Guillemot F. Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. *Genes Dev*. 2013 Aug 15;27(16):1769-86.
143. Furutachi S, Miya H, Watanabe T, Kawai H, Yamasaki N, Harada Y, Imayoshi I, Nelson M, Nakayama KI, Hirabayashi Y, Gotoh Y. Slowly dividing neural progenitors are an embryonic origin of adult neural stem cells. *Nat Neurosci*. 2015 May;18(5):657-65.
144. Holmberg J, He X, Peredo I, Orrego A, Hesselager G, Ericsson C, Hovatta O, Oba-Shinjo SM, Marie SK, Nistér M, Muhr J. Activation of neural and pluripotent stem cell signatures correlates with increased malignancy in human glioma. *PLoS One*. 2011 Mar 31;6(3):e18454.

145. Zhou BB, Zhang H, Damelin M, Geles KG, Grindley JC, Dirks PB. Tumour-initiating cells: challenges and opportunities for anticancer drug discovery. *Nat Rev Drug Discov.* 2009 Oct;8(10):806-23.
146. Polakis P. Wnt signaling in cancer. *Cold Spring Harb Perspect Biol.* 2012 May 1;4(5). pii: a008052.
147. Kormish JD, Sinner D, Zorn AM. Interactions between SOX factors and Wnt/beta-catenin signaling in development and disease. *Dev Dyn.* 2010 Jan;239(1):56-68.
148. Tamashiro DA, Alarcón VB, Marikawa Y. Ectopic expression of mouse Sry interferes with Wnt/beta-catenin signaling in mouse embryonal carcinoma cell lines. *Biochim Biophys Acta.* 2008 Dec;1780(12):1395-402.
149. Kuwabara T, Hsieh J, Muotri A, Yeo G, Warashina M, Lie DC, Moore L, Nakashima K, Asashima M, Gage FH. Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nat Neurosci.* 2009 Sep;12(9):1097-105.