# Molecular Mechanisms of Gene Regulation in Response to Environmental and Developmental Stimuli

Dissertation
submitted to attain the academic degree
"Doctor of Natural Sciences"

at the Department of Biology
of the Johannes Gutenberg University Mainz

by

Sandra Schick

born on 19.01.1987 in Mainz

Mainz, 2016

Dekan:

1. Berichterstatter:
2. Berichterstatter:

Tag der mündlichen Prüfung:

# Table of Contents

# List of abbreviations

| | |
|---|---|
| 5caC | 5-carboxycytosine |
| 5fC | 5-formylcytosine |
| 5hmC | 5-hydroxymethylcytosine |
| 5hmU | 5-hydroxymethyluracil |
| 5mC | 5-methylcytosine |
| 6mA | N6-methyladenine |
| 6-4PP | pyrimidine 6-4 pyrimidone photoproduct |
| ΔCT | delta cycle threshold |
| A | adenine |
| ADAM | disintegrin and metalloproteinase domain-containing protein |
| ADP | adenosine diphosphate |
| AMPK | AMP kinase |
| AP1 | activator protein 1 |
| aRG | apical radial glial |
| ARNTL, BMAL1 | aryl hydrocarbon receptor nuclear translocator-like |
| ARX | homeobox protein ARX |
| ASCL1, MASH1 | achaete-scute homolog 1 |
| ATAC | assay for transposase-accessible chromatin |
| ATM | serine-protein kinase ATM, ataxia telangiectasia mutated |
| ATP | adenosine triphosphate |
| ATR | serine/threonine-protein kinase ATR, ataxia telangiectasia and Rad3-related protein |
| bHLH | basic helix-loop-helix |
| bIP | basal intermediate progenitor |
| BMI1 | polycomb complex protein BMI-1 |
| BMP | bone morphogenic protein |
| bp | base pairs |
| BrdU | bromdeoxyuridine |
| bRG | basal radial glial |
| C | cytosine |
| CA | cellular aggregates |
| $CaCl_2$ | Calcium chloride |
| CAF-1, CNOT7 | chromatin assembly factor 1, CCR4-NOT transcription complex subunit 7 |
| CBNLR | classification by nonlinear regression analysis |
| CBP, CREBBP | CREB-binding protein |
| CCG | clock-controlled genes |
| CCND1 | G1/S-specific cyclin-D1 |

| | |
|---|---|
| CDK4 | cyclin-dependent kinase 4 |
| CHD | chromodomain, helicase, DNA binding |
| ChIP | chromatin immunoprecipitation |
| CK1 | casein kinase 1 |
| CLOCK | circadian locomoter output cycles protein kaput |
| CNS | central nervous system |
| $CO_2$ | carbon dioxide |
| CP | cortical plate |
| CPD | cyclobutane pyrimidine dimer |
| CRY | cryptochrome |
| CT | circadian time |
| CTCF | transcriptional repressor CTCF, CCCTC-binding factor |
| d, D | day |
| DAMT-1 | DNA N6-methyl methyltransferase |
| DAPT | N-[N-(3,5-Difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester |
| DBP | DNA-binding protein or D site-binding protein |
| DDB | DNA damage-binding protein |
| DDR | DNA damage response |
| DFG | Deutsche Forschungsgemeinschaft |
| DLL1 | delta-like protein 1 |
| DMAD | DNA 6mA demethylase |
| DMEM | Dulbecco's modified Eagle's medium |
| DNA | deoxyribonucleic acid |
| DNA-PKcs | DNA protein kinase catalytic subunit |
| DNMT | DNA (cytosine-5)-methyltransferase |
| D-PBS | Dulbecco's phosphate-buffered saline |
| E9.5 | embryonic day 9.5 |
| EB | elution buffer |
| e.g. | for example |
| EMT | epithelial to mesenchymal transition |
| EMX | homebox protein EMX |
| Ep300 | histone acetyltransferase p300 |
| EpiR | epigenetic regulator |
| eRNA | enhancer RNA |
| ESC, ES | embryonic stem cells |
| EZH2 | histone-lysine N-methyltransferase EZH2, enhancer of zeste homolog 2 |
| FACS | Fluorescence-activated cell sorting |
| FAIRE | formaldehyde-assisted isolation of regulatory elements |

| | |
|---|---|
| FGF | fibroblast growth factor |
| Fos, c-Fos | proto-oncogene c-Fos |
| FOXG1, BF-1 | forkhead box protein G1 |
| FUS, TLS | RNA-binding protein FUS |
| G | guanine |
| GEO | gene expression omnibus |
| GFP | green fluorescent protein |
| GG-NER | global-genome nucleotide excision repair |
| GO | gene ontology |
| GSEA | gene set enrichment analysis |
| h | Hill coefficient or hour |
| HAT | histone acetyltransferase |
| HDAC | histone deacetylase |
| HEPES | hydroxyethyl-piperazineethane-sulfonic acid buffer |
| HES | transcription factor HES |
| HIRA | histone chaperone histone regulator A |
| HMGA2 | high mobility group protein HMGI-C |
| hnRNP-K | heterogeneous nuclear ribonucleo-protein K |
| HP1 | heterochromatin protein 1 |
| HP1BP3 | HP1-binding protein 3 |
| HR | homologous recombination |
| HULC | lncRNA highly upregulated in liver cancer |
| i.e. | id est, that is |
| IFT | intraflagellar transport |
| IFT74 | intraflagellar transport protein 74 |
| INO80 | inositol requiring 80 |
| input | input DNA |
| INSM1 | insulinoma-associated protein 1 |
| IP | immunoprecipitated DNA; immune precipitation |
| IPC | intermediate progenitors |
| IQR | interquartile range |
| ISMARA | integrated system for motif activity response analysis |
| ISWI | imitation switch |
| IZ | intermediate zone |
| JAG1 | protein jagged-1 |
| JMJD | jumonji domain containing protein |
| JNK | c-Jun N-terminal kinase |
| JTK | Jonckheere-Terpstra-Kendall |
| Jun, c-Jun | transcription factor AP-1 |

| | |
|---|---|
| KAT2A, GCN5 | histone acetyltransferase KAT2A |
| KDM | lysine demethylase |
| KDM6B, JMJD3 | lysine-specific demethylase 6B |
| KTM | lysine methyltransferase |
| LAD | lamina associated domains |
| LCR | locus control regions |
| LHX2 | LIM/homeobox protein Lhx2 |
| lincRNA | long intergenic non-coding RNA |
| lncRNA | long non-coding RNA |
| LSD | lysine specific demethylase |
| MAML | mastermind-like protein |
| MAPK | mitogen-activated protein kinase |
| MCC | Matthews correlation coefficient |
| MDC1 | mediator of DNA-damage checkpoint protein 1 |
| MeCP2 | methyl-CpG-binding protein 2 |
| MEF | mouse embryonic fibroblasts |
| miRNA | micro RNA |
| MLL | mixed lineage leukemia, histone-lysine N-methyltransferase |
| MRN | Mre11-Rad50-Nbs1 |
| mRNA | messenger RNA |
| Myt1l | myelin transcription factor 1-like protein |
| *n* | number |
| N | neuron |
| NaCl | sodium chloride |
| NAD | nucleolus associated domains |
| nBAF | neuron-specific BAF |
| NCBI | National Center for Biotechnology Information |
| ncRNA | non-coding RNA |
| NE, NPE | neuroepithelial (progenitor) cells |
| NECD | notch extracellular domain |
| NER | nucleotide excision repair |
| NFκB | nuclear factor NF-kappa-B |
| NGN, NEUROG | neurogenin |
| NHEJ | non-homologous end-joining |
| NICD | notch intracellular domain |
| NIH3T3 | fibroblast cells derived from *Mus musculus* |
| NMAD | N6-methyl adenine demethylase 1 |
| NMuMG | mammary glad cells derived from *Mus musculus* |
| npBAF | BAF complex of neuronal progenitors |

| | |
|---|---|
| NPC, NP | neural progenitor cell, neuronal precursor cell |
| NR | not responding, common double-positive class |
| NR1D1, REV-ERBα | nuclear receptor subfamily 1 group D member 1 |
| NR1D2, REV-ERBβ | nuclear receptor subfamily 1 group D member 2 |
| NR2E1, TLX | nuclear receptor subfamily 2 group E member1 |
| NSC | neural stem cell |
| NTC | non-targeting control |
| NuRD | nucleosome-remodeling deacetylase complex |
| ODE | ordinary differential equation |
| *P* | p-value |
| PAX6 | paired box protein Pax-6 |
| PcG | polycomb group |
| PCNA | proliferating cell nuclear antigen |
| PER | period circadian protein homolog |
| PIC | preinitiation complex |
| PIKK | phosphoinositide 3-kinase-related protein kinases |
| piRNA | Piwi-interacting RNA |
| Pou3f2, Brn2 | POU domain, class 3, transcription factor 2 |
| PRC2 | polycomb repressive complex 2 |
| PRMT | arginine methyltransferase |
| PTBP1 | polypyrimidine Tract-Binding Protein 1 |
| PWS | Prader-Willi syndrome |
| *q* | q-value |
| qPCR | quantitative polymerase chain reaction |
| R | replicate |
| RA | retinoic acid |
| RBPJ | recombining binding protein suppressor of hairless |
| REST, NRSF | repressor-element-1-silencing transcription factor |
| RG | radial glia |
| RHT | retinohypothalamic tract |
| RNA | ribonucleic acid |
| RNA Pol II | RNA polymerase II |
| ROR | nuclear receptor ROR, retinoid-related orphan receptor |
| ROS | reactive oxygen species |
| RPA | replication protein A |
| rRNA | ribosomal RNA |
| RT-qPCR | reverse transcription quantitative polymerase chain reaction |
| SCF | Skp1-Cullin-F-box protein |
| SCN | suprachiasmatic nucleus |

| | |
|---|---|
| s.e.m., SEM | standard errors of the mean |
| seq | sequencing |
| *Sey* | small eye |
| sFRP2 | secreted frizzled-related protein 2 |
| SHH | sonic hedgehog |
| shRNA | small hairpin RNA |
| siRNA | small interfering RNA |
| SIRT1 | $NAD^+$-dependent protein deacetylase sirtuin-1 |
| SMARCA2, BRM | probable global transcription activator SNF2L2 |
| SMARCA4, BRG1 | transcription activator BRG1 |
| snoRNA | small nucleolar RNA |
| snRNA | small nuclear RNA |
| SUV39H1 | histone-lysine N-methyltransferase SUV39H1 |
| SVZ | subventricular zone |
| SWI/SNF | switch/sucrose non-fermentable |
| T | thymine or Brachyury |
| TAD | topologically associating domain |
| TBP | TATA-box-binding protein |
| TBR1 | T-box brain protein 1 |
| TBR2, EOMES | T-box brain protein 2, eomesodermin homolog |
| Tcf4 | T-cell factor 4 |
| TC-NER, TCR | transcription coupled nucleotide excision repair |
| TEF | thyrotroph embryonic factor |
| TET | methylcytosine dioxygenase TET, ten-eleven translocation |
| TF | transcription factor |
| TFAP2C, AP2 γ | transcription factor AP-2 gamma |
| TFIID, TFIIH | transcription initiation factor IID, transcription initiation factor IIH |
| TGFβ | transforming growth factor beta |
| TN | terminally differentiated glutamatergic pyramidal neurons |
| TP53, p53 | cellular tumor antigen p53 |
| tRNA | transfer RNA |
| TSS | transcription start site |
| UTR | untranslated region |
| UV | ultraviolet (UVA: 315-380 nm; UVB: 280-315 nm; UVC: 190-280 nm) |
| VZ | ventricular zone |
| WT | wild-type |
| XP | xeroderma pigmentosum |
| XPC | DNA repair protein complementing XP-C cells homolog |
| ZT | Zeitgeber time |

# Abstract

Biological processes such as cellular differentiation or response to environmental stimuli require gene expression adaptations. Such plastic remodeling is achieved by the application of diverse mechanisms that involve transcriptional and epigenetic modulations. In order to gain deeper insights into the molecular changes, the expression and chromatin alterations occurring in response to UV irradiation, circadian rhythm and neuronal cell fate commitment were investigated in this thesis. To this end, molecular experiments and high-throughput sequencing techniques were applied along with bioinformatics and mathematical modeling analyses. These comprehensive investigations have revealed multiple, interconnected processes that regulate diverse gene expression variations upon exposure to stimuli. As a consequence of UV irradiation, the chromatin in fibroblasts underwent drastic alterations. This included a genome-wide loss of chromatin accessibility as well as a reorganization of the histone modification H3K27ac. These epigenetic variations could explain the observed expression changes of numerous genes. In the context of circadian rhythm, transcription factors and epigenetic regulators with a circadian expression pattern were identified. These factors represent putative regulators of the core clock network (e.g., ZFP28) or other genes that display a circadian expression (e.g., ZFP28 and LEO1). Furthermore, this thesis highlights a fate commitment role of PAX6 in neural progenitor cells, where it acts as a gatekeeper by directly or indirectly activating ectodermal genes and repressing genes that are important for other linages. Additionally, a novel PAX6 target, *Ift74*, was demonstrated to be required for *in vivo* neurogenesis. In summary, this thesis revealed diverse, especially chromatin-mediated mechanisms, which cells utilize to respond to stimuli. Moreover, it contributes towards the understanding of the underlying gene regulatory networks.

# Zusammenfassung

Biologische Prozesse, wie beispielsweise zelluläre Differenzierung oder Reaktion auf Umgebungsreize, erfordern Anpassungen der Genexpression. Solche plastischen Veränderungen werden mithilfe verschiedener Mechanismen einschließlich transkriptioneller oder epigenetischer Regulierung erzielt. Um tiefere Einblicke in die zugrunde liegenden molekularen Vorgänge zu gewinnen, wurden im Rahmen dieser Dissertation Expressions- und Chromatinveränderungen untersucht, die in Folge von UV-Bestrahlung, zirkadianer Rhythmik oder Festlegung des neuronalen Zellschicksals auftreten. Hierzu wurden sowohl molekularbiologische Methoden als auch Hochdurchsatz-Sequenziertechniken in Kombination mit bioinformatischen Analysen und mathematischen Modellierungen angewendet. Diese Untersuchungen offenbarten viele miteinander verbundene Prozesse, die reiz-induzierte Expressionsanpassungen regulieren. Als eine Folge von UV Bestrahlung veränderte sich das Chromatin in Fibroblasten drastisch. Dies umfasste eine genom-weite Verringerung der Chromatinzugänglichkeit und eine Umorganisation der Histonmodifizierung H3K27ac. Diese epigenetischen Veränderungen konnten auch die Genexpressionsvariationen, die aus der UV Bestrahlung resultierten, für zahlreiche Gene erklären. Im Rahmen der zirkadianen Rhythmik wurden Transkriptionsfaktoren und epigenetische Regulatoren mit zirkadianer Expression identifiziert. Diese Faktoren repräsentieren mögliche Regulatoren des zugrundeliegenden genetischen Netzwerk der zirkadianen Rhythmik (z.B. ZFP28) oder von anderen zirkadian exprimierten Genen (z.B. ZFP28 und LEO1). Ferner wurde die Rolle von PAX6 auf die Zellschicksalsveränderung von neuralen Vorläuferzellen aufgezeigt. PAX6 übernimmt hierbei eine Weichensteller-Funktion, indem dieser Transkriptionsfaktor direkt oder indirekt ektodermale Gene aktiviert und Gene, die wichtig für andere Entwicklungslinien sind, unterdrückt. Zusätzlich wurde gezeigt, dass *Ift74*, eines der neu identifizierten PAX6 regulierten Gene, für die *in vivo* Neurogenese wichtig ist. Zusammenfassend zeigt diese Dissertation die Vielfalt der insbesondere Chromatin-basierten Mechanismen, die Zellen nutzen, um auf Umgebungsreize zu reagieren. Zudem trägt sie zum Verständnis der zugrunde liegenden genregulierenden Netzwerke bei.

# Acknowledgements

Acknowledgements were deleted for privacy

# 1 Introduction

## 1.1 Gene regulatory mechanisms

### 1.1.1 The organization of the mammalian genome

Since the formation of the earth around 4.5 billion years ago, living organisms have emerged and gained more and more complexity over time. For instance, multicellular organisms have evolved several times such as the first multicellular eukaryote around one billion years ago [1, 2]. Mammals like humans are very complex, multicellular organisms consisting of more than 200 different cell types [3]. These cluster together to form organs, which fulfill different functions in the body. Even though the cell types and their functions are highly diverse, each cell contains nearly the same genetic information. This is stored in the sequence of the deoxyribonucleic acid (DNA), which represents the heritable material encoding the composition of an organism.

DNA was first isolated from the nuclei of eukaryotic cells by Friedrich Miescher in 1869, but proven to represent the genetic material only in the next century [4-6]. Deoxyribonucleic acid is composed of four nucleotides containing the nucleobases adenine (A), cytosine (C), guanine (G) and thymine (T), which are chemically connected to build a polymer. This chain forms a complex with an inverted, complementary strand of nucleotides, whereby the nucleobases project towards inside and form hydrogen bonds (A with T, C with G). These two nucleotide chains are twisted around each other to create a double helix [7]. This structure enables the replication of DNA in a semiconservative manner [8].

To fit the human genome containing around $3.2 \times 10^9$ base pairs (bp) or the mouse genome with around $2.6 \times 10^9$ bp (~ 1.7 m of DNA) in a 5 µm nucleus, the DNA is compacted with the help of specialized proteins called histones (Figure 1) [9]. The core unit, a nucleosome, is a wedge-shaped histone octamer surrounded by approximately 1.7 turns of 146 bp DNA [10, 11]. Each nucleosome consists of two copies of the four core histones: H2A, H2B, H3 and H4. The length of the linker DNA between the nucleosomes can vary (around 10-100 bp [12], the mean in vertebrates is around 35 bp [13]). DNA wrapped around nucleosomes leads to the so-called “beads-on-a-string”

structure. The DNA is further compacted with the help of other proteins, including histone H1 that binds to the linker DNA. This complex of DNA, proteins and associated ribonucleic acid (RNA) molecules is termed as chromatin. The three-dimensional folding of chromatin is largely determined by its primary structure (DNA sequence and nucleosome composition (see sections 1.1.3.2 and 1.1.3.3)) and is mainly a result of polymer physics, but also influenced by biochemical mechanisms such as local compaction, long-range interactions and anchoring to the nuclear scaffold (see sections 1.1.3.4, 1.1.3.6 and 1.1.3.7) [14, 15].

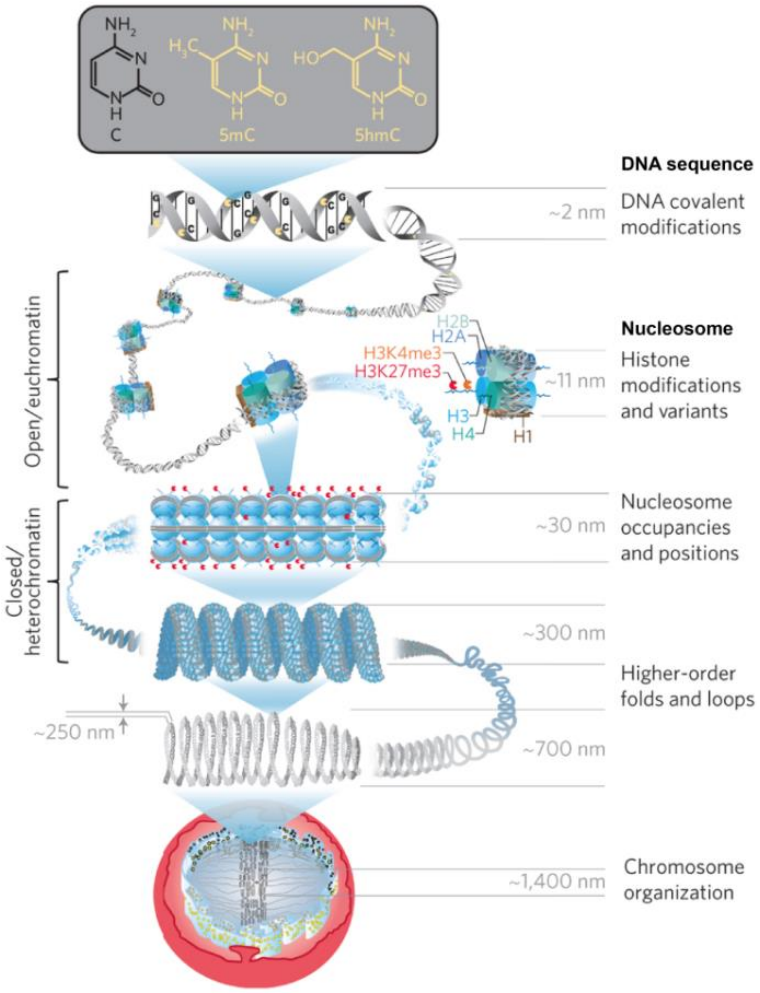



***Figure 1: Organization of the mammalian genome***

The DNA wraps around nucleosomes, which are composed of two copies of the histone H2A, H2B, H3 and H4. Histone H1 and other proteins regulate the further compaction of the genome in the nucleus. Chromatin can occur open (euchromatin) or closed (heterochromatin). Epigenetic features that modulate chromatin structure and function are indicated (see section 1.1.3). Source: adapted from [16]

Chromatin undergoes large structural changes throughout the cell cycle with its strongest compaction (~10,000- to 20,000-fold) during the metaphase of the mitosis and meiosis [13]. In 1928, Emil Heitz observed that chromatin occurs in different forms and introduced the binary division of chromatin into hetero- and euchromatin [17]. Euchromatin decondenses during interphase and contains regions that accommodate most transcriptional activity (see section 1.1.2). The latter is minimal in the compacted DNA of heterochromatic regions. Two types of heterochromatin have been described. Constitutive heterochromatin, mainly occurring at telomeres and centromeres, is condensed throughout the cell cycle and replicates usually later than euchromatin [18-21]. Facultative heterochromatin can decondense temporarily to allow the transcription

of genes, for example during certain developmental stages [22]. With the progress of chromatin research and the genome-wide mapping of chromatin features, more insights into the complex chromatin composition have been attained. These analyses have deciphered that there exist more fine structures and types of chromatin, since certain chromatin features cluster together in different combinations [23-25]. Their exact compositions and functions are currently explored.

### 1.1.2 The regulation of gene expression

In 1958, Francis Crick proposed the central dogma of molecular biology that describes the flow of the genetic information [26]. It states that DNA is transcribed into RNA, which is then translated into proteins that accomplish the majority of cellular functions [26]. As each cell contains nearly the same genetic information, the cellular identity is given by its set of genes being expressed or turned off at a given time. This expression status has not only to be established during development, but also needs to be maintained. Therefore, a tightly controlled temporal and spatial regulation of gene expression is essential for the proper development and physiology of different cell types. Furthermore, the gene expression needs to be flexible and reversible to allow the cell to react to stimuli such as environmental changes or stress exposure.

The DNA segments that encode functional RNAs or proteins are called genes. A gene consists of the coding regions (exons) mostly separated by non-coding regions (introns) and is flanked by regulatory sequences. Latter are named the 5′ and 3′ untranslated regions (UTR), which harbor information regulating the messenger RNA (mRNA) stability, localization and translation into proteins [27, 28]. Usually around the transcription start site (TSS) is the promoter region with specific sequence elements to which regulatory proteins such as transcription factors and RNA polymerases, the enzymes transcribing the genes, bind to control and initiate the transcription of the genes [29]. Furthermore, there are genomic regions distal from the genes that can influence the expression of a gene (see section 1.1.3.6).

Internal and external signals enable gene expression programs that regulate which genetically encoded information is used to give rise to the cellular phenotypes. As the expression of a certain set of genes is essential for the cellular identity including proper structure and function, gene expression is a tightly controlled multilayer process (Figure 2). Initially, specific as well as general transcription factors bind to the DNA by recognizing specific DNA sequence motives [29]. Binding of so-called pioneer transcription factors, which are able to bind to their target sequences on nucleosome, can precede the binding of the other transcription factors (TFs) to mediate the accessibility of the regions for the other factors [30, 31]. The transcription factors recruit the RNA polymerases as well as other regulatory proteins to the promoter of the genes and build the transcription initiation complex (Figure 3) [29]. RNA polymerase II transcribes the protein-coding as well as some non-coding genes (see section 1.1.3.5), while the structurally different RNA polymerases I and III are transcribing non-coding genes [32].

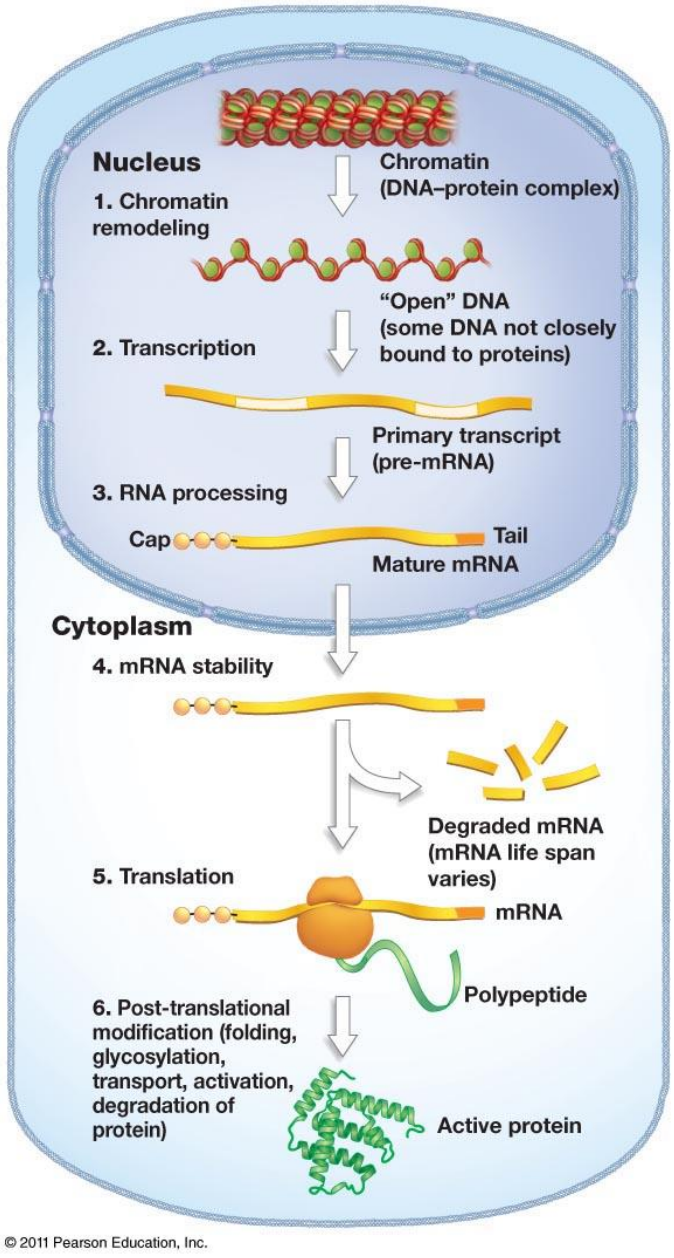



***Figure 2: The multistep process of gene expression***
Chromatin undergoes certain modifications (1.) to allow the transcription of the genes by RNA polymerases, which is coordinated by the binding of transcription factors (2). The primary transcript (pre-mRNA) is then further processed (3.) and translocated to the cytoplasm, where the transcript is translated into an amino acid chain that folds into a protein (5.). The functionality of proteins can further be modulated by post-translational modifications, which can influence their activity or degradation (6.). Source: http://www.uic.edu/classes/bios/bios100/lectures/18_01_gene_expression-L.jpg

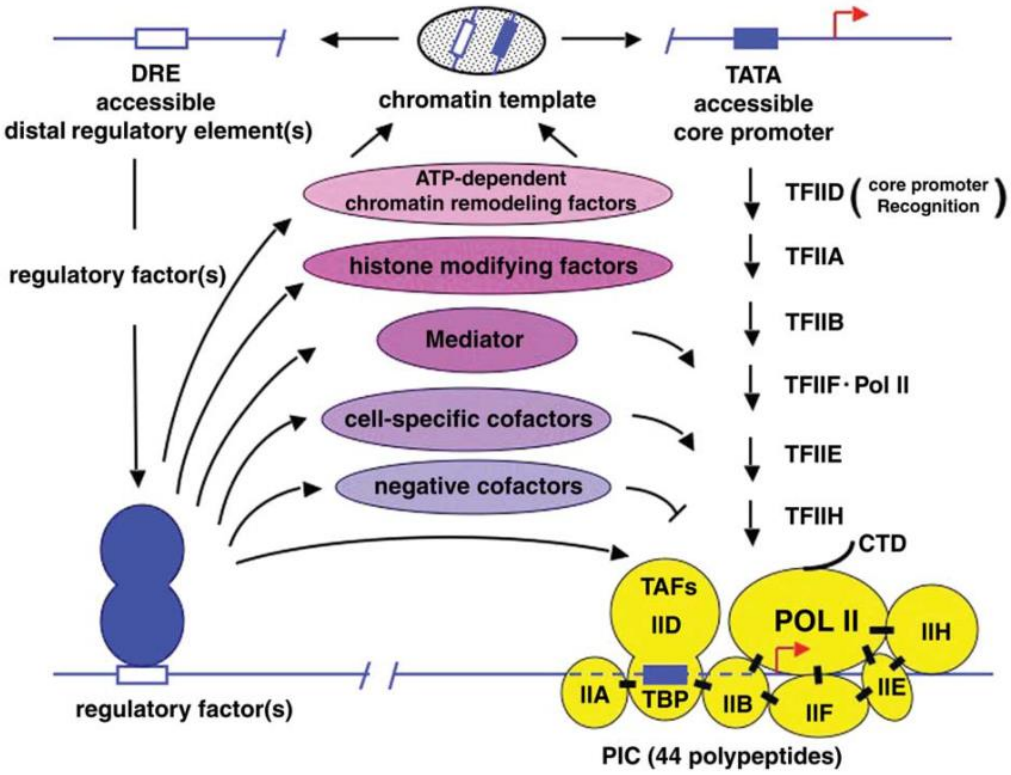



***Figure 3: Establishment of the RNA polymerase II preinitiation complex***

Transcription factors (blue) bind to distal regulatory regions and can recruit further cofactors such as chromatin modifiers (purple). TFIID bind to the TATA element in the promoter and initiates the binding of the other general transcription factors and the RNA polymerase II that results in the formation of the preinitiation complex (PIC). Source: [33]

When the paused RNA polymerase is released, the transcription elongation process starts, whereby the gene is transcribed into a complementary RNA molecule by the RNA polymerase [34]. This pre-mRNA is further processed by removing intron sequences and ligating exons together [35]. This process can result in different transcript isoforms when different splice sites are utilized (alternative splicing). Furthermore, the ends of eukaryotic transcripts are modified by capping the 5'-end and in many cases by adding a poly-A tail at the 3'-end [36]. In case of protein-coding genes, this mature mRNA is transported to the cytoplasm. The coding region of the mRNA is then translated into a protein. This is based on the genetic code, where triplets of nucleotides (called codons) code either for an amino acid or for a signal of translation initiation or termination [37, 38]. Ribosomes with the aid of transfer RNAs (tRNAs) execute this conversion of the nucleotide sequence into an amino acid sequence [39]. The amino acid sequences fold into functional proteins, which can further be chemically modified in order to influence their function and stability.

### 1.1.3 Epigenetic mechanisms regulate gene expression

In 1942, Conrad Hal Waddington introduced the term "epigenetics" for the "branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being" [40-42]. As with time more insights into the mechanisms were achieved, the definition of epigenetics was adjusted. In 1958, David Nanney suggested that the expression potential of the genes could persist through cell divisions by some stable epigenetic mechanisms and Robin Holliday later on included heritability as a defining feature of epigenetics in 1994 [43, 44]. The currently most widely used definition of epigenetics describes the study of mechanisms that cause, independent of alterations in the DNA sequence, mitotically and/or meiotically heritable changes, which introduce variations in the gene expression or function and thereby the phenotype [41, 45, 46]. Nevertheless, the term epigenetics is often also applied to chromatin changes influencing gene expression, which are not necessarily inherited.

Many epigenetic mechanisms have evolved to modulate chromatin, which will be further described in the following sections (Figures 1 and 4). These mechanisms influence in a combinatorial manner the spatial, temporal and quantitative control of the gene expression and consequently the cellular phenotype. These "epigenetic" mechanisms can, however, also be subjected to genetic variations. Several studies have indicated that genetic variations impact on transcription factor binding, which in turn alters histone modifications (see section 1.1.3.3) and enhancer choice (see section 1.1.3.6) and can (not necessarily) lead to gene expression and phenotypic changes [47-52]. The importance of epigenetic mechanisms is also visible considering that many diseases are induced, for example, by mutations in epigenetic control regions or genes encoding for epigenetic regulators [53, 54].

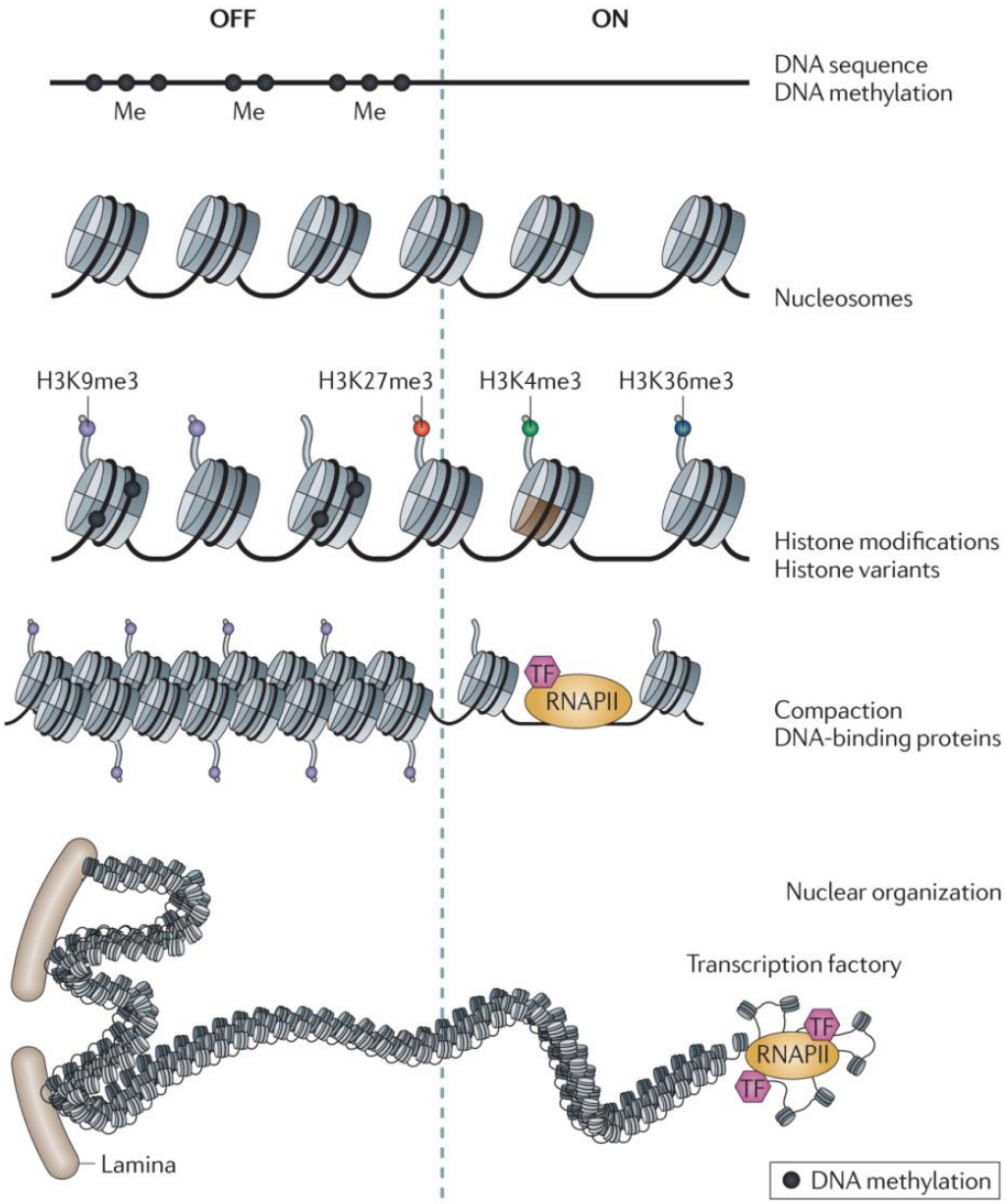



***Figure 4: Possible epigenetic mechanisms that influence the transcriptional status of a gene***

Chromatin is a complex structure whose composition and functional state can be modified on multiple layers: DNA modifications, histone variants and post-translational modifications, chromatin remodeling, binding of RNA and proteins and three-dimensional organization. These alterations influence if a gene is transcriptionally inactive (off) or active (on). Source: [55]

### 1.1.3.1 DNA modifications

DNA consists of only four bases; however, they can be chemically modified to influence diverse biological processes including heritable gene expression changes. One such abundant modification in mammals is the methylation at the fifth carbon position of cytosine (5-methylcytosine, 5mC) (Figures 1 and 5) [56, 57]. The modification is

especially enriched at CpG dinucleotides, where the methylation on both strands offers a mechanism for semiconservative propagation through the DNA replication [58-61]. The DNA (cytosine-5)-methyltransferase 1 (DNMT1) recognizes hemi-methylated DNA and methylates the cytosine on the newly synthesized strand [62]. While DNMT1 is important for the propagation and maintenance of DNA methylation, *de novo* methylation is established by the DNA (cytosine-5)-methyltransferases 3A and 3B (DNMT3A and DNMT3B) [62, 63]. CpGs are usually found methylated with the exception of specific CpG-rich regions, named CpG islands, that occur in active regulatory regions [64, 65]. 5mC is mainly known for repressing gene expression, especially when it occurs at promoters or enhancers, while in certain contexts (e.g., 5mC in gene bodies) it is also associated with active transcription [66, 67]. The effect of 5mC might result from its interference on transcription factor binding or the recruitment of factors that mediate modulations of chromatin features [57, 65, 68, 69]. Besides the influence of DNA methylation on gene regulation or as a consequence, DNA methylation has been shown to play a role in development and cell fate specification, genomic imprinting, X-chromosome inactivation and transposon control; furthermore, missregulated methylation patterns have been ascribed to cancer [57, 70-76].

5mC has been considered as a stable epigenetic mark until the discovery of the Methylcytosine dioxygenase TET (TET, Ten-Eleven Translocation) that can actively demethylate DNA (Figure 5) [77]. The first metabolic product of the demethylation process is 5-hydroxymethylcytosine (5hmC), which itself is relatively stable and might also be considered as an epigenetic mark [78, 79]. 5hmC might play a role in transcription regulation at promoters, gene bodies or enhancers [80, 81]. For example, the levels of 5hmC increase during neuronal differentiation and 5hmC is enriched at intragenic regions of many active neuron-specific genes [82-85]. 5hmC has been found in diverse cell types with higher enrichment in neuronal cells [83, 86, 87]. 5hmC can be further oxidized by TET enzymes to 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) [88, 89]. These modifications are very low abundant, but have been shown to influence gene expression and the DNA damage response [57, 90].

Also other DNA modifications can occur and may accomplish specific functions. For example, TET proteins can also oxidize thymine to 5-hydroxymethyluracil (5hmU) and furthermore, N6-methyladenine (6mA) has been found in eukaryotes (Figure 5) [57].

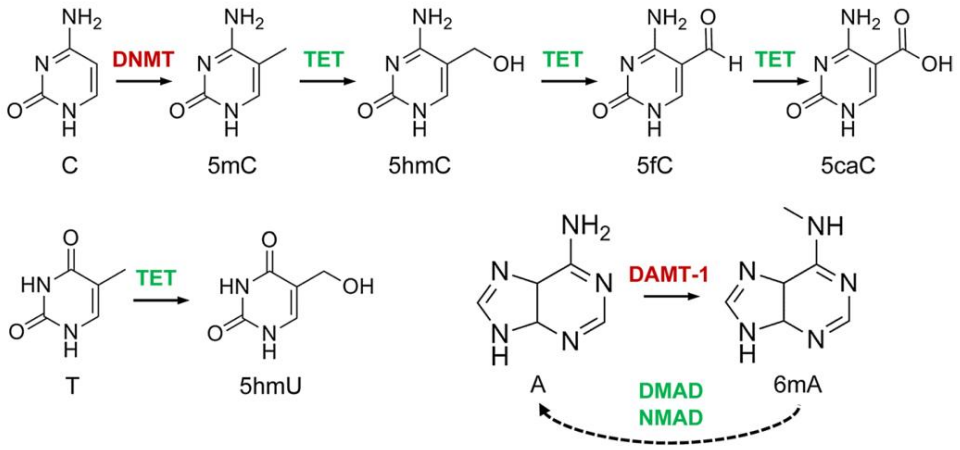



***Figure 5: Chemical modifications of DNA***

DNA methyltransferases (DNMT) can methylate cytosine (C) at the 5th carbon position (5mC), which can then be further oxidized to 5hmC, 5fC and 5caC by methylcytosine dioxygenases (TET). TET can also oxidize thymine (T) to 5-hydroxyuracil (5hmU). N6 of adenine (A) can be methylated by DNA N6-methyl methyltransferase (DAMT-1) to 6mA, which can be reverted by the demethylases DNA 6mA demethylase (DMAD) or N6-methyl adenine demethylase 1 (NMAD). Source: [57]

### 1.1.3.2 Histone variants

There is a high demand of histones when DNA is replicating during the S-phase of the cell cycle. In order to accomplish the synthesis of around $10^8$ molecules of each of the core histone proteins during this short period, several genes encode for the canonical histones and are organized in one large and two smaller clusters [91]. The canonical histones are primarily synthesized and deposited into chromatin during the S-phase of the cell cycle and encode RNAs that contain a short 3' stem-loop instead of a poly-A tail and no introns, probably to achieve their fast expression [91, 92].

Non-allelic isoforms of the canonical histone proteins exist and are named non-canonical histones or histone variants. These histone variants are synthesized throughout the cell cycle in smaller amounts and often incorporated independent of the replication into chromatin with the help of chaperones and ATP-dependent chromatin remodeling factors (see section 1.1.3.4) to substitute their canonical counterpart or to fill their places after histone eviction [93, 94]. The mRNAs of histone variants contain mostly the conventional poly-A tails and often introns; the latter increase the variety of histones due to possible different isoforms by alternative splicing [12].

Histone variants contain sequence variations and can therefore change the properties of nucleosomes and consequently of the chromatin structure [12, 95]. They also extend

the variability of nucleosome composition by increasing the number of possible combinations of canonical and non-canonical histones inside one nucleosome.

Many histone variants are highly conserved (e.g., H2A.Z and CenH3), which might reflect their important cellular functions, while others evolve rapidly in order to adapt fast and to fulfill tissue-/cell-specific transcriptional roles [15]. Furthermore, the replication independent synthesis of histone variants allows their incorporation into chromatin in response to environmental stimuli [96].

Histone variants are shown to play critical roles in diverse biological processes including transcription regulation, chromosome segregation, DNA repair and recombination, meiotic recombination, chromatin remodeling, germline-specific DNA packaging and activation, extra-nuclear acrosomal function and regulation of cellular/ developmental plasticity [97, 98].

H2A.Z is an essential and relatively conserved histone variant with ~60% similarity to H2A (Figure 6) [99]. The functions of this variant are versatile [94, 96]. For example, it has been described in gene activation as well as in heterochromatic silencing [100-103].

H2A.X can be expressed with the properties of a canonical histone during S-phase and polyadenylated during G1 phase, ensuring an even expression throughout the cell cycle [104]. It represents around 2-25% of the total H2A in mammalian cells [105]. H2A.X has diverse biological functions, but is mainly known for its role in the DNA damage response, when its distinctive C-terminal sequence becomes phosphorylated around the sites of DNA double-strand breaks (therefore it is also named the "histone guardian of the genome") (see sections 1.1.3.3 and 1.2.3) [106-109].

MacroH2A is a large histone variant due to an additional fold domain in its C-terminal tail (Figure 6) [110]. It has been implicated in transcription regulation and is found to be concentrated in the mammalian inactive X chromosome [111, 112].

The two most studied H3 variants are H3.3 and CenH3. H3.3 has only four amino acid substitutions compared to the canonical H3, nevertheless, this leads to crucial changes in its properties [113-115]. H3.3 is often found in nucleosomes of actively transcribed genes as well as in active regulatory regions associated with H2A.Z containing nucleosomes (Figure 6) [116, 117]. CenH3 (CENPA) is a histone H3 variant occurring at centromeric regions and is, for example, essential for the assembly of the kinetochore complex (Figure 6) [118]. The DNA around nucleosomes containing CenH3 is wrapped

right-handed instead of the usual left-handed orientation [119].

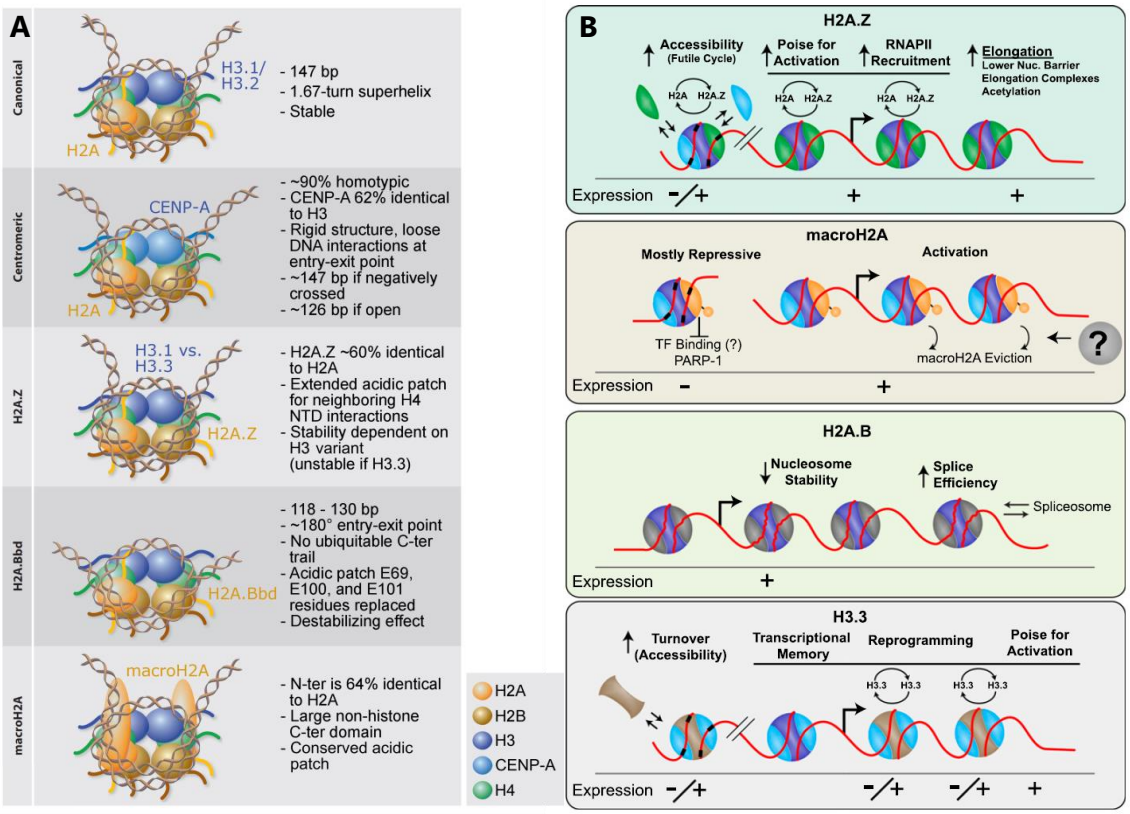



***Figure 6: The influence of histone variants on nucleosome structure and gene expression***

A) Properties of different histone variants and their influence on nucleosome structure. B) Models illustrating the deposition of certain histone variants into chromatin and their impact on transcriptional regulation (-/+: negative/positive influence on expression). Source: adapted from [95, 120]

### 1.1.3.3 Post-translational histone modifications

The variety of nucleosome composition is not only increased by the incorporation of histone variants, but also by post-translational modifications of specific amino acids of the histones. These chemical modifications are so far mostly described at the N-terminal tail of histones, which protrude outside of the nucleosome octamer resulting in high flexibility and accessibility, but can also occur at the globular histone core [15, 121]. Specific amino acids can, for example, be acetylated, phosphorylated, methylated, ubiquitinylated or sumoylated [121].

Histone modifications have not only been shown to be implicated in the regulation of transcription, but also in many other biological processes such as DNA repair or replication [122]. Some modifications influence the chromatin structure directly due to altered electrostatic or hydrophobic interactions with the DNA or chromatin-associated proteins [123]. On the other hand, different combinations of histone modifications at

one or several histones can recruit specific regulatory effector molecules that change the functional state of the surrounding chromatin and consequently influence cellular processes [124, 125]. This "histone code" postulated by Brian D. Strahl and C. David Allis constitutes another epigenetic mechanism regulating gene expression.

The occurrence and effect of histone modifications is controlled by the combinatorial activity of proteins termed histone "writers", "readers" or "erasers" [126]. "Writers" or "erasers" modify the histones by adding chemical modifications or by removing them. These histone-modifying enzymes are recruited via diverse molecules such as site-specific DNA-binding factors, co-activators and co-repressors, RNA polymerase II, preceding histone modifications (called histone crosstalk) or long non-coding RNAs [15, 127-139]. "Readers" recognize the occurrence or absence of specific modifications at distinct amino acids with special binding domains, whereby neighboring post-translational modifications on the same or a different nucleosome can impact on the binding [140-142]. Readers mostly recruit further regulatory factors, which mediate the biological outcome of the epigenetic mark such as gene expression modulation. The recruited factors can be chromatin remodelers, chromatin modifiers, architectural proteins or components of machineries that mediate biological processes such as transcription, DNA damage repair or replication (Figure 7) [141].

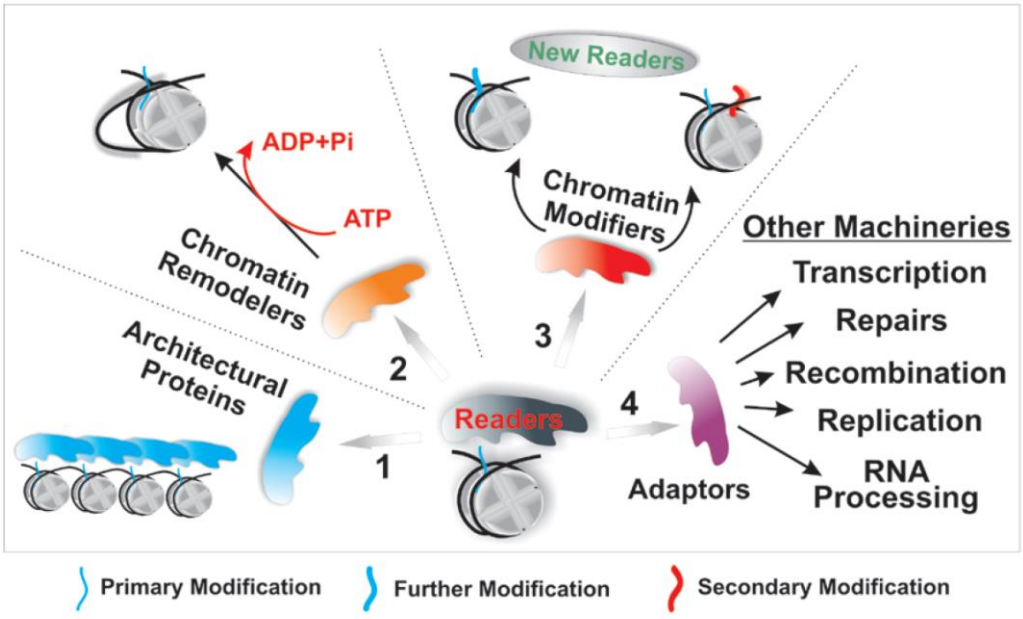



***Figure 7: The reading of modified histones can result in different functional outcomes***

Readers interact with different kind of proteins such as architectural proteins (1), chromatin remodelers (2), chromatin modifiers (3) and other adaptor proteins that coordinate different biological processes (4). Readers mediate via these interaction partners the functional outcomes of the histone modification. For example, they can influence the transcriptional output of a gene. Source: [141]

Mono-, di- or trimethylation of lysine residues is mediated by Lysine methyltransferases (KTMs), while Arginine methyltransferases (PRMTs) lead to mono- and symmetric or asymmetric dimethylation of arginine residues [143, 144]. S-adenosyl-methionine often serves as the methyl donor [143, 144]. Methylation is removed by Lysine specific demethylases (LSDs) or Jumonji domain containing proteins (JMJDs) (KDMs) [145, 146]. Histone methylation does not change the charge of the residue in contrast to histone acetylation or phosphorylation [122]. Histone methylation can result, dependent on the methylated histone residue, in active or inactive chromatin. For example, in the absence of repression-associated histone modifications are trimethylation of H3 lysine 4 (H3K4me3) at the transcriptions start site or of H3 lysine 36 (H3K36me3) within gene bodies associated with active gene transcription [147, 148]. Trimethylation of lysine 27 of H3 (H3K27me3), lysine 9 of H3 (H3K9me3) or lysine 20 of H4 (H4K20me3) denote inactive, heterochromatic chromatin regions [148, 149]. H3K27me3, deposited by the Polycomb group (PcG) proteins, marks facultative heterochromatin [150-152]. H3K9me3 and the catalyzing Histone-lysine N-methyltransferase Suv39h lead to the binding of the Heterochromatin protein 1 (HP1), which might then recruit the Histone-lysine N-methyltransferase Suv420H enzymes that catalyze the trimethylation of H4K20 to establish constitutive heterochromatin [153-155].

Histone acetylation, mediated by Histone acetyltransferases (HATs), causes the loss of the positive charge of the lysines and thereby weakens the interactions of the histones with the negatively charged DNA in their surrounding [122, 156]. Consequently, histone acetylation of the different lysine residues at regulatory regions is usually associated with active transcription [156, 157]. Histone acetylation is usually recognized by readers containing a bromodomain and erased by Histone deacetylases (HDACs), which are often part of a bigger repression complex [141, 158, 159]. Lysine 27 of histone H3 is acetylated by the HAT family CREB-binding protein (CREBBP, CBP)/ Histone acetyltransferase p300 (Ep300) and marks active promotors and active enhancers (see section 1.1.3.6) [160-162].

Histone phosphorylation of serine, threonine and tyrosine residues is mediated by kinases and removed by phosphatases [126]. For example, the histone variant H2A.X is phosphorylated rapidly in response to DNA double-strand breaks at serine 139 by Phosphoinositide 3-kinase-related protein kinases (PIKK), a kinase family which includes

the Serine-protein kinase ATM (ATM, Ataxia telangiectasia mutated), the Serine/threonine-protein kinase ATR (ATR, Ataxia telangiectasia and Rad3-related protein) and the DNA protein kinase catalytic subunit (DNA-PKcs) [105, 163, 164]. The phosphorylation of H2A.X, referred to as γH2A.X, is an early event after DNA damage and is responsible for the recruitment of DNA damage response factors to the breakage [163, 165].

These examples illustrate the spectrum of histone modifications, which increase together with the histone variants the variety of nucleosome composition immensely. The chromatin modifications are possibly important for the storage and integration of different signals induced by diverse stimuli [166]. Their combinatorial and cumulative action (histone crosstalk) regulates the transcriptional output based on all inputs and is an important regulatory mechanism during the cellular processes (Figure 8) [166, 167].

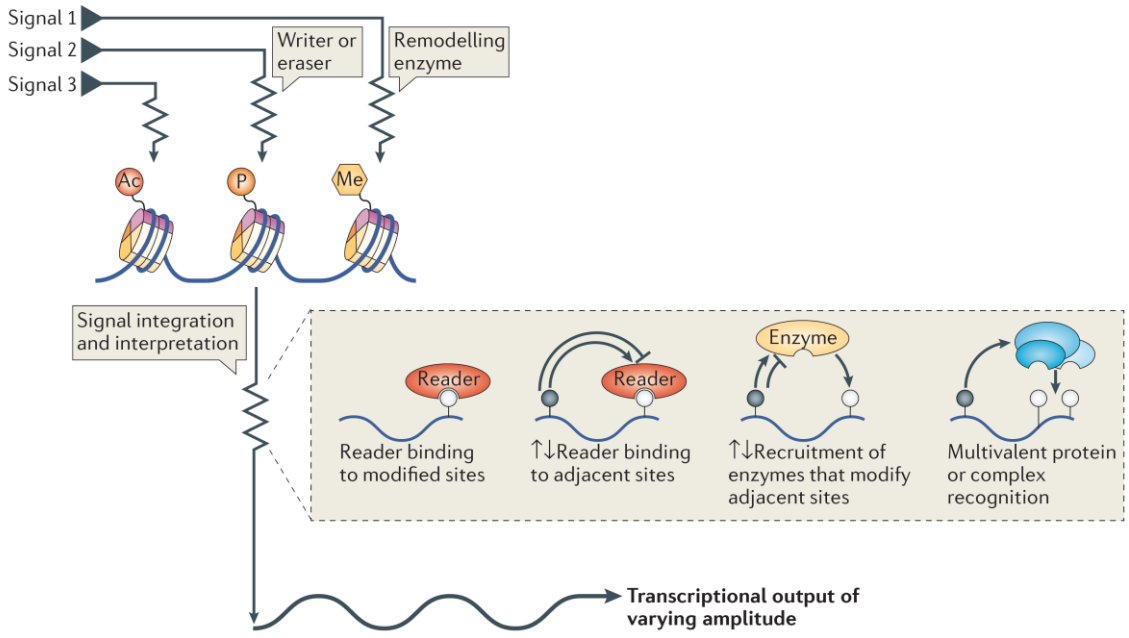



***Figure 8: Chromatin modifications serve as a signal integrator and storage***

Different signals lead via "writer", "eraser" and remodeling enzymes to alterations in the chromatin such as post-translational histone modifications. "Readers" recognize these modifications and recruit additional proteins that mediate further changes. These influence the transcriptional output based on the integrated signals. Source: [166]

### 1.1.3.4 Chromatin remodeling

DNA is highly packed into chromatin in order to fit inside the nucleus. This compaction influences all DNA-dependent processes as it can reduce the access of DNA for regulatory proteins. The exact positioning of nucleosomes can hereby have an immense impact, as many regulatory factors can only bind if the DNA is accessible, i.e. depleted of nucleosomes [168, 169]. For instance, free linker DNA can be bound by a DNA-binding factor more easily than DNA, which is wrapped around a histone octamer [170-172]. Exceptions are pioneer transcription factors that can bind to DNA on nucleosome [30, 31].

Nucleosome rearrangements can actively be catalyzed by adenosine triphosphate (ATP)-dependent nucleosome remodeling complexes. They utilize the energy released from the hydrolysis of ATP to slide nucleosomes along the DNA and to evict, insert or exchange nucleosomes or parts of it, respectively (Figure 9) [173-176].

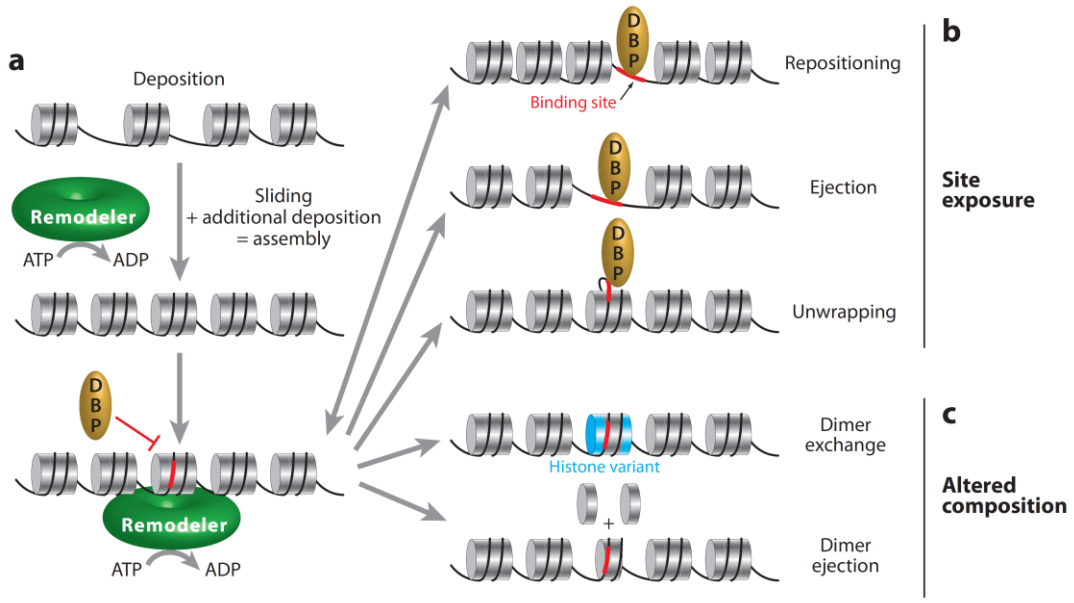



***Figure 9: Possible actions of ATP-dependent chromatin remodeling complexes***

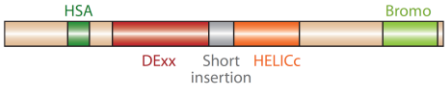
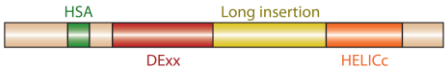
ATP-dependent chromatin remodeling complexes (green) can modify chromatin by insertion, eviction or moving nucleosomes or nucleosome components. This can result in the assembly of chromatin (A), alterations of DNA accessibility to DNA-binding factors (e. g., the DNA-binding protein (DBP)) (B) or changed chromatin composition (C). Source: [176]

Up to date, many different chromatin remodeling complexes have been discovered that are classified into four major classes: SWI/SNF (switch/sucrose non-fermentable), INO80 (inositol requiring 80), ISWI (imitation switch) and CHD (chromodomain, helicase, DNA binding) [174, 176, 177]. They are involved in a wide range of biological processes,

whereby the chromatin remodelers of the different classes might have partly specialized functions and mechanisms (Table 1) [178].

***Table 1: Families of the ATP-dependent chromatin remodelers***

The illustrations show the ATPase domain structure of the respective family: ATPase domain = DExx + HELICc, HSA = helicase-SANT, Bromo = bromodomain, tandem chromo = tandem chromodomain, SANT + SLIDE = nucleosome recognition module
Source: [176, 177, 179-188]

| Family | ATPase | Biological function |
|---|---|---|
| **SWI/SNF**; BAF; PBAF; SMARC | BRM, BRG1<br>  | - Transcription regulation<br>- Development/ differentiation<br>- Chromatin organization<br>- DNA damage response/repair |
| **CHD**; NuRD | CHD1-9<br>  | - Transcription regulation<br>- Development<br>- DNA damage response/ repair<br>- Tumor suppressor |
| **ISWI**; NURF; CHRAC; ACF | SNF2H, SNF2L<br>  | - Transcription regulation<br>- Chromatin assembly<br>- Heterochromatin formation<br>- Replication<br>- ES cell pluripotency<br>- DNA damage response/ repair |
| **INO80**; SRCAP; TRRAP/ Tip60 | INO80, Domino, SRCAP<br>  | - Transcription regulation<br>- Replication<br>- DNA damage response/ repair<br>- Insertion of histone variants<br>- Cell cycle checkpoint recovery<br>- Regulation of telomere length |

Chromatin remodelers are important for the integrity and composition of the chromatin by regulating nucleosome assembly and spacing as well as the exchange of histone variants [189]. Their action at regulatory elements further influences gene expression [189]. Chromatin remodeler can be recruited to their target genes by transcription factors, RNA polymerases, transcription elongation factors or chromatin modifications and can influence the transcription initiation, elongation and termination positively or negatively [182, 189]. The catalyzed reaction depends on the type of chromatin remodeler, the complex composition, post-translational modifications and the chromatin composition [190, 191]. Thus, chromatin remodeling displays another important regulatory layer in the control of transcription.

#### 1.1.3.5 Non-coding RNAs

Against the classical dogma, where genes are encoding proteins and RNA is only an intermediate to produce functional proteins, there is increasing evidence that RNAs also fulfill additional critical cellular functions. While only around 2 % of the genome is encoding proteins, most of the genome is transcribed [192-196]. The transcripts, which are not translated into proteins, are named non-coding RNAs (ncRNAs). The ncRNAs can occur as separate genes or overlap in sense or antisense direction with protein-coding genes [197-199]. The relative amount of ncRNAs is increasing the higher an organism is developed [200]. This suggests that ncRNAs may contribute to their gain in complexity and functionality.

Classical examples for non-coding RNAs are tRNAs or ribosomal RNAs (rRNAs) that are essential for the translation of mRNAs. These ncRNAs accomplish housekeeping function and are therefore expressed at higher levels. However, there have been many more ncRNAs identified with often low and tissue-specific expression, which fulfill diverse and essential functions in various cellular processes [201-205]. Their importance is reflected in the fact that mutations or a deregulation of these ncRNAs can lead to various types of diseases such as neurological diseases, cardiovascular diseases and cancer [206-209].

Many different kinds of ncRNAs exist and a broad classification is based on the length of ncRNAs: small/short non-coding RNAs are shorter than 200 nucleotides, while long non-coding RNAs (lncRNAs) are longer than 200 nucleotides [210, 211].

Several families of small non-coding RNAs exist such as tRNA, micro RNA (miRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA) and Piwi-interacting RNA (piRNA) [204, 212]. They control many biological processes often by regulating gene expression post-transcriptionally [204]. For instance, miRNAs recognize their target mRNAs by base pairing and can repress the mRNA expression by different mechanisms such as mRNA degradation or translation inhibition [213, 214].

The lncRNAs are involved in diverse biological processes like development, X-chromosome inactivation, genomic imprinting or transcription regulation [199, 215-218]. In order to achieve the latter, lncRNAs mediate changes in chromatin or in the activity of transcription factors [199, 217]. For example, many long non-coding RNAs, which are transcribed from intergenic regions (long intergenic non-coding RNAs,

lincRNAs), interact with chromatin-modifying complexes and ultimately influence gene expression [219]. These lncRNAs can apply different mechanisms: They can tether the chromatin-modifying complex *in cis* (e.g., ANRIL [220]) or *in trans* (e.g., HOTAIR [221, 222]) to specific genomic regions, they can mediate intrachromosomal looping (e.g., HOTTIP RNA [223]) or, for example, mediate the binding of chromatin modifiers by forming a triple helix with specific DNA sequences (e.g., rDNA-pRNA triplexes [224]) (Figure 10). Moreover, lncRNA could allosterically regulate the activity of the protein they bind to by inducing a conformation change (e.g., G1/S-specific cyclin-D1 (*CCND1*) promoter RNA [225]) or act as a scaffold to link proteins together (e.g., HOTAIR [226]). Furthermore, lncRNA can regulate transcription factors or the transcription machinery directly, for example, by preventing their binding to DNA or influencing their post-transcriptional modifications [217, 227, 228]. In general, the RNA secondary structure and/or the RNA sequence specificity to pair with DNA may mediate these processes [229].

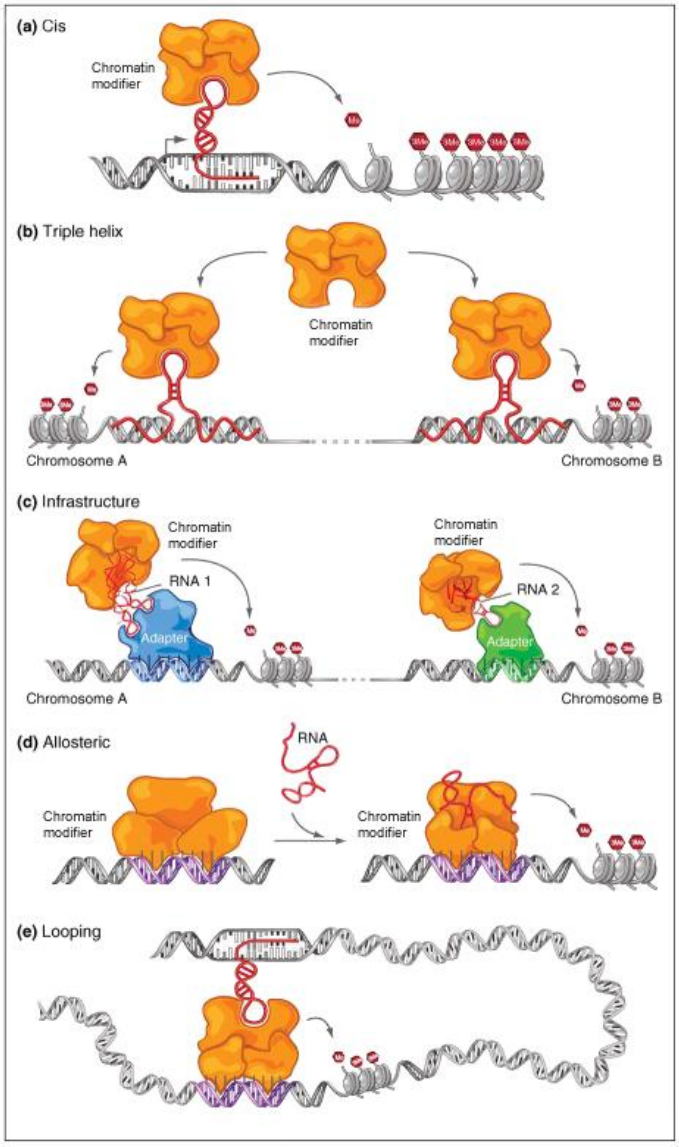



***Figure 10: LncRNAs apply diverse mechanisms to regulate gene expression***

(a-c) lncRNA can act by eviction or recruitment of proteins, such as chromatin modifiers, to DNA. (d) They can also act as scaffolds to bring different proteins together or by allosterically modifying their activity or DNA binding capacity. (e) Furthermore, they can influence chromatin looping, for example, between promoters and enhancers (see section 1.1.3.6). Source: adapted from [230]

Many more types of ncRNAs have been proposed and also post-transcriptional modifications have been described for RNAs, which increase the variety of these regulatory molecules immensely [231, 232]. Their relevance and functional mechanisms are currently under investigation.

### 1.1.3.6 Distal regulatory regions

Promotors have long been known to be essential for the regulation of gene expression. However, increasing evidence has been obtained that also distal regulatory regions can silence or enhance gene expression (Figure 11) [234]. Enhancers have originally been defined as genomic elements, which act over a long distance (in average over ~20kb) to positively regulate the transcription of their target genes [235, 236]. The spatiotemporal activities of enhancers have been shown to be essential for the development, including the linage commitment and the cell type specific transcription control, and the response to stimuli [237-246].

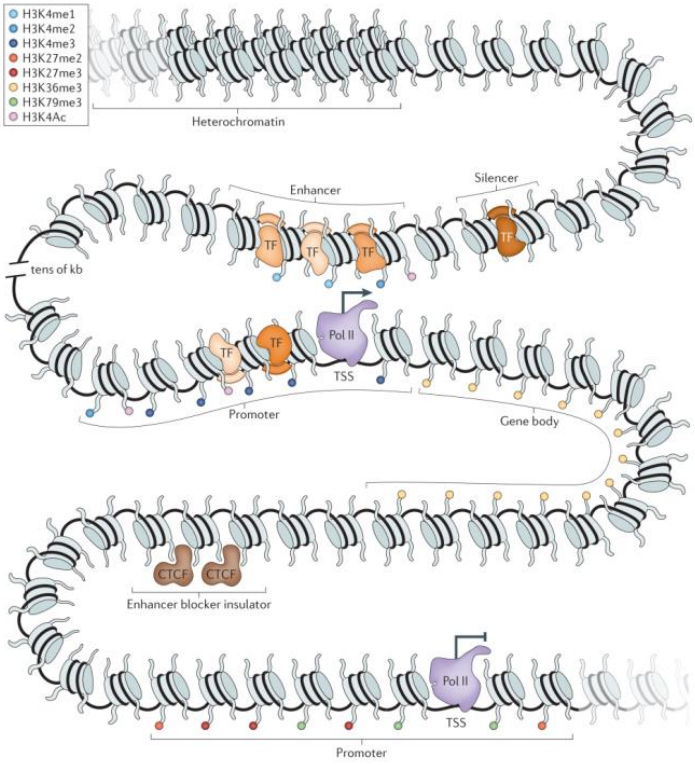



***Figure 11: Gene regulatory regions***

Besides promoters, distal regulatory regions, such as enhancers, silencers or insulators, impact on the transcriptional outcome of a gene. Source: [233]

Enhancers contain binding sites for sequence-specific transcriptions factors, which in turn can recruit co-activators and co-repressors to define the activity of the enhancer [247]. Thereby the composition as well as the sequential binding of the factors can be crucial [50, 248-250]. How this is controlled in a spatial and temporally organized manner during development is not yet fully understood.

Closed or poised enhancers are marked with monomethylated H3K4 (H3K4me1) and/or H3K27me3, while active enhancers are characterized to be accessible as well as marked by H3K27ac and H3K4me1 (Figure 12) [160, 244, 251]. Enhancers can also be transcribed by RNA polymerase II resulting in the expression of enhancer RNAs (eRNAs) [252]. Some studies suggest a functional role for the eRNAs [250, 253, 254]. For instance, there are hints that eRNAs might be involved in the formation or stabilization of chromatin loops between enhancers and their target genes or mediate chromatin remodeling [255-259]. A recent publication suggests that eRNAs may build a positive-

feedback loop by stabilizing the transcription factor occupancy at the gene regulatory elements [260].

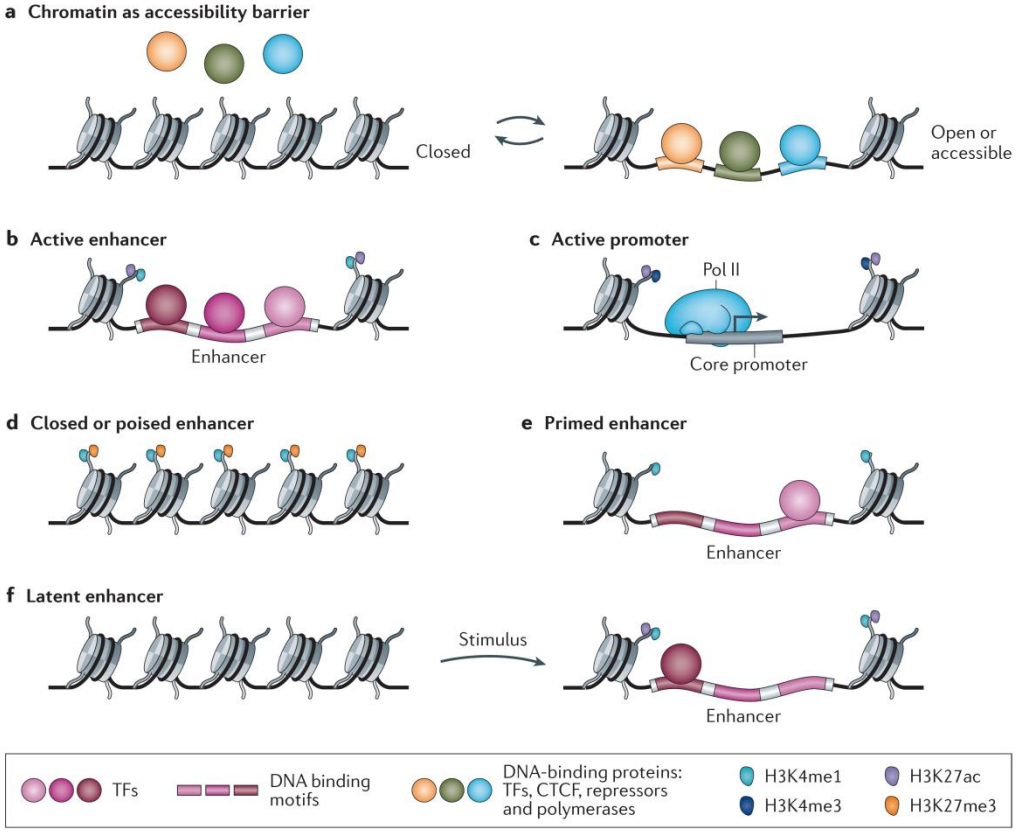



***Figure 12: Properties of different enhancer states***

Active enhancers are accessible for transcription factors (TFs) and RNA polymerase II and marked by H3K4me1 and H3K27ac, while closed or poised enhancers are marked by H3K4me1 and H3K27me3. Source: [247]

Mechanistically, experimental data suggest that enhancers are brought in close proximity to the genes they activate by chromatin looping (Figure 13) [261-263]. Latter seems to be mediated by transcription factors, the Transcriptional repressor CTCF (CTCF, CCCTC-binding factor), Cohesin, Mediator and the transcription machinery [258, 264-266]. The long-range enhancer-promoter interactions enable the enhancers to influence transcription initiation or elongation [258, 267]. Promotors as well as enhancers can be engaged in several long-range interactions suggesting a potential combinatorial or cumulative regulatory mechanism [268-270]. Moreover, the promoter-enhancer pairs may vary from cell-to-cell as there is a variability in the chromosomal structure between cells [271].

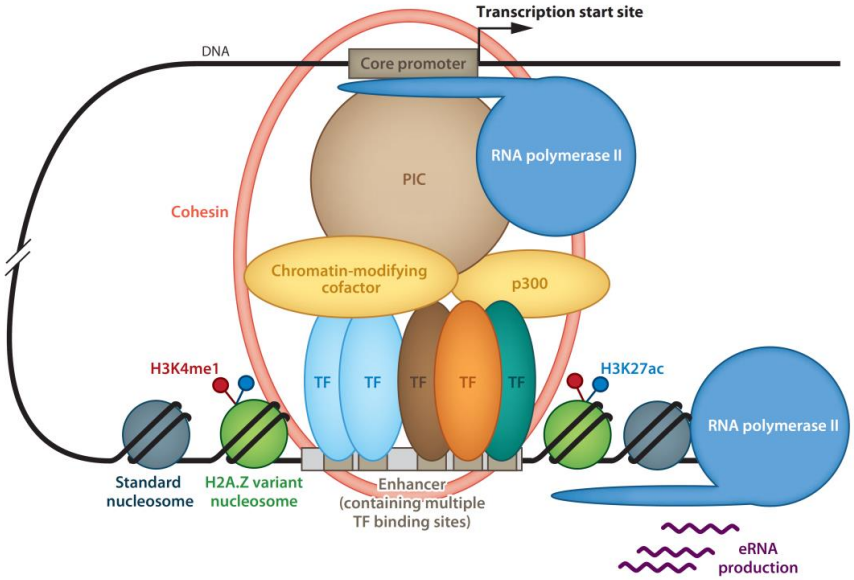



***Figure 13: Schematic model of promoter-enhancer looping***

Transcription factors and chromatin modifiers bind to enhancers and complex with the preinitiation complex (PIC) and RNA polymerase II that bind to the promoter of a gene. To allow this complexation, chromatin undergoes looping so that a promoter and an enhancer are in close proximity. This structure can be stabilized by Cohesin. Enhancers can also be transcribed resulting in the production of enhancer RNAs (eRNAs). Source: [272]

Special types of enhancers have been reported. For example, locus control regions (LCRs) are defined "by their ability to enhance the expression of linked genes to physiological levels in a tissue-specific and copy number-dependent manner at ectopic chromatin sites" [273]. They can strongly enhance the transcription of its targets as shown for the β-globin LCR [274]. Super-enhancers are large clusters of enhancers occupied by key transcription factors and Mediator [275, 276]. They are controlling the expression of cell identity genes and are enriched at oncogenes in cancer cells [275, 277, 278]. Furthermore, super-enhancers also respond to stimuli and are therefore important for the cellular reactions to environmental changes [279-281].

### 1.1.3.7 Chromatin organization

Evidences are increasing that also the dynamic spatial organization and the compartmentalization of the genome play an important role in the regulation of gene expression [13, 25, 282].

Interphase chromosomes occupy distinct regions in the nucleus that are called chromosome territories (Figure 14) [283-285]. The chromosomes are not freely arranged in the genome, as major nuclear structural elements constrain some interactions and different domains within the chromosome territories are likely occupied by active or inactive chromatin, respectively [13]. For instance, centromeres,

telomeres and genes encoding ribosomal RNA are located at particular regions in the nucleus [286-288]. In addition, large, evolutionarily conserved transcriptionally inactive chromatin regions have been described to interact with the nuclear lamina (lamina associated domains, LADs) or with the nucleolus (nucleolus associated domains, NADs) [289-292]. These regions can be dynamically rearranged according to the expression change of the genes, for example during the differentiation of a cell [293-295]. Furthermore, transcribed genes were observed to cluster together in so-called transcription factories [296-298]. A recently developed model for the nuclear landscape proposes that the active and inactive nuclear compartments are spatially co-aligned [299].

Not only more intra-chromosomal than inter-chromosomal interactions occur, but also certain regions within the chromosome show higher contract frequencies compared to inter-domain interactions [300, 301]. For instance, the promoter-enhancer looping mostly occurs within these topologically associating domains (TAD) [258, 302-304]. The borders of the TADs are enriched for CTCF, Cohesin and the Mediator complex as well as for highly transcribed housekeeping genes [302, 305].

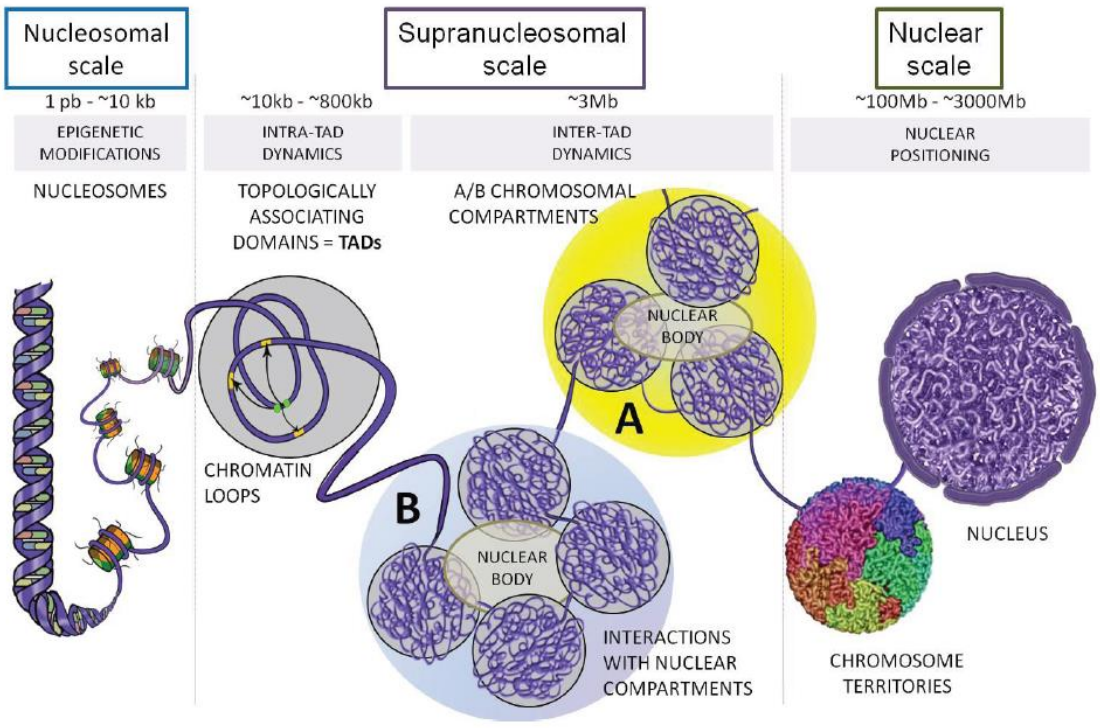



***Figure 14: Three-dimensional organization of the mammalian genome***

The DNA is wrapped around nucleosomes (nucleosomal scale). Due to chromatin interactions between different regions, for example by chromatin looping, the chromatin is further folded. These interactions mainly occur within topologically associating domains (TADs). Furthermore, transcriptionally active regions (A) can cluster together and transcriptionally inactive regions have been found to interact with nuclear compartments such as the nuclear laminar or the nucleolus (B). Chromosomes occupy distinct regions in the nucleus called chromosome territories (nuclear scale). Source: [305]

## 1.2 The effect of ultraviolet light on cells

### 1.2.1 UV light induces DNA damage

Cellular DNA is continuously confronted by endogenous or exogenous stimuli, which could directly or indirectly induce DNA damage. Endogenous DNA damage agents are, for example, reactive oxygen species (ROS) or other reactive molecules produced by biochemical reactions like metabolic processes inside the cell [307]. In addition, environmental cues can damage DNA such as ionizing or ultraviolet (UV) radiation. These damages range from base modifications to DNA double-strand breaks [308]. UV-induced DNA damages are primarily cyclobutane pyrimidine dimers (CPDs), pyrimidine 6-4 pyrimidone photoproducts (6-4PPs), and their Dewar isomers, but also DNA double-strand breaks (Figure 15) [306]. UV light can also damage the DNA indirectly by the generation of reactive oxygen species [309]. The solar UV light that is not absorbed by the ozone layer can induce under strong sunlight conditions around 100,000 lesions per hour in an exposed cell [310].

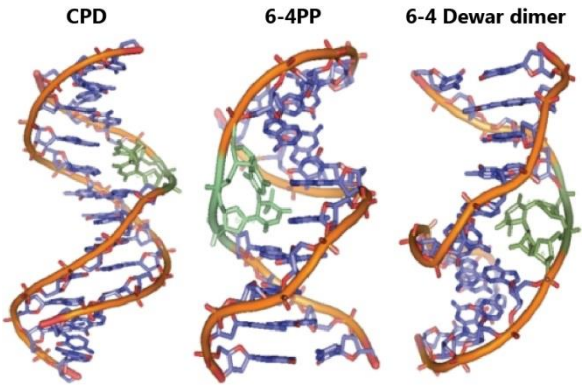



***Figure 15: Structure of DNA duplexes with common UV-induced DNA lesions***

Structures of CPD, 6-4PP and 6-4 Dewar dimer lesions are shown in green. Hydrogen atoms are not shown. Used PDBs: 1TTD, 1CFL, 1QKG. Source: [306]

DNA damages pose an enormous risk for the cell as they can have profound consequences on the cellular function. The damaged DNA may lead, for example, to aberrant or even non-functional gene products, which would consequently affect the cellular processes they are involved in. Thus, preserving the genome integrity is essential for the function and viability of cells. In order to reverse the damages, several DNA repair pathways have evolved, each of them specialized for specific DNA lesions. The UV-induced mutagenic photodimers or other helix-distorting base lesions are repaired by the nucleotide excision repair (NER) (Figure 16) [311, 312]. The NER system is variable, but contains common steps including the recognition of the DNA damage, unwinding of the DNA, cleavage and excision of the damaged sites and finally the new synthesis of the removed parts with ligation of the cut sites [313, 314]. The global-genome NER (GG-NER) is able to repair DNA lesions throughout the genome independent of its structural and functional status and prevents mutagenesis [311, 315].

Transcription-coupled NER (TC-NER, TCR) is instead specialized for an efficient repair of DNA lesions that occur in the template DNA strand of actively transcribed genes [314, 316-320]. The DNA lesion leads to a stalling of the RNA polymerase II during the transcription elongation, which triggers the repair of the damage and then allows a fast resumption of the transcription after the successful repair [315]. The stalled RNA polymerase can not only activates TC-NER, but also DNA damage signaling or, if the blockage persists, the Cellular tumor antigen p53 (TP53) dependent apoptosis [318, 321, 322]. DNA double-strand breaks are mainly repaired by non-homologous end-joining (NHEJ), homologous recombination (HR) or alternatively by more error-prone repair pathways [323, 324].

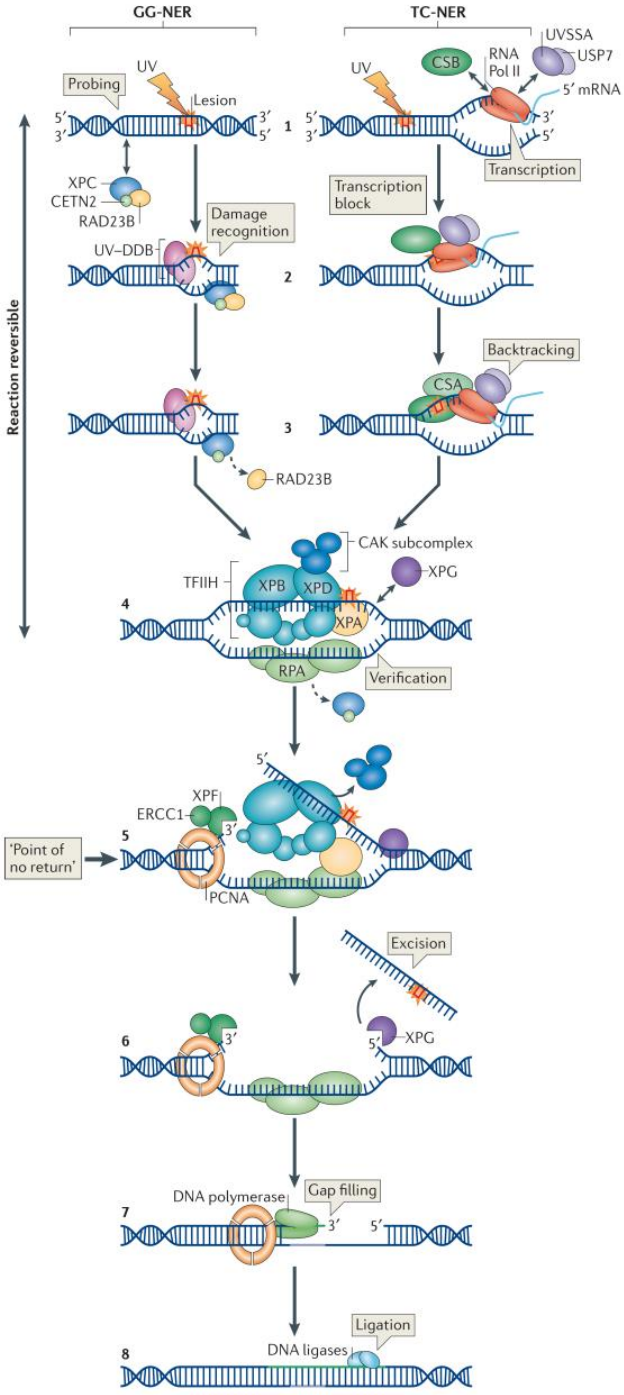



***Figure 16: Nucleotide excision repair***

A schematic illustration of the global-genome NER (GG-NER, left) and the transcription-coupled NER (TC-NER, right). A complex containing the damage sensor XPC (DNA repair protein complementing XP-C cells homolog) and the UV-DDB (ultraviolet (UV) radiation- DNA damage-binding protein) complex are recognizing the damages induced by UV irradiation and initiate the GG-NER (1-3 left). TC-NER is started by the indirect recognition of the DNA damage by a stalled RNA polymerase II (RNA Pol II) at a lesion during transcription elongation (1-3 right). Following the damage recognition, the transcription initiation factor IIH (TFIIH) complex is recruited to the DNA lesion in both pathways and its helicase activity helps unwinding the DNA around the lesion (4). The Replication protein A (RPA) is recruited to coat the undamaged strand. Recruited endonucleases induce incisions 5′ to the lesion (5) and 3′ to the lesion (6), which leads to the excision of a 22-30 nucleotide-long DNA strand including the DNA lesion (6). The 5′ endonucleases recruit the Proliferating cell nuclear antigen (PCNA) that in turn recruits DNA polymerases. The DNA polymerase fills the gap by synthesizing DNA using the non-damaged strand as a template (7). The nicks are finally sealed by DNA ligases (8). Source: [311]

If the cells fail to repair the damages, mutations in the cells will arise and increase the risk for malignant transformation towards carcinogenesis [325-327]. For example, the risk for skin cancer is 2,000- to 10,000-fold higher in xeroderma pigmentosum (XP) patients with defective NER compared to the general population and the cancer also occurs at a significantly younger age in these patients [328].

The DNA damage response (DDR) comprises the recognition of the damage that lead via mediators and signaling cascades to the activation of effector molecules, which induce cellular responses such as DNA repair, chromatin remodeling and control of transcription, cell cycle progression and cell fate (Figure 17) [310, 329]. For example, the cells might arrest in their cell cycle to protect the propagation of damaged or incorrect genetic information and induce expression and chromatin changes [329]. UV irradiation can also induce programmed cell death, which is important to prevent mutagenesis and the establishment of disease [330].

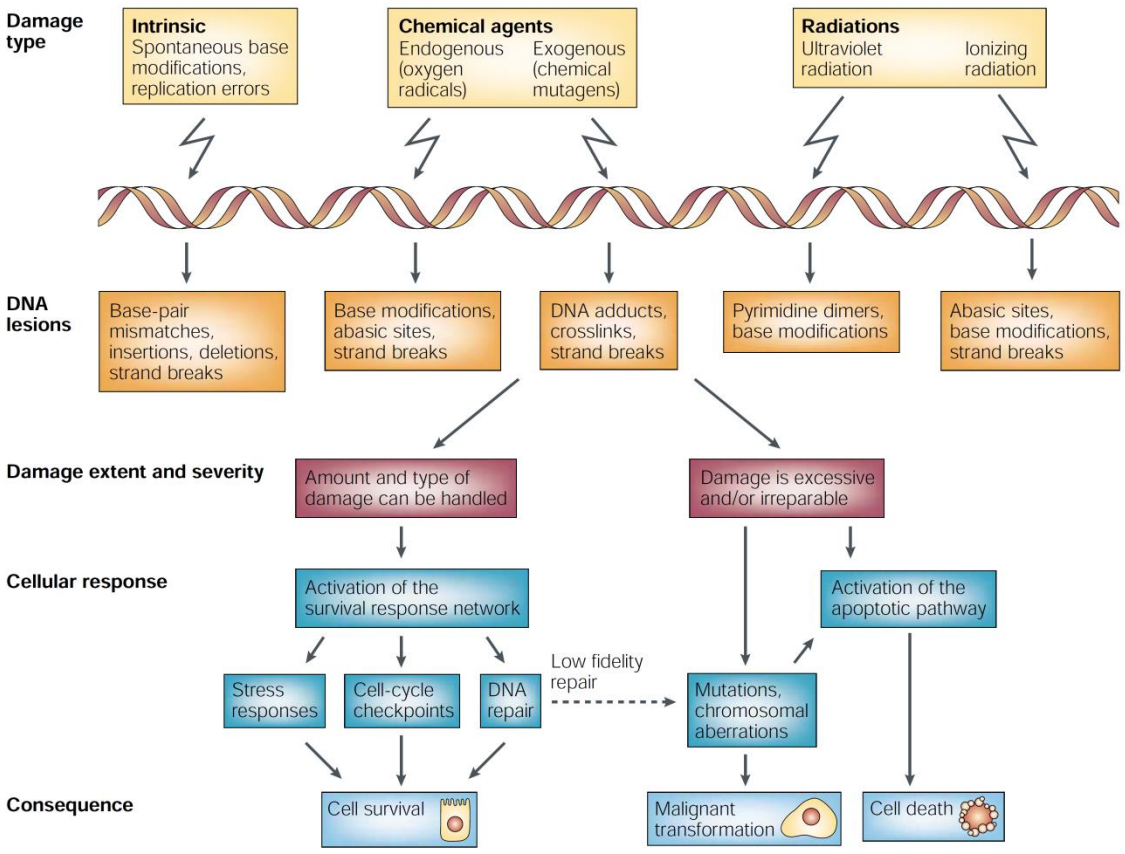



***Figure 17: A scheme of the DNA damage response***

Different intrinsic and extrinsic cues lead to various types of DNA lesions. The process of DNA damage response comprises transcription modulation, cell cycle regulation and DNA damage repair. In case the damage is too severe, cells will undergo cell death or the mutations may lead to malignant transformations. Source: [331]

### 1.2.2 Influence of UV exposure on gene expression

Gene expression is widely affected in cells, which have been exposed to UV light. For example, bulky DNA lesions are able to hinder RNA polymerase elongation leading to transcription arrest and induction of TC-NER [315, 332, 333]. However, the RNA polymerase can sometimes also bypass the lesions causing transcriptional mutagenesis (translesion bypass synthesis) [334-336]. During nucleotide excision repair, the complex of the general transcription initiation factor IIH (TFIIH) is recruited to the DNA damage site [337]. Furthermore, the basal transcription factor TFIID/TATA-box-binding protein (TBP) can bind to DNA lesions caused by cisplatin treatment or UV irradiation due to their structural similarity with the TATA boxes [338, 339]. The recruitment of basal transcription factors to damaged DNA may contribute, at least locally, in transcription inhibition of non-damaged genes by reducting the pool of available transcription factors [337, 338].

Besides the DNA damage mediated transcription changes, many genes are de-regulated as part of the UV-induced DNA damage response. For example, the so-called primary response genes are rapidly induced in their expression after cellular stimulation by pre-existing transcription factors [340, 341]. Among the UV-induced immediate early genes are the transcription factors Proto-oncogene c-Fos (*c-Fos, Fos*) and Transcription factor AP-1 (*c-Jun, Jun*), which dimerize to the activator protein 1 (AP1) complex that activate the transcription of other genes [342, 343]. The importance of the AP1 complex in cellular protection and survival rather than DNA repair in response to UVC irradiation has been demonstrated in mouse 3T3 fibroblasts lacking *c-Fos* [344]. After UV irradiation, these cells strongly reduced transcriptional induction of the tested AP1 target genes and, while still being able to repair UV-damaged DNA, exhibited increased cell death and delayed re-entering into the cell cycle of the UV-arrested cells [344].

The transcriptional or post-transcriptional activation of UV-induced transcription factors can be mediated via signaling cascades initiated by the UV light [345-347]. For example, UV irradiation induces different Mitogen-activated protein kinases (MAPKs) and Nuclear factor NF-kappa-B (NFκB) signaling pathways (Figure 18) [348-351]. These MAPK signaling pathways, especially the c-Jun N-terminal kinase (JNK1-3, MAPK8-10) activating pathway, can result into the transcription activation of AP1 components and furthermore increase the activity of AP1 by post-translational phosphorylation of the newly synthesized as well as the pre-existing AP1 components [342, 345, 347, 351-361].

The MAPKs are important for the regulation of cell growth, chromatin remodeling and cell death [309].

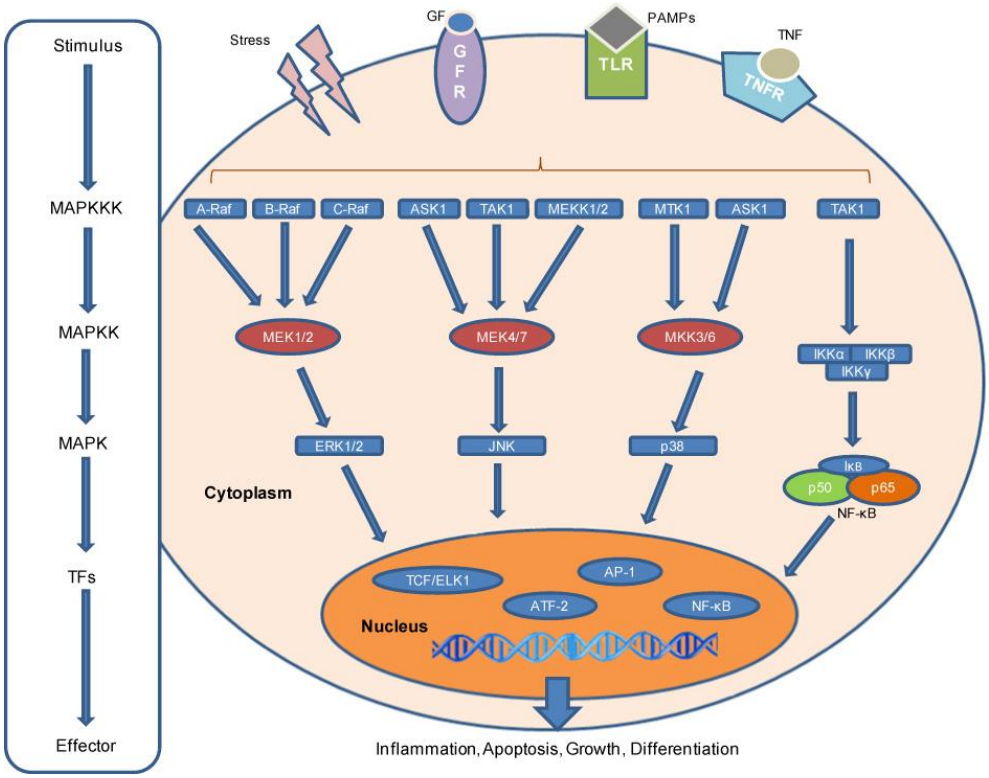



***Figure 18: UV light induces MAPK and NFκB signaling cascades***

UV light induces signaling cascades in the cells. The signal is transferred by kinases that sequentially phosphorylate each other and finally result in the activation of certain transcription factors. These regulate the expression of genes important for the stress response.
Source: [362]

Different kinases, such as JNK1 (MAPK8), p38 (MAPK14), ERK1/2 (MAPK3/Mapk1), and ATR, can phosphorylate the Cellular tumor antigen p53 in response to UV radiation, which increases the activity of the transcription factor [363-366]. TP53 protein levels are also increased in response to UV exposure [367, 368]. TP53 is a tumor-suppressor that can mediate in response to cellular stress G1 and G2 cell cycle arrest [369-373]. The G1 cell cycle arrest allows DNA repair before the DNA replicates. However, UV-induced G1 arrest can occur even independent of TP53 [374]. TP53 can further induce apoptosis to eliminate damaged cells that might become carcinogenic [375-377]. Whether TP53 mediates cell cycle arrest or cell death depends on many factors such as the cellular properties, the intensity of the stress signal and the molecular regulation of the TP53 target choice [309, 371, 378]. Furthermore, TP53 might contribute to the NER [309]. UV irradiation can also induce mutations in the *Tp53* gene, which can lead to skin cancer;

this further underlines the protective function of TP53 in response to UV irradiation [309, 379-381].

TP53 can also regulate the expression of large intergenic non-coding RNAs [382]. One of them, lincRNA-p21, has been shown to associate with the Heterogeneous nuclear ribonucleo-protein K (hnRNP-K) and to localize the repressive complex to its targets (Figure 19) [382]. The function of the lincRNA-p21 is required for TP53-dependent apoptosis [382]. Also other non-coding RNAs have been reported to be modulated in their expression in response to DNA damage and to fulfill important functions during the DDR [383-389]. The lncRNAs can mediate their response by interaction with transcription regulators or chromatin modifiers [382-385]. Recently, a significant deregulation of the expression of many long non-coding RNAs in UVB-irradiated melanocytes has also been reported [390].

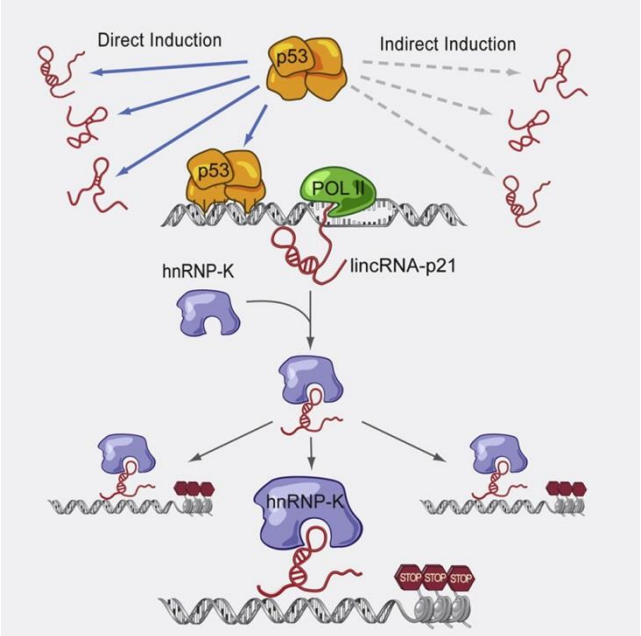



***Figure 19: TP53-induced lincRNA-p21 mediates gene repression required for apoptosis***

TP53 induces the expression of the large intergenic non-coding RNA lincRNA-p21, which interacts with hnRNP-K. The lincRNA-p21 helps recruiting the repressive complex to its targets. This process is involved in TP53-induced apoptosis during the DDR. Source: [382]

### 1.2.3 Chromatin changes following UV radiation

UV exposure and the induced DNA damage lead to a wide range of chromatin modulations. For example, nucleosome rearrangement following UV treatment have been reported which resulted in the access/prime-repair-restore model (Figure 20) [391-393]. This implies that the damaged DNA becomes accessible and thereby primed for the binding of repair factors and after successful repair, the chromatin is restored. ATP-dependent chromatin remodeling factors are important for mediating these chromatin changes as it has been shown that these alterations require ATP hydrolysis and members of all four families of ATP-dependent chromatin remodelers can be recruited to the damaged sites [394, 395]. For instance, it has been demonstrated that members of the INO80 or SWI/SNF families catalyze the chromatin relaxation required

for the initial steps of the nucleotide excision repair (Figure 20) [396-398]. The chromatin remodelers contribute to efficient DNA repair and can thereby reduce the risk for apoptosis [399].

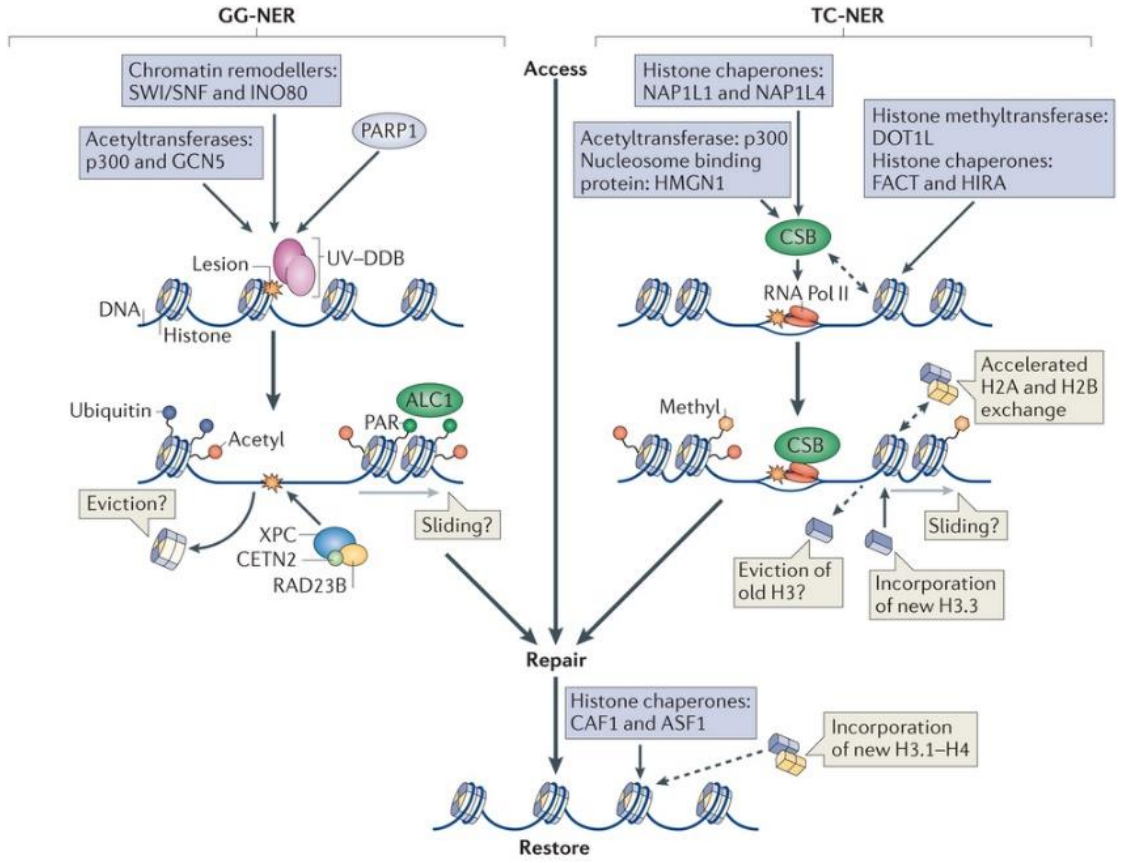



***Figure 20: Epigenetic mechanisms involved during the repair of UV-induced DNA damage***

To allow DNA repair, e.g. by GG-NER (left) or TC-NER (right), the chromatin is opened with the help of histone modifiers, histone chaperones and chromatin remodelers. This allows DNA repair factors to bind and to efficiently repair the damage. Afterwards, the chromatin is restored ("access, repair, restore" model). Source: [311]

The chromatin changes during the DDR include the exchange of histone variants (Figure 21) [392, 400]. For instance, the histone chaperone histone regulator A (HIRA) deposits H3.3 at the UV-damaged sites, which seems to prime the chromatin for transcription restart after the DNA repair and to be essential for the replication fork progression after UV damage [401, 402]. The histone chaperone Chromatin assembly factor 1 (CAF-1, CNOT7, CCR4-NOT transcription complex subunit 7), which is recruited in a NER-dependent manner to the damaged sites, incorporates the histone variant H3.1 after the repair of UV-induced DNA damage [403, 404]. Also the histone variants H2A.X and H2A.Z show enhanced dynamics following UV microirradiation [405, 406]. The histone chaperone FACT (facilitates chromatin transcription) mediates H2A and H2B exchange at UV-induced DNA damage sites, which is important for the transcription restart of the stalled RNA Polymerase II [407]. These examples show the

dynamics of histone exchange in response to UV and illustrate that the histone variants can fulfill different roles during the DNA damage response.

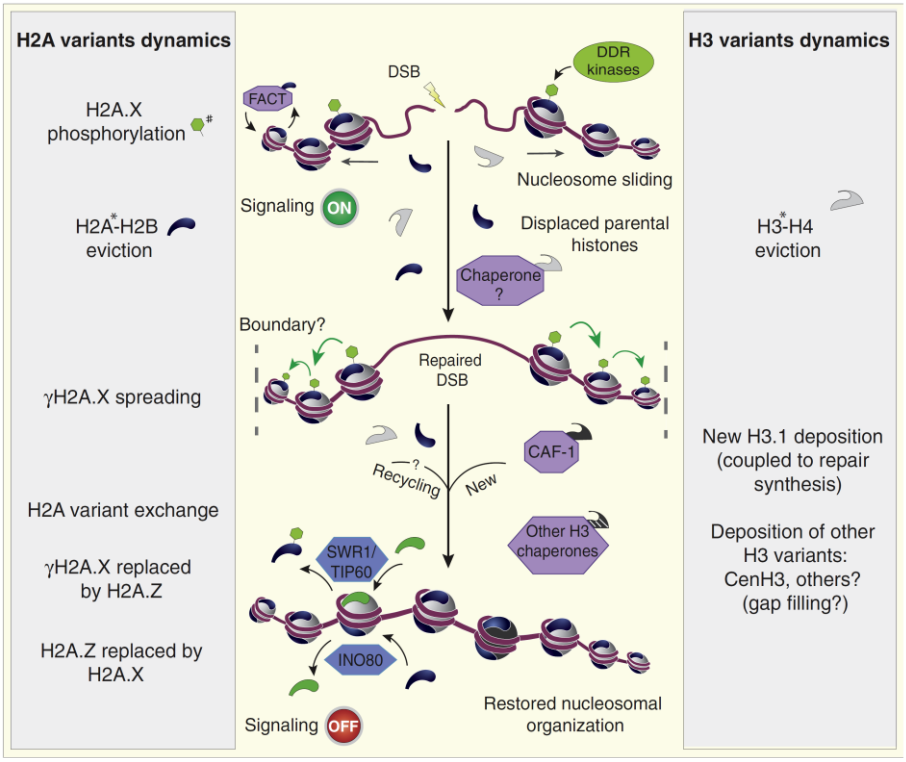



***Figure 21: Dynamics of H2A and H3 variants in response to DNA damage***

DNA damage response kinases phosphorylate H2A.X in response to DNA double-strand breaks. This initiates a bidirectional spreading of the γH2A.X and inhibits the H2A.X/H2A exchange by the histone chaperone FACT. The nucleosomal organization becomes restored after the repair with the help of histone chaperones (purple) and chromatin remodelers (blue). For instance, new H3.1 is deposited by the Chromatin assembly factor 1 (CAF-1). Source: [392]

Also post-translational histone modifications are crucial players in the spatiotemporal regulation of the DDR [408]. The most well-known example for a DNA damage induced histone modification is the phosphorylation of H2A.X by kinases of the phosphoinositide 3-kinase-like kinase family [409]. The establishment of γH2A.X around DNA double-strand breaks is an early and important step in the DNA damage response and leads to checkpoint activation, cell cycle arrest, recruitment of repair proteins and DNA repair [409]. The Mediator of DNA-damage checkpoint protein 1 (MDC1) binds to γH2A.X and recruits important DNA damage response proteins to the damage sites such as the MRN (Mre11-Rad50-Nbs1) complex or the ATM kinase [410-414]. γH2A.X can spread up to around 1.7 x $10^6$ bp around the break and can be induced by UV irradiation [415-419].

Besides the phosphorylation of H2A.X, many histone modifications are modulated upon DNA damage and mediate important DNA damage responses [420-422]. For instance,

the histone marks H3K9ac and H3K56ac have been reported to be reversibly reduced in response to DNA damage [420]. H3S28 phosphorylation has been found to occur at 50% of all stress-induced genes in 3T3 fibroblast cells [422]. The phosphorylation of H3S28 contributes to the dissociation of histone deacetylases, which results in an enhanced transcription of the respective genes [422]. The histone acetyltransferase KAT2A (KAT2A, GCN5) and p300 are implicated in mediating acetylation changes in response to UV irradiation [423]. Also further post-translational modifications such as methylation, ubiquitylation, SUMOylation, neddylation, poly(ADP-ribosyl)ation are important for the DNA damage response [408, 423].

The changes in histone modifications, histone variants and chromatin remodeling are occurring highly interconnected. For instance, phosphorylation of H2A.X facilitates the FACT-mediated exchange of H2A.X [424]. In response to ionizing radiation, γH2A.X stimulates the acetylation of H3 on the γH2A.X nucleosome to which the SWI/SNF chromatin remodeler binds with its bromodomain [425]. The SWI/SNF chromatin remodeler is further required for the phosphorylation of H2A.X and the DNA repair, establishing a feedback loop between these features to promote the DNA double-strand repair (Figure 22) [425].

Distal regulatory regions might also play a role in the DDR as genome-wide analyses of TP53 occupancy in human and mouse cells revealed a predominant binding of TP53 to enhancer regions [387]. Furthermore, spatial chromatin reorganization has been observed in response to DNA damage [185, 427, 428]. The role of distal regulatory elements and of potential chromatin movements during UV-induced DDR need to be further explored in mammalian cells.

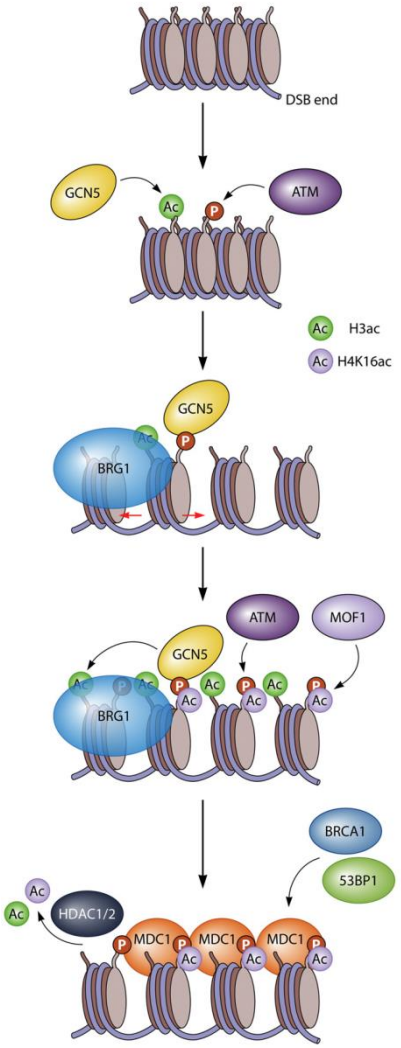



***Figure 22: The interplay of chromatin modifications to promote DNA repair***

Histone acetylation by GCN5 and the SWI/SNF chromatin remodeler (Transcription activator BRG1 (Smarca4)) promote H2A.X phosphorylation by ATM in a feedback loop. γH2A.X and MOF1-catalzyed H4K16ac recruit MDC1. MDC1 mediated recruitment of factors important for the DNA damage response and histone deacetylation by HDACs promote DNA repair. Source: [426]

## 1.3 Circadian rhythm

### 1.3.1 Adaptation of organisms to circadian rhythm

Many organisms, from prokaryotes to eukaryotes, have evolved an endogenous clock to conform their behavior and physiological processes to the 24 h day-night cycle caused by the rotation of the earth on its axis [429-432]. This not only allows these processes to occur at the appropriate time of the day, but also entails the organism to foresee and react in accordance with environmental changes [433, 434]. For example, the circadian rhythm plays a role in the regulation of the cardiovascular system, the renal system, the metabolism, the endocrine system, the immune system, the body temperature and the reproductive system and is connected with the cell cycle regulation and the DNA damage response [435-438]. These diverse functions also become visible in the variety of diseases, including sleep disorders, metabolic syndromes, cardiovascular as well as psychiatric diseases and cancer, which are linked to disruptions of this timekeeping mechanism (Figure 23) [435, 439-441].

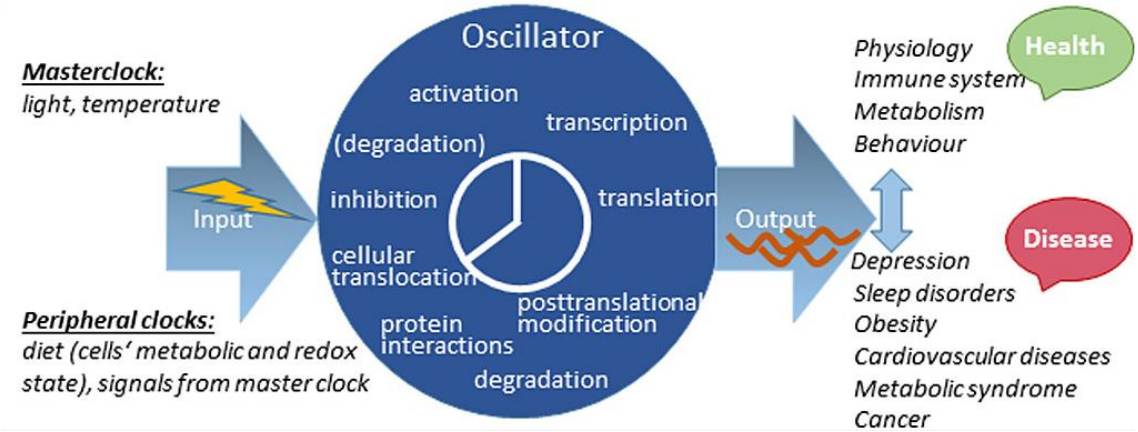



***Figure 23: Schematic view on the input and output signals of circadian oscillations***

The central oscillator (masterclock) receives its input signals (Zeitgeber) from the environment, which are mainly light, but also temperature. After integrating them, the masterclock in turn sends signals to the peripheral tissues (peripheral clocks) to synchronize their cell-autonomous oscillations. Furthermore, the oscillations of the peripheral clocks are also influenced by the metabolism. The cellular oscillations are based on transcription-translation feedback loops and additional regulatory mechanisms. A disruption of these oscillations can result in deregulation of normally circadianly occurring physiological processes and ultimately in a large variety of diseases. Source: [442]

The cellular circadian rhythms are self-sustainable, but synchronized to and entrainable by the environment. Therefore, the organisms recognize external cues that are termed "Zeitgeber". The main stimuli are photic cues that are recognized by the eyes and are transmitted mainly by the retinohypothalamic tract (RHT) to the suprachiasmatic

nucleus (SCN), which is located in the anterior hypothalamus of the central nervous system. The SCN is the master circadian peacemaker in mammals, which entrains the intrinsic circadian rhythm of the other cells/tissues in the body via neuronal or hormonal pathways (Figure 24) [443, 444]. These peripheral clocks can furthermore be influenced by food intake or temperature [444-448].

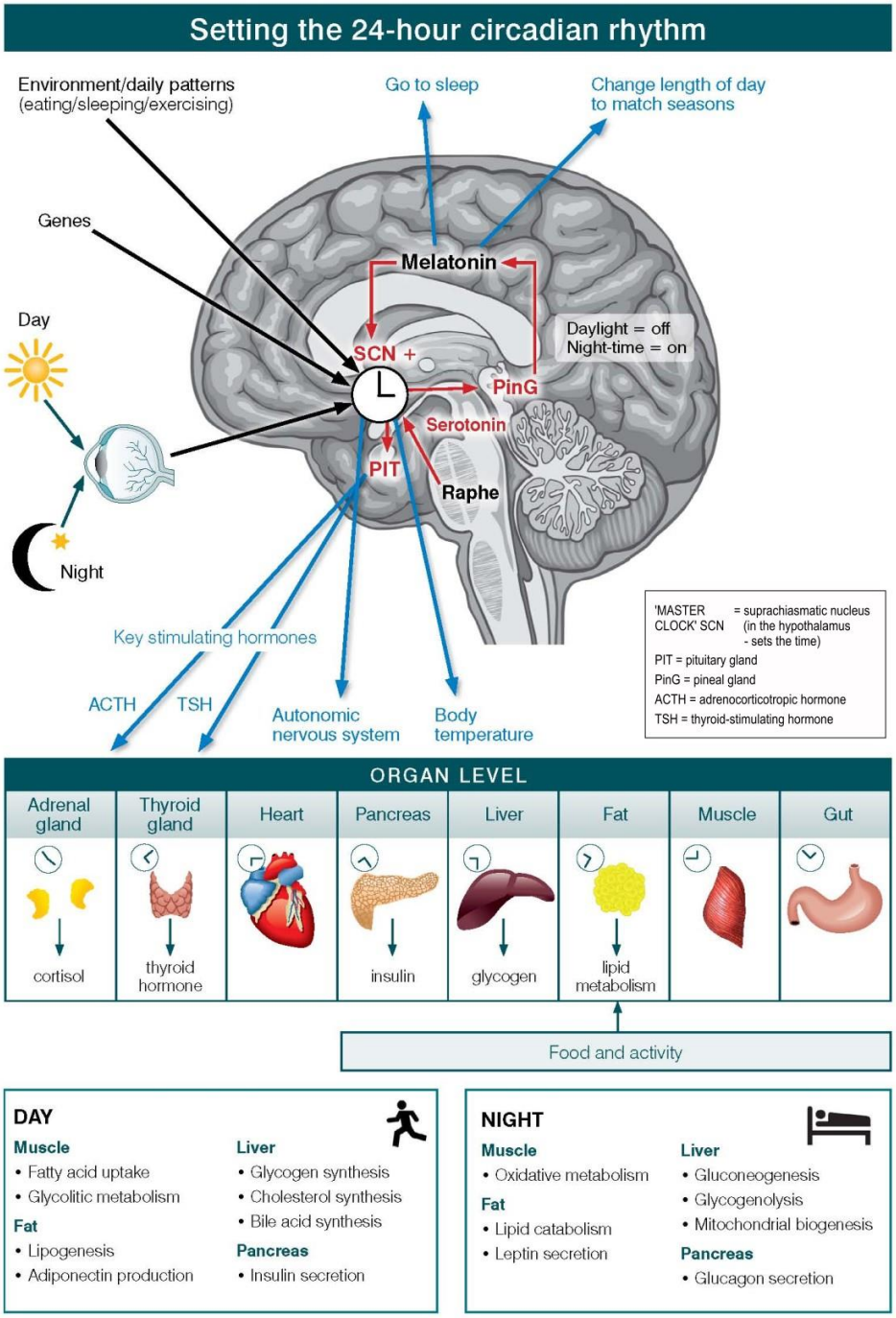



***Figure 24: The SCN entrains the peripheral clocks to the day-night cycle***

The SCN integrates external stimuli and coordinate via neural and humoral signals the circadian rhythm of the peripheral tissues. Essential biological processes of the different tissues are under circadian regulation. Source: adapted from [449]

### 1.3.2 Transcription-translation feedback loops establish the cellular circadian rhythm

The cell intrinsic circadian rhythm is regulated by the core clock factors, which act in interlocking transcriptional feedback loops with post-transcriptional and post-translational regulation (Figure 25) [450]. The helix-loop-helix transcription factors Circadian locomoter output cycles protein kaput (CLOCK) and Aryl hydrocarbon receptor nuclear translocator-like (ARNTL, BMAL1) heterodimerize, bind mainly to E-box DNA motifs and activate the transcription of a high number of genes, called the clock-controlled genes (CCGs) [451-453]. They also activate the transcription of Period circadian protein homolog 1-3 *(Per1, Per2, Per3)* and Cryptochrome *(Cry1, Cry2)* genes, which build the repressive components of one transcription-translation feedback loop [454-459]. PER and CRY proteins build a heterodimer, translocate from the cytoplasm into the nucleus, where they interact with the CLOCK-BMAL1 dimer to inhibit its transcription activation ability [458]. Ubiquitin-dependent degradation of PER and CRY

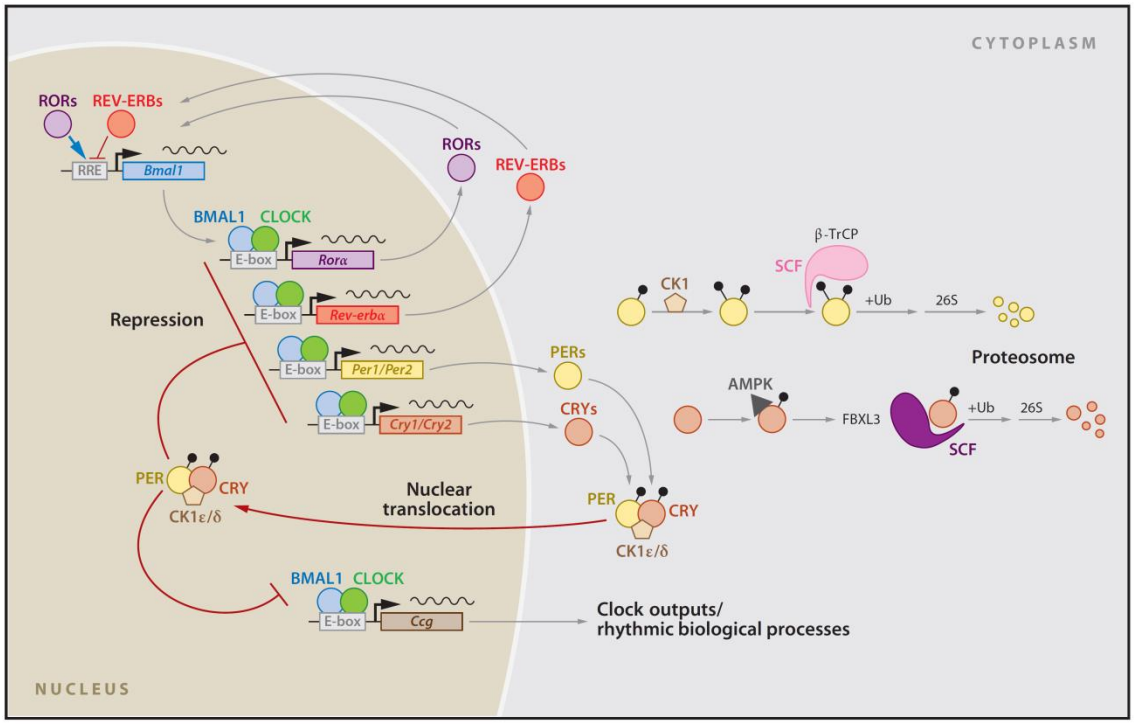



***Figure 25: Molecular regulation of the mammalian circadian clock***

The circadian rhythmicity is established by two cell-intrinsic autoregulatory transcription-translation feedback loops. One feedback loop is composed of the activating dimer CLOCK-BMAL1 and their target genes *Per* and *Cry*, whose products close the negative feedback loop. PER and CRY stability are regulated post-translationally by the E3 ubiquitin ligase SCF (Skp1-Cullin-F-box protein) and the kinases CK1 (Casein kinase 1) and AMPK (AMP kinase). Another feedback loop contains the CLOCK-BMAL1 targets *Ror* and *Rev-erb*. Furthermore, the CLOCK-BMAL1 dimer activates more target genes that are referred to as clock-controlled genes (CCGs). Source: [460]

lead to the derepression of the CLOCK-BMAL1 complex, so that the cycle with a periodicity of around 24 h can proceed [461-464]. This feedback loop is strongly post-translationally regulated. For example, kinases/phosphates, (de)acetylases or E3 ubiquitin ligases target PER and CRY proteins to control their cellular localization and stability [464-469]. Also post-translational regulated proteosomal degradation of CLOCK-BMAL1 seems to impact on robust circadian transcription of their downstream targets [470]. In a second interlocked feedback loop, the CLOCK-BMAL1 targets Nuclear receptor ROR (Retinoid-related orphan receptor, RORα, β, γ; activators) and Nuclear receptor subfamily 1 group D members (NR1D1 (REV-ERBα), NR1D2 (REV-ERBβ); repressors) assure the rhythmic expression of BMAL1 [471-476].

These transcription-translation feedback loops are required for the maintenance of the cell intrinsic circadian rhythm. CLOCK-BMAL1 as well as the other components of the core clock network further control the transcription of other genes [479-481]. The down-stream targets can themselves be transcription factors inducing rhythmic expression of their targets or be involved in diverse cellular processes, thereby establishing the circadian phenotype of the cells (Figure 26). While the expression of nearly half of all protein-coding genes have been observed to oscillate in some tissue of the body, the set of clock controlled genes in a cell are largely cell-type specific [482-485].

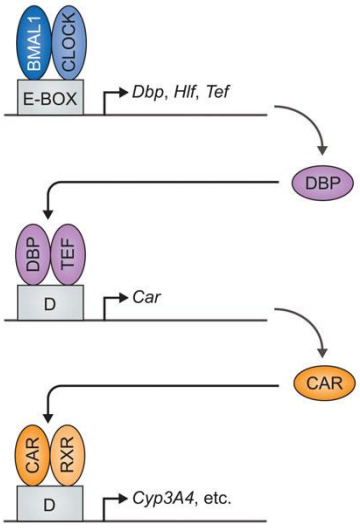



***Figure 26: Circadian transcriptional cascades***

CLOCK-BMAL1 induce circadian expression of the transcription factors DBP (D site-binding protein) and TEF (Thyrotroph embryonic factor), which themselves regulate the circadian transcription of the Constitutive androstane receptor *(Car, Nr1i3)*; CAR controls circadian xenobiotic metabolism via its downstream targets [477]. Source: [478]

Genome-wide binding studies for the core clock components and for the RNA polymerase II during circadian rhythm suggest a global circadian regulation of thousands of expressed genes, independent of a detectable cycling transcript, with temporally separated phases along the day [479]. The cycle contains a transcriptionally poised state around the circadian time (CT) 23 h to 4 h with the initiation form of RNA polymerase II and CRY1 binding at the promoters, an activation phase at CT 5 h to 10 h with CLOCK and BMAL1 binding, which then activates transcription peaking at CT 15 h and a final repression phase at around CT 15 h to 22 h with decreased CLOCK-BMAL1

binding and highest PER1, PER2 and CRY2 occupancy (Figure 27) [479]. While these results were obtained for the internal clock, the authors of another study reported bimodal transcription phases of RNA polymerase II in mice kept under 12 h light and 12 h darkness conditions [486]. More studies regarding polymerase II activity will be required to fully understand the transcriptional regulation during diurnal cycles.

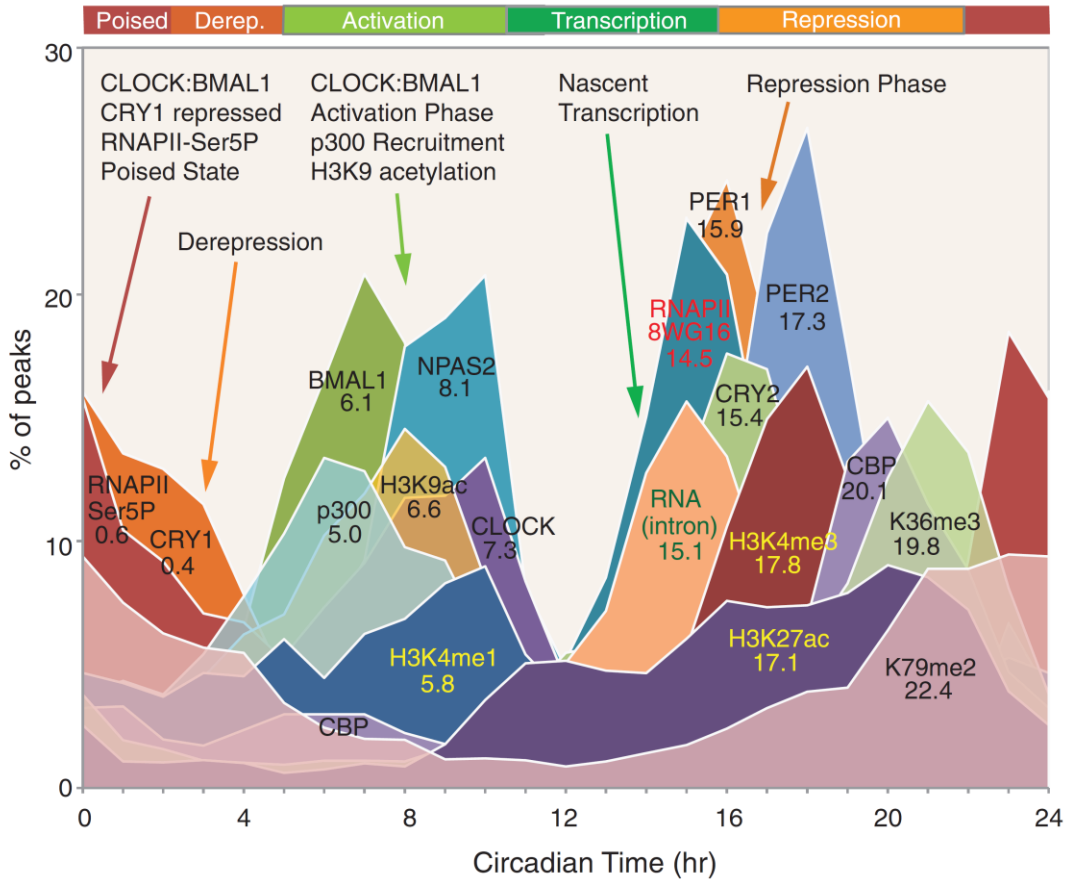



***Figure 27: Circadian transcriptional cycle***

Core clock factors, RNA polymerase II, some chromatin modifiers as well as some histone modifications occur at the DNA at specific times during the day and thereby lead to circadian transcriptional cycles. Source: [479]

In line with many other reports, both studies further point to the importance of RNA half-lives and post-transcriptional mechanisms in the establishment of the circadian expression variety that includes cycling of mature RNAs distributed across all phases [479, 486-489]. These post-transcriptional mechanisms can modulate the expression patterns during the day so that circadian transcribed genes may finally not oscillate on mRNA or protein level or that a constantly transcribed gene is cyclically expressed on mRNA or protein level [490]. Non-transcriptional mechanisms have been shown to be sufficient to sustain circadian rhythm for a certain time and circadian post-translational

modifications also occur in red blood cells, in which no transcription is performed due to the lack of DNA [491, 492].

In summary, these results indicate that transcription is widely circadianly regulated, while rhythmic gene expression is regulated on top via additional mechanisms.

### 1.3.3 Epigenetic modulations can occur in a circadian pattern

Associated with the observed circadian transcriptional cycles are rhythmic chromatin alterations. Similar to pioneer transcription factors, CLOCK-BMAL1 can bind to its targets sites within nucleosomes and then stimulates nucleosome removal to further allow binding of other transcription factors that influence the transcriptional output and contribute likely to the tissue specificity of the circadian expression (Figure 28) [493]. As the TSS are in general more nucleosome depleted, these changes are especially predominant if CLOCK-BMAL1 binding occurs in gene bodies or intergenic regions [493]. These rhythmic changes in chromatin accessibility are associated with chromatin modulations such as oscillating occurrence of H2A.Z and histone modifications at the CLOCK-BMAL1 binding sites [493].

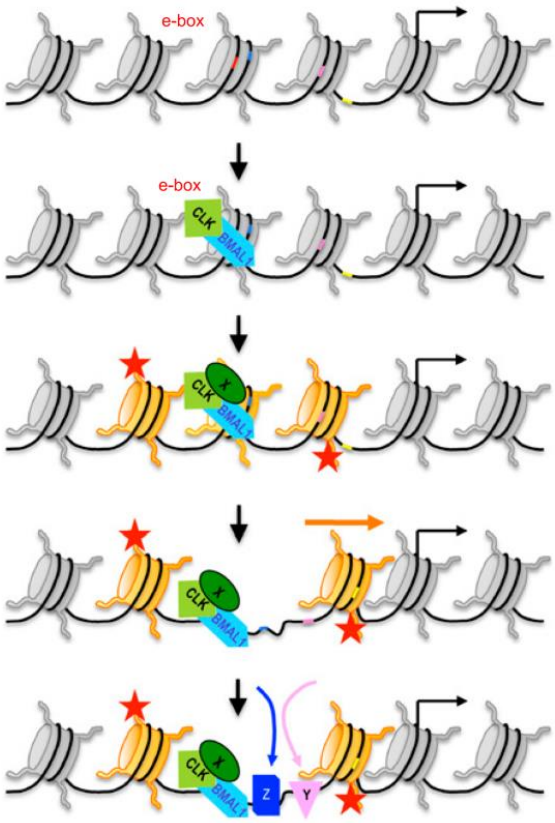



***Figure 28: CLOCK-BMAL1 binding induces rhythmic chromatin changes***

CLOCK-BMAL1 binding to E-box elements results in rhythmic chromatin opening, histone variant exchange and post-translational modifications. Source: [493]

For example, the presence of the histone modifications H3K9ac, H3K4me3 and H3K27ac at the transcription start site of most expressed genes shows robust cycles with highest enrichment at the activation or transcription/repression phase [479]. The occupancy of the transcription factors, histone modifications and the RNA polymerase II at intergenic enhancers was observed to be circadian modulated like for promoters, but may differ for some histone modifications in phase [479, 493]. The occurrence of

other histone marks, such as the enhancer mark H3K4me1, the with inactive chromatin associated mark H3K9me3 or the transcription elongation marks H3K36me3 and H3K79me2, are barely or less pronounced circadianly affected [479, 486, 494].

The establishment of the circadian histone modification occurrence is regulated by various mechanisms, as it can be exemplified for H3K4me3. Two histone methyltransferases, Mixed lineage leukemia 1 and 3 (MLL1 and MLL3; also named Histone-lysine N-methyltransferase 2A and 2C (KMT2A and KMT2C)) have been connected to the ability to establish circadian H3K4me3 patterns; however, they function in different manners. Both methyltransferases require an intact clock network, but only MLL3 is circadian expressed [494, 495]. Recruitments of both enzymes to promoters are circadian, but only MLL1 is co-recruited with the CLOCK-BMAL1 complex and consequently directly associated to histone acetylation, since the transcription factor CLOCK itself is a histone acetyltransferase [494-496]. The activity of MLL1 relies on its acetylation level, which is controlled in a circadian fashion by the $NAD^+$-dependent protein deacetylase sirtuin-1 (SIRT1), providing a link of circadian histone modifications to the energy metabolism [497]. SIRT1 associates with CLOCK and deacetylates acetylated BMAL1, H3K9 and H3K14 in a circadian manner [498].

Post-translational histone modifications seem also to be involved in the proper process of the transcription-translation feedback loops underlying circadian rhythm. It has been described that during the negative feedback loop, Per proteins recruit first the histone deacetylase HDAC1 and 4 h later the HP1γ-associated Histone-lysine N-methyltransferase SUV39H1 (SUV39H1), which catalyzes the methylation of H3K9, to the *Per1* and *Per2* promoters to repress the transcription of these genes [129, 499]. Furthermore, CLOCK-BMAL1 mediated H2B mono-ubiquitinylation may link the positive and negative feedback loops by establishing chromatin changes that promote PER complex association and CLOCK-BMAL1 complex dissociation [500].

The acetyltransferase CBP/Ep300 has been reported to co-occur with Clock and to contribute to the CLOCK-BMAL1 mediated transcription of Period and Cryptochrome genes [501]. Furthermore, Ep300-mediated modulation of the CLOCK-BMAL1 transcriptional activity might occur in a cell-type specific manner based on its interaction with different co-factors [502].

Also the histone deacetylase HDAC3 has been shown to be critical for circadian rhythm. The circadian HDAC3 occupancy is regulated by REV-ERBα and is phase-shifted with the RNA polymerase II recruitment to the respective sites [503]. HDAC3 contributes independent of its enzymatic activity to the robustness of circadian gene expression by modulating the function of key core clock factors during the repression as well as activation phase of the circadian rhythm [504].

Studies of the clock-controlled gene D site-binding protein *(Dbp)* illustrate the interplay of transcriptional mechanisms with epigenetic regulation and furthermore exemplify that besides post-translational histone modifications, also other epigenetic mechanisms are involved in the regulation of circadian rhythm. The binding of CLOCK-BMAL1 to the E-box elements of *Dbp* is accompanied by H3K9 acetylation, H3K4 trimethylation and a gain of chromatin accessibility, while during the repressive phase H3K9 is dimethylated, Heterochromatin protein 1α (HP1α) binds and the nucleosome density increases [505, 506]. Moreover, not only local chromatin remodeling has been described at the *Dbp* locus, but also changes in its long-range interactome during the circadian rhythm [507]. This indicates that the nuclear topological organization may undergo diurnal alterations.

DBP is a D-box binding transcription factor that acts mainly downstream of the core clock network and can induce cyclical transcription of its target genes likely by its cyclical occupancy at promoters and distal regulatory elements [481, 505, 508-510]. The impact of enhancers in circadian transcription regulation have been highlighted in a study analyzing rhythmic enhancer RNA expression during circadian rhythm in mouse liver and comparing it to nearby gene transcription [481]. They revealed six enhancer groups whose activities are controlled by distinct transcription factors enriched in the particular phase such as the binding of D-box binding transcription factors between ZT 9 and ZT 15 or Rev-erbα between ZT 18 and ZT 24 (Figure 29) [481]. The enrichment of the E-box motif between ZT 6 and ZT 9 is moreover in line with a previous report of BMAL1 occupancy at enhancer around CT 8 [479, 481].

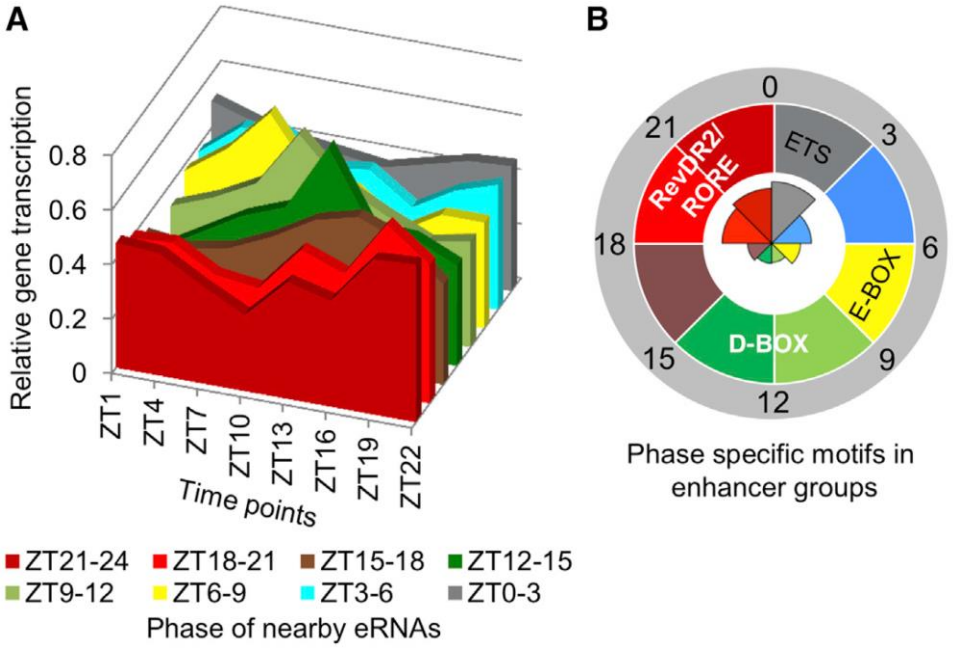



***Figure 29: Phase-specific transcription factors activate enhancers resulting in circadian transcription of nearby genes***

A) Oscillating eRNAs in mouse liver are grouped based on their phase and the relative gene transcription of the closest genes are illustrated (within 200 kb of TSS). B) Motifs that are specifically enriched in certain eRNA groups are highlighted in a clock diagram.
Source: [481]

Besides eRNAs, other non-coding RNAs are also involved in circadian rhythm. For example, miRNAs can adjust the rhythmic transcriptome post-transcriptionally, also of core clock members [511-513]. Also long non-coding RNAs with circadian expression have been reported, among them are also antisense transcripts of the core clock genes *Per1*, *Per2* and *Per3* [479, 482, 514-516]. In Neurospora, the sense and antisense transcription of a core clock component establishes a regulatory mechanism influencing the core clock network [516-519]. However, the role of lncRNAs in the regulation of circadian gene expression in mammals is so far not explored in detail.

DNA methylation has been implicated in the entrainment of the circadian rhythm. The promoter DNA methylation as well as the transcription has been found to reversibly change in the SCN under altered environmental conditions, indicating that DNA methylation might be involved in regulating circadian clock plasticity in the SCN [520].

The provided examples elucidate that epigenetic regulators themselves are often regulated in a circadian manner and that the sustainment of circadian rhythm is regulated by several feedback-loops encompassing transcriptional, epigenetic and post-transcriptional mechanisms.

## 1.4 Embryonic neurogenesis

### 1.4.1 Development of the central nervous system

The life of a multicellular organism starts with the fertilization of an egg cell by a sperm. This diploid zygote undergoes multiple divisions to form the blastula. The inner cell mass of the blastula contains pluripotent cells, which can self-renew and can also give rise to the different cell types of the body. During gastrulation, the cells are arranged into distinct cell layers, which will give rise to the three different germ layers, namely ecto-, meso- and endoderm [521]. Under the control of certain stimuli, the cells of the different layers will then differentiate spatiotemporally tightly organized into distinct cell types. The ectoderm undergoes neurulation and results into three major parts, the epidermal ectoderm, the neural crest cells that are precursors of the peripheral nervous system and the neural tube cells that will become the central nervous system (CNS) [522]. The neural tube develops into distinct regions of the CNS, namely the pros-, mes- and rhombencephalon, which will give rise to the brain and the spinal cord [523]. The neocortex, which fulfills sensory, motoric and cognitive functions, emerges from the telencephalon, a part of the prosencephalon, and forms histologically a six-layered structure in mammals [524]. The neocortex consists of two major classes of neurons originating from different regions of the telencephalon [525-527]. The inhibitory, GABAergic interneurons are largely generated in the subpallial (ventral) proliferative zone of the telencephalon (ganglionic eminences) and then migrate tangential into the neocortex, where they regulate local circuits [528-530]. The excitatory, glutamatergic projection neurons are instead derived from the cortical, pallial (dorsal) neuroepithelium, migrate radial towards the pial (basal) surface and project to cortical and subcortical brain regions [531, 532].

The cortical neurogenesis in a mouse starts at E10.5, when neuroepithelial (progenitor) cells (NE, NPE; neural stem cells, NSCs) not only proliferate to expand the neuroepithelial layer, but also give rise to radial glial cells (Figure 30) [524, 533, 534]. These cells exhibit, like the neuroepithelial cells, apical-basal polarity with processes to the apical (ventricular) and basal (pial) surface, but also additional astroglial properties [534, 535]. Radial glial cells further divide most of the time asymmetrically to give rise to another radial glial cell and to a more committed neural progenitor cell or directly to a neuron [536-538]. Radial glial cells, for example, generate basal (outer) radial glial

cells or intermediate progenitors that can further proliferate and differentiate into neurons [539-547].

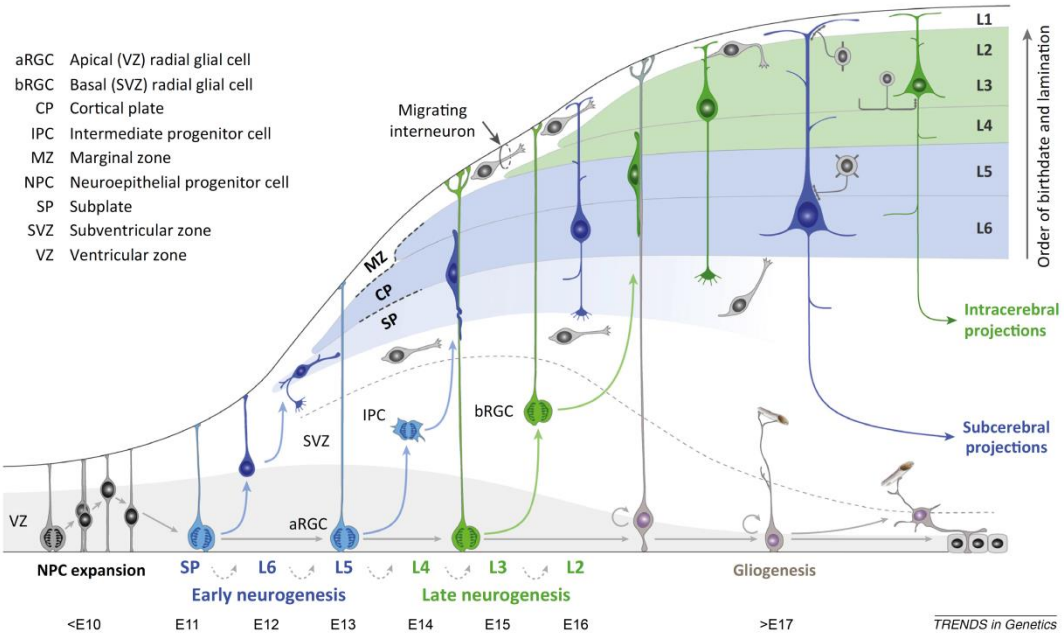



***Figure 30: A schematic illustration of cortical neurogenesis***

Radial glial cells and their derivates differentiate into neurons, which migrate along the processes of radial glial cells to the outer layer. This process results into the six-layered structure of the cortex with an "inside-out" fashion. The embryonic day is mentioned below (e.g., E11 for embryonic day 11).
Source: [548], adjust according to [549]

These processes are not only temporally, but also spatially well defined. The neuroepithelial and apical radial glial cells build the lowest layer, which is named the ventricular zone (VZ). The progenitors emerged from radial glial cells such as basal radial glial cells and intermediate progenitors, which have retracted their processes, move dorsal of the VZ and build the subventricular zone (SVZ) (Figure 31). Neurons finally migrate along the processes of the radial glial cells to the cortical plate (CP), whereby the newer neurons are always moving on the top layer of the cortical plate ("inside-out fashion") [533, 550]. Furthermore, different subtypes of neurons are produced at different developmental stages, whereby the molecular mechanisms of defined linage commitments remain widely unclear [524, 544]. Also spatial information introduced by morphogen and signaling molecule gradients contributes to differential transcription factor expression in progenitors from different regions, which is important for regionalization of the cortex and may also influence neuron type specification [524, 551, 552]. After the embryonic neurogenesis process is accomplished at around E17.5, radial glial cells will give rise to glial cell types like astrocytes or oligodendrocytes to

finally complete the complex network of the CNS [532, 533, 536, 539, 553, 554]. The development of the brain in terms of cell number and cell types is therefore highly dependent on the controlled proliferation and differentiation of neural stem cells and progenitor cells [534, 544].

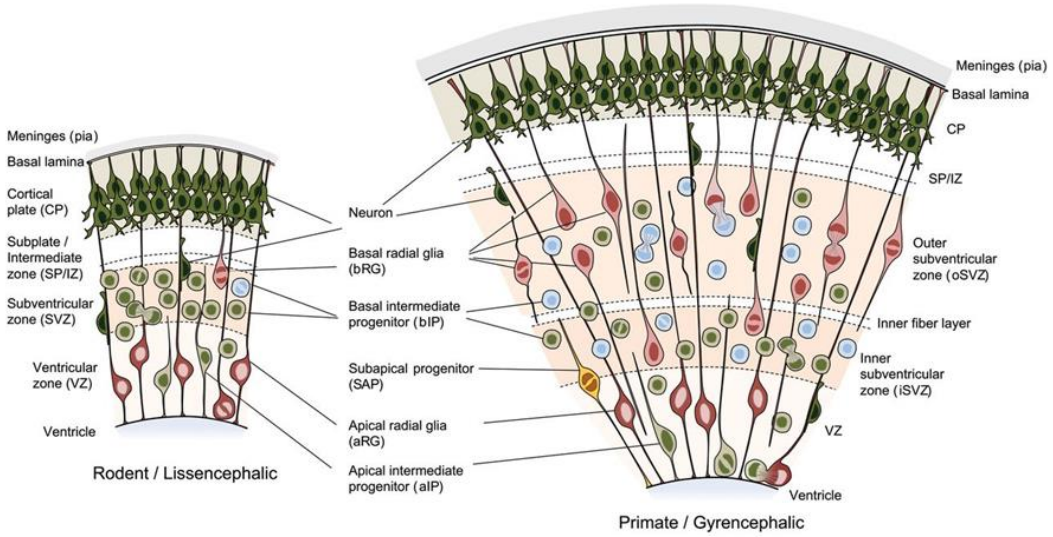
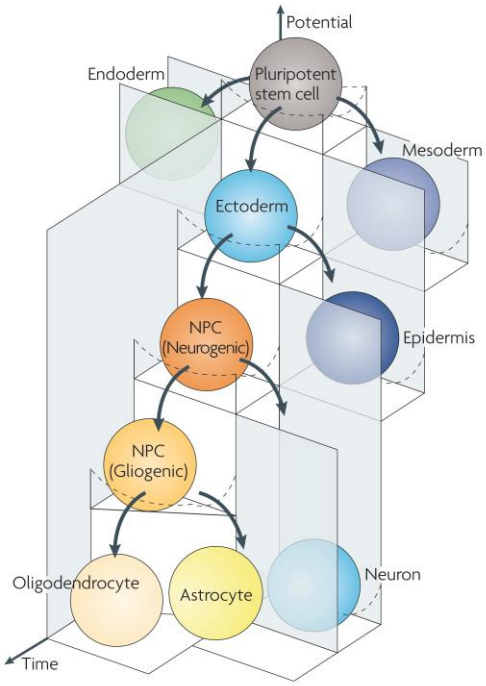



***Figure 31: Neural progenitor types of the developing neocortex in rodents and primates***

The graphics shows different progenitor types that occur during development in the diverse layers of the neocortex in rodents (lissencephalic cortex) and in primates (gyrencephalic cortex). Source: [544]

### 1.4.2 Commitment of neural progenitors to neurogenesis




During the embryonic development pluripotent stem cells first limit their differentiation potential towards the different linages and then undergo further stepwise restrictions in their potencies to differentiate into other cell types likely by long-term repression of developmental genes required for other linages and cell types (Figure 32) [555]. For example, they generate neuroepithelial cells, which give rise to progenitor cells of neurons, astrocytes and oligodendrocytes. These neural progenitor

***Figure 32: Stepwise restriction of the differentiation potential of stem cells during development***

Pluripotent stem cells commit first to the different linages, namely endo-, meso- and ectoderm and then further differentiate in the different cell types. For example, ectodermal stem cells differentiate into epidermal cells or into neural progenitor cells (NPCs). These give first rise to neurons and then later on to oligodendrocytes and astrocytes.

Source: [555]

cells proliferate further, in order to be able to produce the large amount of brain cells. However, at distinct time points these progenitors commit to neural cell fate and differentiate directly or via a more committed progenitor cell into functional, post-mitotic neurons. In order to coordinate the proliferation, cell cycle exit and differentiation of neural progenitors in a spatiotemporal manner to produce the appropriate number of neurons and glial cells, the gene expression needs to be tightly controlled by signaling, transcriptional and epigenetic mechanisms (Figure 33).

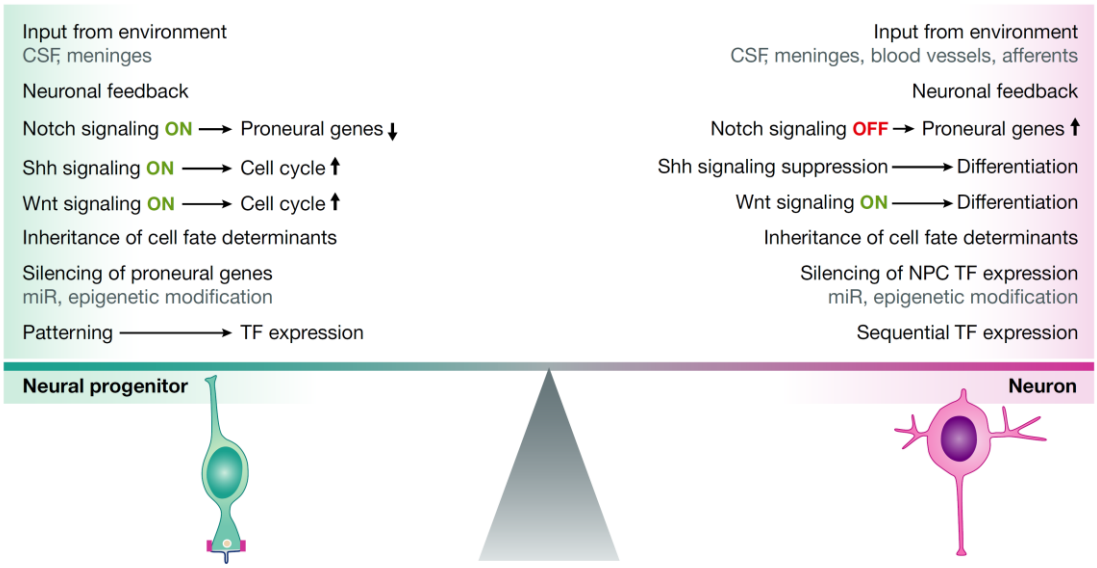



***Figure 33: Extra- and intracellular factors regulate the balance between neural progenitor proliferation and differentiation***

Signals from the environment as well as internal factors regulate the fate of neural progenitor cells. Among them are different signaling pathways, epigenetic mechanisms and transcription factors (TF). They determine if the progenitor cells keep proliferating or differentiate into neurons. Source: [550]

### 1.4.3 Transcriptional control of neural progenitor fate

The precise transcriptional gene regulation in neural progenitor cells is essential for controlling the coordinated expansion of the cortex together with the timely onset of differentiation. Several transcription factors fulfill key functions in the regulation of cell cycle, self-renewal and neuronal differentiation of neural progenitor cells.

The Paired box protein Pax-6 (Pax6) is one essential transcription factor that controls among other functions the balance between self-renewal and differentiation of neural progenitors [556]. This factor comprises two contradictory functions: on the one hand it promotes the proliferation of neural progenitor cells and on the other hand it induces neurogenesis [557]. It has been suggested that this dual function of Pax6 is achieved by distinct subdomain activity, interaction of PAX6 with varying transcription factors or by

induction of different target genes based on the expression level of PAX6 [557-560]. For instance, PAX6 targets genes promoting neural progenitor self-renewal (e.g., Cyclin-dependent kinase 4 *(Cdk4)*, High mobility group protein HMGI-C *(Hmga2)*); however, at high concentration, Pax6 induces genes driving neurogenesis such as the proneural gene Neurogenin 2 *(Ngn2)* (Figure 34) [558, 559]. Murine cortical progenitors depleted for PAX6 or for other factors like LIM/homeobox protein Lhx2 (LHX2), Homeobox protein ARX (ARX), Homebox protein EMX1 and EMX2 (EMX1, EMX2), Forkhead box protein G1 (FOXG1, BF-1) and Nuclear receptor subfamily 2 group E member 1 (NR2E1, TLX) exhibit defects in progenitor propagation and brain growth [561-569]. Therefore, these factors seem to be important for neural progenitor proliferation.

Also the initiation of the transition from radial glial cells to more committed progenitor cells or to neurons needs to be coordinated. Basic helix-loop-helix (bHLH) proneural factors like Neurogenin (NGN, NEUROG) and Achaete-scute homolog 1 (ASCL1, MASH1) are essential for the neuronal fate decision as well as neurogenesis and start to be expressed around the time of radial glia generation [570-573]. These transcription factors are even sufficient to induce neuron formation by ectopic overexpression and have further been shown to exhibit the potential to transdifferentiate or reprogram non-neuronal cells into neurons [574-580]. While ASCL1 is especially required for the induction of GABAergic interneurons in the ventral telencephalon, NGN2 drives neurogenesis of glutamatergic projection neurons in the dorsal telencephalon and suppresses the differentiation into GABAergic neurons [573, 581-583]. Similar to PAX6, also the bHLH factors can fulfill different functions regulated via post-translational modifications and expression patterns [573]. For example, oscillating NGN2 or ASCL1 expression promote the cell cycle progression of neural precursor cells and sustained expression goes along with cell cycle exit and neuronal differentiation [584-587].

Radial glial cells express the transcription factor PAX6, which induces *Ngn2* expression that in turn activates the expression of the T-box brain protein 2 (*Tbr2*, *Eomes;* Eomesodermin homolog) [558, 588]. The up-regulation of TBR2 is associated with the transition of radial glial cells to intermediate progenitors or basal progenitors [541, 589, 590]. Also the Insulinoma-associated protein 1 (Insm1) and the Transcription factor AP-2 gamma (Tfap2c, AP2γ) induce the expression of *Tbr2* and thereby regulate the generation and the cell fate of basal progenitors (Figure 34E) [591, 592]. NGN1 and

NGN2 further stimulate the expression of genes that promote cell cycle exit and neuronal differentiation; among them are bHLH differentiation factors such as Neurod transcription factors [582, 583, 593, 594]. These factors and T-box brain protein 1 (TBR1) are critical regulators during the generation of post-mitotic neurons [589, 595-597].

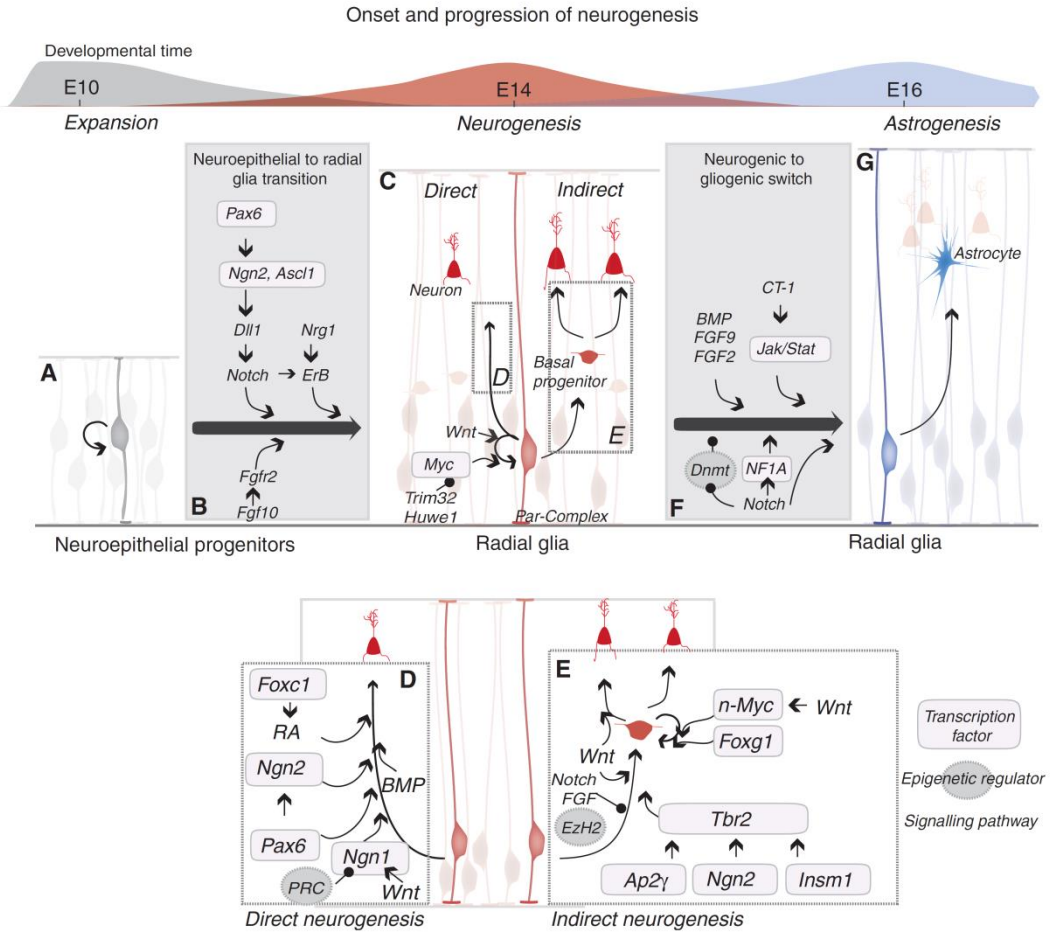



***Figure 34: Illustration of the molecular regulation of the onset, progression and termination of neurogenesis in the rodent cerebral cortex***

During cortical neurogenesis neuroepithelial progenitors (A) give rise to radial glial cells (C), which are, for example, strongly regulated by the transcription factor PAX6 and the Notch signaling pathway (B). Radial glial cells then generate neurons either directly (D) or indirectly via another more committed progenitor cell (E). Here, the proneural factor NGN2 plays an essential role. After the termination of neurogenesis, radial glial cells produce glial cells (F, G). Each of these steps is regulated via signaling pathways, transcription factors and epigenetic mechanisms of which some crucial factors are depicted in the illustration. Source: [567]

Another layer constitutes signaling pathways that regulate the identity and fate of neural progenitor cells. Among these are, for example, Notch, Sonic hedgehog (SHH), Wnt/β-catenin, bone morphogenic protein (BMP), fibroblast growth factor (FGF) and retinoic acid (RA) signaling [550, 567]. Gradients of morphogens from these signaling

pathways lead to regional specific expression patterns in the neuronal precursor cells and furthermore influence the production of specific neuron types [567, 598, 599].

The Notch signaling pathway (Figure 35A) is induced at the time of the transition from neuroepithelial cells to radial glial cells (Figure 34B) [567]. It is important for the maintenance of radial glial cells and its repression promotes proneural genes and consequently direct or indirect neurogenesis (Figure 33 and 34C) [600-608]. The Notch signaling activates Hes factors, which in turn repress the expression of proneural genes such as *Ascl1* and *Ngn2,* thereby inhibiting neurogenesis and promoting the radial glia maintenance [606, 609, 610]. Furthermore, it has been shown that the Notch target gene Transcription factor HES-1 *(Hes1)* exhibits an oscillating expression due to a negative feedback loop leading to an anti-phase oscillation with respect to the Notch ligand Delta-like protein 1 (Dll1), Ngn2 and Ascl1 [584-586, 611]. At the onset of neurogenesis, the oscillations stop due to still unknown reasons and while *Hes1* keeps repressed, *Ngn2* and *Dll1* show constant high expression levels [585]. During the asymmetric division of radial glial cells, the daughter cell with the high Notch signaling remains a radial glial cell, while the other one with the low Notch signaling has high expression levels of Dll1 and proneural genes and starts to differentiate (Figure 35B) [550, 585, 588, 612]. The lateral inhibition of the cell adjacent to the differentiating cell ensures that an appropriate number of progenitors is maintained throughout the embryonic development [613] .

These as well as several other factors and signaling pathways build a complex network regulating the identity and cell fate change of neural progenitor cells, which need to integrate all these signals to the proper response that will ultimately lead to the formation of the complex cortex structure. Nevertheless, more research is necessary to completely understand the regulatory relationships between the involved factors.

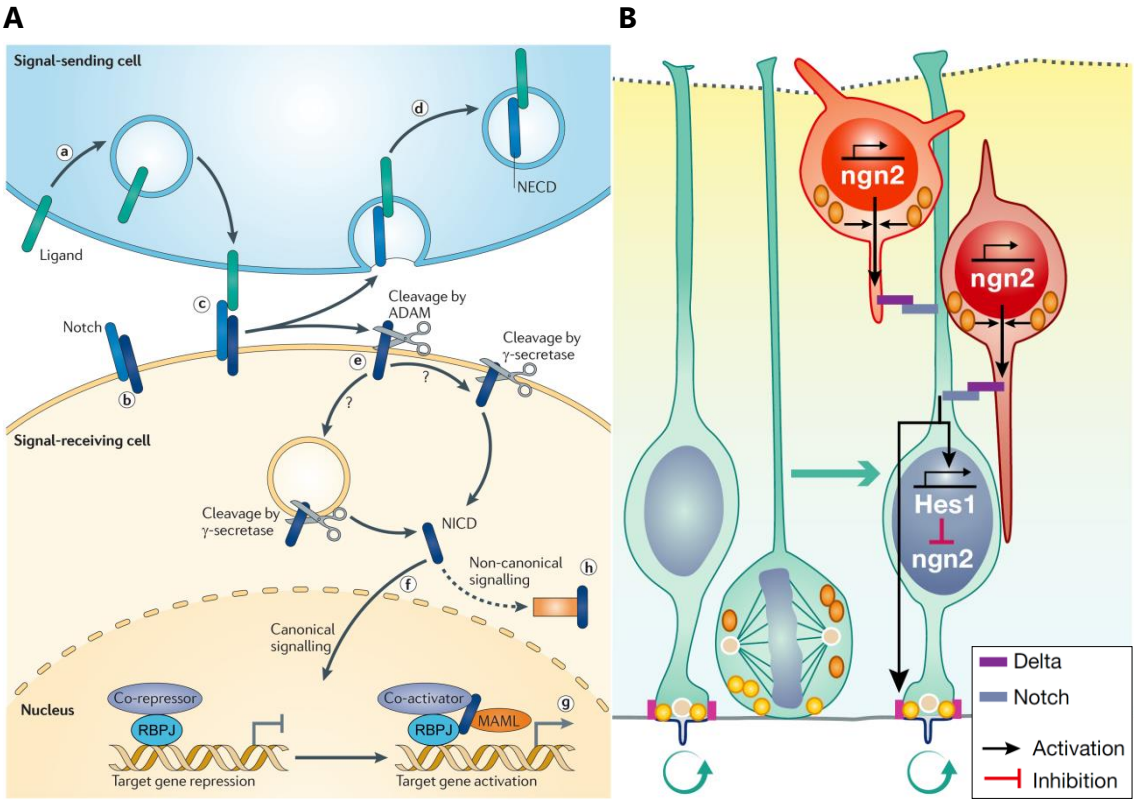



***Figure 35: The Notch signaling pathway is important for neural progenitor proliferation***

A) Notch ligands (green, e.g., Delta-like protein 1 (DLL1) or Protein jagged-1 (JAG1)) are endocytosed from the membrane of the 'signal-sending cell' (a). After re-establishing of the ligand at the membrane (c), it can bind to Notch (blue, heterodimer) that is in the cell membrane of the signal-receiving cell (b). According to one model, the ligand and the Notch extracellular domain (NECD) are endocytosed by the 'signal-sending cell' (d). The Notch intracellular domain (NICD) is cleaved in the 'signal-receiving cell' by Disintegrin and metalloproteinase domain-containing protein (ADAM) and γ-secretase (e). The cleaved NICD translocates to the nucleus (f), where it leads to the dissociation of the repressor complex from the Recombining binding protein suppressor of hairless (RBPJ). NICD becomes stabilized to RBPJ by the Mastermind-like proteins (MAML) and recruits co-activator complexes to initiate the transcription of the Notch target genes (e.g., Transcription factor HES-1 *(Hes1)* or HES-5 *(Hes5)*) (g). Notch signaling can also act via non-canonical pathways (h). B) During asymmetric division of radial glial cells, the daughter cell with high Notch signaling expresses Hes1 and remains a radial glial cells, while the daughter cell with low Notch signaling expresses the proneural gene Ngn2 and becomes a more committed progenitor cell or a neuron. Source: A adapted from [614], B [550]

### 1.4.4 Epigenetic control of the onset of neurogenesis

Epigenetic regulators also fulfill, besides and in cooperation with transcription factors, essential functions in the regulation of neural progenitor identity and during neurogenesis.

A number of genes are essential for the development and lineage specification of a cell and their expression is known to be epigenetically regulated. Many of these genes have been reported to occur in an epigenetically poised state in embryonic stem cells. This so-called bivalent state is characterized by the occurrence of H3K4me3 and H3K27me3 at the promoter of these genes, resulting in repression or only low expression of the genes [615, 616]. During *in vitro* neurogenesis, it was observed that new bivalent states

were obtained while others are lost at the transition from embryonic stem cells to neuronal progenitors as well as from neuronal progenitors to neurons (Figure 36) [617]. Examples for an activation of bivalent genes during neurogenesis are, for example, *Pax6*, *Ascl1* and *Ngn1* [148, 555]. Genes important for other lineages lose H3K4me3 during neurogenesis and become repressed [618].

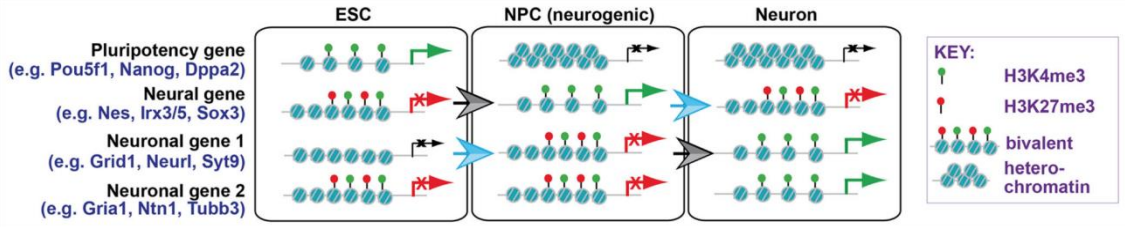



***Figure 36: Post-translational histone modification status of cell state specific genes during in vitro neurogenesis***

H3K4me3 and H3K27me3 occurrence at cell state specific genes in embryonic stem cells (ESC), in neuronal precursor cells (NPC) and in neurons. Gray arrow: transition from bivalent to active state, blue arrow: acquirement of bivalent state. Source: [619], based on data from [617] (Note: [617] measured H3K4me2 and not H3K4me3 as shown in the figure, but they showed that H3K4me2 positive promoters highly overlap with H3K4me3 positive promoters in stem cells.)

The importance of H3K27me3 for the regulation of the transition from neural precursor state towards neurogenesis has also been demonstrated *in vivo* by depleting the Histone-lysine N-methyltransferase EZH2 (Enhancer of zeste homolog 2, EZH2), which is part of the Polycomb repressive complex 2 (PRC2) and catalyzes the methylation of H3K27, in cortical progenitor cells from embryonic day 9.5 (E9.5) onwards [620]. The absence of the repressive mark causes an up-regulation of many transcripts and leads to a premature initiation of neurogenesis [620]. Also the Lysine-specific demethylase 6B (Kdm6b, Jmjd3), which is the demethylase of H3K27, seems to play a role in neuronal fate decision [621, 622].

The poised state of developmental genes in stem or progenitor cells maintains their capacity for future activation and thereby determines the differentiation potential of the cell [555]. With the progression of differentiation and fate restriction, the pluripotency, developmental and tissue-specific genes undergo long-term or permanent silencing, for example by methylation of H3K27, H3K9 and DNA [555, 623-627].

DNA methyltransferases seem to be critical regulators of neural cell fate commitment. For example, DNMT1 and DNMT3A seem to be important in neural progenitor cells to suppress astroglial differentiation during the neurogenic period [628, 629]. DNMT3A

knock-out neural stem cells derived from mouse embryonic stem cells furthermore exhibit increased cell proliferation [629]. DNMT3B is transiently expressed during embryonic development and its expression in neural progenitor cells might be important during early neurogenesis for the timing of neuronal differentiation and maturation [71, 630-632]. On the other hand, DNA demethylation of the Notch signaling target Hes5 has been shown to be required for the transition from neuroepithelial cells to neural precursor cells [633]. The expression of the Methylcytosine dioxygenase TET3 in neural progenitor cells seems to be important for their maintenance as well as for their differentiation into neurons and is controlled by the microRNA-15b [634, 635].

Also many other studies have implicated non-coding RNAs such as miRNAs and lncRNAs in the regulation of neurogenic commitment as well as brain development and function [636-647]. For example, the Let-7 miRNAs were described to regulate proliferation and differentiation of neural precursor cells [648-651]. The lncRNA *Pnky,* in cooperation with the splicing regulator Polypyrimidine Tract-Binding Protein 1 (PTBP1), promotes the proliferative state of neural stem cells by regulating the expression and alternative splicing of transcripts that are key for the differentiation [652].

Furthermore, local and global chromatin rearrangements have been observed to occur during the differentiation of cells. For example, changes in the chromatin three-dimensional structure such as reorganization of higher-order domains (metaTADs) and of the chromatin regions that interact with the nuclear lamina contribute to the expression changes during cellular differentiation [295, 653-655]. Also chromatin accessibility changes, especially at distal regulatory regions such as enhancers, have been demonstrated to play an essential role in the gene expression regulation during linage commitment and neurogenesis [239, 655, 656]. Dynamic enhancer utilization during neurogenesis is essential for the process and the brain development [238, 239, 242-244].

It has been described that different subunits of the BAF (mSWI/SNF) chromatin remodeling complexes are assembled in a cell-type specific fashion in embryonic stem cells, neural progenitors and neurons, which seems to be crucial for the transition of the different cell stages during neurogenesis and for fulfilling stage specific functions [657-661]. The BAF complex of neuronal progenitors (npBAF) contributes to the proliferation

of progenitor cells, possibly by enhancing the Notch signaling pathway and/or interaction with PAX6, and thus to the regulation of the size and thickness of the neocortex [657, 661-664]. The catalytic subunit of the BAF complexes BRG1 has also been shown to mediate the transcriptional activities of the proneural bHLH genes [665]. The switch from npBAF to neuron-specific BAF (nBAF) is accompanied by the subunit switch from BAF53a to BAF53b, which is mediated by the derepression of the microRNAs mi-R9* and miR-124 due to reduced activity of the Repressor-element-1-silencing transcription factor (REST, NRSF) in neurons [666]. REST forms complexes with chromatin-binding proteins (e.g., Methyl-CpG-binding protein 2 (MeCP2)) and chromatin modifiers (e.g., histone deacetylases) to repress neuronal genes in non-neuronal cells and it regulates neurogenesis by the subsequent repression of its expression from stem cells to neuronal progenitor cells to neurons [555, 667-671].

Not only the expression of *Pax6* is epigenetically regulated, for example by post-translational histone modifications, chromatin remodeling or by the lncRNA *Paupar*, but also PAX6 functions in cooperation with epigenetic players [555, 556, 672, 673]. In this connection, PAX6 binds several chromatin modifiers such as the histone acetyltransferase Ep300 and subunits of the BAF complexes [556].

A recent report demonstrated a link between the transcription factor action and the epigenetic regulation. This study shows that epigenetic silenced neuronal genes can be activated by the transcription factor Neurod1 thereby initiating neurogenesis [674]. Neurod1 induces reprogramming of the chromatin landscape of these genes from an inactive into an active status and these features are maintained even after the expression of Neurod1 is reduced again [674].

These provided examples of epigenetic mechanisms and their connections between each other and to transcription factors as well as signaling pathways show the complexity of the gene expression regulation during neurogenesis. They further illustrate that multiple mechanisms act together to ensure gene expression patterns during development necessary for proper cell differentiation.

## 1.5 Aims of the study

The aims of this thesis were to elucidate the gene expression control during three fundamental biological processes:

- To reveal the influence of UV exposure on chromatin accessibility and on histone acetylation, especially at regulatory regions and in relation to gene expression (chapter 1).
- To identify novel transcription factors and epigenetic regulators with circadian expression patterns and to reveal their potential gene regulatory function on circadian rhythm and on other cyclically expressed genes (chapter 2).
- To uncover the genomic targets and gene regulatory mechanisms underlying Pax6 function during neurogenesis (chapter 3).

The implications of the observed findings and perspectives for further studies were also discussed (discussion).

# 2 References


1. Rokas, A., *The origins of multicellularity and the early history of the genetic toolkit for animal development.* Annu Rev Genet, 2008. **42**: p. 235-51.
2. Grosberg, R.K. and R.R. Strathmann, *The evolution of multicellularity: A minor major transition?* Annual Review of Ecology Evolution and Systematics, 2007. **38**: p. 621-654.
3. Medvedev, S.P., A.I. Shevchenko, and S.M. Zakian, *Molecular basis of Mammalian embryonic stem cell pluripotency and self-renewal.* Acta Naturae, 2010. **2**(3): p. 30-46.
4. Dahm, R., *Friedrich Miescher and the discovery of DNA.* Dev Biol, 2005. **278**(2): p. 274-88.
5. Avery, O.T., C.M. Macleod, and M. McCarty, *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii.* J Exp Med, 1944. **79**(2): p. 137-58.
6. Hershey, A.D. and M. Chase, *Independent functions of viral protein and nucleic acid in growth of bacteriophage.* J Gen Physiol, 1952. **36**(1): p. 39-56.
7. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature, 1953. **171**(4356): p. 737-8.
8. Meselson, M. and F.W. Stahl, *The Replication of DNA in Escherichia Coli.* Proc Natl Acad Sci U S A, 1958. **44**(7): p. 671-82.
9. Moran, U., R. Phillips, and R. Milo, *SnapShot: key numbers in biology.* Cell, 2010. **141**(7): p. 1262-1262 e1.
10. Kornberg, R.D., *Chromatin structure: a repeating unit of histones and DNA.* Science, 1974. **184**(4139): p. 868-71.
11. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.
12. Cheema, M.S. and J. Ausio, *The Structural Determinants behind the Epigenetic Role of Histone Variants.* Genes (Basel), 2015. **6**(3): p. 685-713.
13. Woodcock, C.L. and R.P. Ghosh, *Chromatin higher-order structure and dynamics.* Cold Spring Harb Perspect Biol, 2010. **2**(5): p. a000596.
14. van Steensel, B., *Chromatin: constructing the big picture.* EMBO J, 2011. **30**(10): p. 1885-95.
15. Luger, K., M.L. Dechassa, and D.J. Tremethick, *New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?* Nat Rev Mol Cell Biol, 2012. **13**(7): p. 436-47.
16. Aguilar, C.A. and H.G. Craighead, *Micro- and nanoscale devices for the investigation of epigenetics and chromatin dynamics.* Nat Nanotechnol, 2013. **8**(10): p. 709-18.
17. Passarge, E., *Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago.* Am J Hum Genet, 1979. **31**(2): p. 106-15.
18. Bailis, J.M. and S.L. Forsburg, *It's all in the timing: linking S phase to chromatin structure and chromosome dynamics.* Cell Cycle, 2003. **2**(4): p. 303-6.
19. Probst, A.V. and G. Almouzni, *Pericentric heterochromatin: dynamic organization during early development in mammals.* Differentiation, 2008. **76**(1): p. 15-23.
20. Almouzni, G. and A.V. Probst, *Heterochromatin maintenance and establishment: lessons from the mouse pericentromere.* Nucleus, 2011. **2**(5): p. 332-8.
21. Craig, J.M., *Heterochromatin--many flavours, common themes.* Bioessays, 2005. **27**(1): p. 17-28.
22. Trojer, P. and D. Reinberg, *Facultative heterochromatin: is there a distinctive molecular signature?* Molecular Cell, 2007. **28**(1): p. 1-13.
23. Filion, G.J., et al., *Systematic protein location mapping reveals five principal chromatin types in Drosophila cells.* Cell, 2010. **143**(2): p. 212-24.
24. Ernst, J. and M. Kellis, *Discovery and characterization of chromatin states for systematic annotation of the human genome.* Nat Biotechnol, 2010. **28**(8): p. 817-25.
25. de Graaf, C.A. and B. van Steensel, *Chromatin organization: form to function.* Curr Opin Genet Dev, 2013. **23**(2): p. 185-90.
26. Crick, F.H., *On protein synthesis.* Symp Soc Exp Biol, 1958. **12**: p. 138-63.
27. Mignone, F., et al., *Untranslated regions of mRNAs.* Genome Biol, 2002. **3**(3): p. REVIEWS0004.
28. Matoulkova, E., et al., *The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells.* RNA Biol, 2012. **9**(5): p. 563-76.
29. Haberle, V. and B. Lenhard, *Promoter architectures and developmental gene regulation.* Semin Cell Dev Biol, 2016.

30. Zaret, K.S. and J.S. Carroll, *Pioneer transcription factors: establishing competence for gene expression.* Genes Dev, 2011. **25**(21): p. 2227-41.
31. Spitz, F. and E.E. Furlong, *Transcription factors: from enhancer binding to developmental control.* Nat Rev Genet, 2012. **13**(9): p. 613-26.
32. Carter, R. and G. Drouin, *Structural differentiation of the three eukaryotic RNA polymerases.* Genomics, 2009. **94**(6): p. 388-96.
33. Roeder, R.G., *Transcriptional regulation and the role of diverse coactivators in animal cells.* FEBS Lett, 2005. **579**(4): p. 909-15.
34. Kwak, H. and J.T. Lis, *Control of transcriptional elongation.* Annu Rev Genet, 2013. **47**: p. 483-508.
35. Lee, Y. and D.C. Rio, *Mechanisms and Regulation of Alternative Pre-mRNA Splicing.* Annu Rev Biochem, 2015. **84**: p. 291-323.
36. Darnell, J.E., Jr., *Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture.* RNA, 2013. **19**(4): p. 443-60.
37. Nirenberg, M.W. and J.H. Matthaei, *The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides.* Proc Natl Acad Sci U S A, 1961. **47**: p. 1588-602.
38. Goldberg, A.L. and R.E. Wittes, *Genetic code: aspects of organization.* Science, 1966. **153**(3734): p. 420-4.
39. Leung, E.K., et al., *The mechanism of peptidyl transfer catalysis by the ribosome.* Annu Rev Biochem, 2011. **80**: p. 527-55.
40. Waddington, C.H., *The epigenotype. 1942.* Int J Epidemiol, 2012. **41**(1): p. 10-3.
41. Deans, C. and K.A. Maggert, *What do you mean, "epigenetic"?* Genetics, 2015. **199**(4): p. 887-96.
42. Waddington, C.H., *The Basic Ideas of Biology.* Biological Theory, 2008. **3**(3): p. 238-253.
43. Nanney, D.L., *Epigenetic Control Systems.* Proc Natl Acad Sci U S A, 1958. **44**(7): p. 712-7.
44. Holliday, R., *Epigenetics: an overview.* Dev Genet, 1994. **15**(6): p. 453-7.
45. Wu, C. and J.R. Morris, *Genes, genetics, and epigenetics: a correspondence.* Science, 2001. **293**(5532): p. 1103-5.
46. Berger, S.L., et al., *An operational definition of epigenetics.* Genes Dev, 2009. **23**(7): p. 781-3.
47. Kasowski, M., et al., *Extensive variation in chromatin states across humans.* Science, 2013. **342**(6159): p. 750-2.
48. Kilpinen, H., et al., *Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription.* Science, 2013. **342**(6159): p. 744-7.
49. McVicker, G., et al., *Identification of genetic variants that affect histone modifications in human cells.* Science, 2013. **342**(6159): p. 747-9.
50. Heinz, S., et al., *Effect of natural genetic variation on enhancer selection and function.* Nature, 2013. **503**(7477): p. 487-92.
51. Furey, T.S. and P. Sethupathy, *Genetics. Genetics driving epigenetics.* Science, 2013. **342**(6159): p. 705-6.
52. Stower, H., *Gene regulation: from genetic variation to phenotype via chromatin.* Nat Rev Genet, 2013. **14**(12): p. 824.
53. Brazel, A.J. and D. Vernimmen, *The complexity of epigenetic diseases.* J Pathol, 2015.
54. Mirabella, A.C., B.M. Foster, and T. Bartke, *Chromatin deregulation in disease.* Chromosoma, 2016. **125**(1): p. 75-93.
55. Zhou, V.W., A. Goren, and B.E. Bernstein, *Charting histone modifications and the functional organization of mammalian genomes.* Nat Rev Genet, 2011. **12**(1): p. 7-18.
56. Hotchkiss, R.D., *The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography.* J Biol Chem, 1948. **175**(1): p. 315-32.
57. Breiling, A. and F. Lyko, *Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond.* Epigenetics Chromatin, 2015. **8**: p. 24.
58. Doskocil, J. and F. Sorm, *Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids.* Biochim Biophys Acta, 1962. **55**: p. 953-9.
59. Holliday, R. and J.E. Pugh, *DNA modification mechanisms and gene activity during development.* Science, 1975. **187**(4173): p. 226-32.
60. Riggs, A.D., *X inactivation, differentiation, and DNA methylation.* Cytogenet Cell Genet, 1975. **14**(1): p. 9-25.
61. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond.* Nat Rev Genet, 2012. **13**(7): p. 484-92.
62. Goll, M.G. and T.H. Bestor, *Eukaryotic cytosine methyltransferases.* Annu Rev Biochem, 2005. **74**: p. 481-514.
63. Jurkowska, R.Z., T.P. Jurkowski, and A. Jeltsch, *Structure and function of mammalian DNA methyltransferases.* Chembiochem, 2011. **12**(2): p. 206-22.
64. Bird, A.P., *CpG-rich islands and the function of DNA methylation.* Nature, 1986. **321**(6067): p. 209-13.

65. Smith, Z.D. and A. Meissner, *DNA methylation: roles in mammalian development.* Nat Rev Genet, 2013. **14**(3): p. 204-20.
66. Shukla, S., et al., *CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing.* Nature, 2011. **479**(7371): p. 74-9.
67. Spruijt, C.G. and M. Vermeulen, *DNA methylation: old dog, new tricks?* Nat Struct Mol Biol, 2014. **21**(11): p. 949-54.
68. Klose, R.J. and A.P. Bird, *Genomic DNA methylation: the mark and its mediators.* Trends Biochem Sci, 2006. **31**(2): p. 89-97.
69. Baubec, T. and D. Schubeler, *Genomic patterns and context specific interpretation of DNA methylation.* Curr Opin Genet Dev, 2014. **25**: p. 85-92.
70. Franchini, D.M., K.M. Schmitz, and S.K. Petersen-Mahrt, *5-Methylcytosine DNA demethylation: more than losing a methyl group.* Annu Rev Genet, 2012. **46**: p. 419-41.
71. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.
72. Li, E., T.H. Bestor, and R. Jaenisch, *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.* Cell, 1992. **69**(6): p. 915-26.
73. Li, E., et al., *DNA methylation, genomic imprinting, and mammalian development.* Cold Spring Harb Symp Quant Biol, 1993. **58**: p. 297-305.
74. Li, E., C. Beard, and R. Jaenisch, *Role for DNA methylation in genomic imprinting.* Nature, 1993. **366**(6453): p. 362-5.
75. Panning, B. and R. Jaenisch, *DNA hypomethylation can activate Xist expression and silence X-linked genes.* Genes Dev, 1996. **10**(16): p. 1991-2002.
76. Walsh, C.P., J.R. Chaillet, and T.H. Bestor, *Transcription of IAP endogenous retroviruses is constrained by cytosine methylation.* Nat Genet, 1998. **20**(2): p. 116-7.
77. Tahiliani, M., et al., *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.* Science, 2009. **324**(5929): p. 930-5.
78. Spruijt, C.G., et al., *Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives.* Cell, 2013. **152**(5): p. 1146-59.
79. Iurlaro, M., et al., *A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation.* Genome Biol, 2013. **14**(10): p. R119.
80. Wu, H. and Y. Zhang, *Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation.* Genes Dev, 2011. **25**(23): p. 2436-52.
81. Pastor, W.A., L. Aravind, and A. Rao, *TETonic shift: biological roles of TET proteins in DNA demethylation and transcription.* Nat Rev Mol Cell Biol, 2013. **14**(6): p. 341-56.
82. Mellen, M., et al., *MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system.* Cell, 2012. **151**(7): p. 1417-30.
83. Song, C.X., et al., *Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine.* Nat Biotechnol, 2011. **29**(1): p. 68-72.
84. Szulwach, K.E., et al., *5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging.* Nat Neurosci, 2011. **14**(12): p. 1607-16.
85. Hahn, M.A., et al., *Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis.* Cell Rep, 2013. **3**(2): p. 291-300.
86. Kriaucionis, S. and N. Heintz, *The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain.* Science, 2009. **324**(5929): p. 929-30.
87. Globisch, D., et al., *Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates.* PLoS One, 2010. **5**(12): p. e15367.
88. He, Y.F., et al., *Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA.* Science, 2011. **333**(6047): p. 1303-7.
89. Ito, S., et al., *Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.* Science, 2011. **333**(6047): p. 1300-3.
90. Ito, S. and I. Kuraoka, *Epigenetic modifications in DNA could mimic oxidative DNA damage: A double-edged sword.* DNA Repair (Amst), 2015. **32**: p. 52-7.
91. Marzluff, W.F., et al., *The human and mouse replication-dependent histone genes.* Genomics, 2002. **80**(5): p. 487-98.
92. Marzluff, W.F., E.J. Wagner, and R.J. Duronio, *Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail.* Nat Rev Genet, 2008. **9**(11): p. 843-54.

93. Burgess, R.J. and Z. Zhang, *Histone chaperones in nucleosome assembly and human disease.* Nat Struct Mol Biol, 2013. **20**(1): p. 14-22.

94. Talbert, P.B. and S. Henikoff, *Histone variants--ancient wrap artists of the epigenome.* Nat Rev Mol Cell Biol, 2010. **11**(4): p. 264-75.

95. Campos, E.I. and D. Reinberg, *Histones: annotating chromatin.* Annu Rev Genet, 2009. **43**: p. 559-99.

96. Talbert, P.B. and S. Henikoff, *Environmental responses mediated by histone variants.* Trends Cell Biol, 2014. **24**(11): p. 642-50.

97. Talbert, P.B., et al., *A unified phylogeny-based nomenclature for histone variants.* Epigenetics Chromatin, 2012. **5**: p. 7.

98. Santoro, S.W. and C. Dulac, *Histone variants and cellular plasticity.* Trends Genet, 2015. **31**(9): p. 516-27.

99. Zlatanova, J. and A. Thakar, *H2A.Z: view from the top.* Structure, 2008. **16**(2): p. 166-79.

100. Hu, G., et al., *H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation.* Cell Stem Cell, 2013. **12**(2): p. 180-92.

101. Hardy, S., et al., *The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition.* PLoS Genet, 2009. **5**(10): p. e1000687.

102. Weber, C.M., S. Ramachandran, and S. Henikoff, *Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase.* Molecular Cell, 2014. **53**(5): p. 819-30.

103. Fan, J.Y., et al., *H2A.Z alters the nucleosome surface to promote HP1alpha-mediated chromatin fiber folding.* Molecular Cell, 2004. **16**(4): p. 655-61.

104. Nagata, T., et al., *Polyadenylated and 3' processed mRNAs are transcribed from the mouse histone H2A.X gene.* Nucleic Acids Res, 1991. **19**(9): p. 2441-7.

105. Rogakou, E.P., et al., *DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139.* J Biol Chem, 1998. **273**(10): p. 5858-68.

106. Fernandez-Capetillo, O., et al., *H2AX: the histone guardian of the genome.* DNA Repair (Amst), 2004. **3**(8-9): p. 959-67.

107. Turinetto, V. and C. Giachino, *Multiple facets of histone variant H2AX: a DNA double-strand-break marker with several biological functions.* Nucleic Acids Res, 2015. **43**(5): p. 2489-98.

108. Scully, R. and A. Xie, *Double strand break repair functions of histone H2AX.* Mutat Res, 2013. **750**(1-2): p. 5-14.

109. Redon, C., et al., *Histone H2A variants H2AX and H2AZ.* Curr Opin Genet Dev, 2002. **12**(2): p. 162-9.

110. Chakravarthy, S., et al., *Structural characterization of the histone variant macroH2A.* Mol Cell Biol, 2005. **25**(17): p. 7616-24.

111. Gamble, M.J. and W.L. Kraus, *Multiple facets of the unique histone variant macroH2A: from genomics to cell biology.* Cell Cycle, 2010. **9**(13): p. 2568-74.

112. Costanzi, C. and J.R. Pehrson, *Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals.* Nature, 1998. **393**(6685): p. 599-601.

113. Hake, S.B., et al., *Serine 31 phosphorylation of histone variant H3.3 is specific to regions bordering centromeres in metaphase chromosomes.* Proc Natl Acad Sci U S A, 2005. **102**(18): p. 6344-9.

114. Elsasser, S.J., et al., *DAXX envelops a histone H3.3-H4 dimer for H3.3-specific recognition.* Nature, 2012. **491**(7425): p. 560-5.

115. Wen, H., et al., *ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression.* Nature, 2014. **508**(7495): p. 263-8.

116. Jin, C., et al., *H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions.* Nat Genet, 2009. **41**(8): p. 941-5.

117. Yukawa, M., et al., *Genome-wide analysis of the chromatin composition of histone H2A and H3 variants in mouse embryonic stem cells.* PLoS One, 2014. **9**(3): p. e92689.

118. Santaguida, S. and A. Musacchio, *The life and miracles of kinetochores.* EMBO J, 2009. **28**(17): p. 2511-31.

119. Furuyama, T. and S. Henikoff, *Centromeric nucleosomes induce positive DNA supercoils.* Cell, 2009. **138**(1): p. 104-13.

120. Weber, C.M. and S. Henikoff, *Histone variants: dynamic punctuation in transcription.* Genes Dev, 2014. **28**(7): p. 672-82.

121. Huang, H., et al., *SnapShot: histone modifications.* Cell, 2014. **159**(2): p. 458-458 e1.

122. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.

123. Rice, J.C. and C.D. Allis, *Histone methylation versus histone acetylation: new insights into epigenetic regulation.* Curr Opin Cell Biol, 2001. **13**(3): p. 263-73.

124. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**(6765): p. 41-5.

125. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.
126. Zhang, G. and S. Pradhan, *Mammalian epigenetic mechanisms.* IUBMB Life, 2014. **66**(4): p. 240-56.
127. Chittka, A., et al., *Transcription factor positive regulatory domain 4 (PRDM4) recruits protein arginine methyltransferase 5 (PRMT5) to mediate histone arginine methylation and control neural stem cell proliferation and differentiation.* J Biol Chem, 2012. **287**(51): p. 42995-3006.
128. Lhoumaud, P., et al., *Insulators recruit histone methyltransferase dMes4 to regulate chromatin of flanking genes.* EMBO J, 2014. **33**(14): p. 1599-613.
129. Duong, H.A. and C.J. Weitz, *Temporal orchestration of repressive chromatin modifiers by circadian clock Period complexes.* Nat Struct Mol Biol, 2014. **21**(2): p. 126-32.
130. Brownell, J.E., et al., *Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation.* Cell, 1996. **84**(6): p. 843-51.
131. Ng, H.H., et al., *Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity.* Molecular Cell, 2003. **11**(3): p. 709-19.
132. Krogan, N.J., et al., *The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation.* Molecular Cell, 2003. **11**(3): p. 721-9.
133. Lo, W.S., et al., *Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14.* Molecular Cell, 2000. **5**(6): p. 917-26.
134. Cheung, P., et al., *Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation.* Molecular Cell, 2000. **5**(6): p. 905-15.
135. Shilatifard, A., *Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression.* Annu Rev Biochem, 2006. **75**: p. 243-69.
136. Zippo, A., et al., *Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation.* Cell, 2009. **138**(6): p. 1122-36.
137. Sanchez-Elsner, T., et al., *Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax.* Science, 2006. **311**(5764): p. 1118-23.
138. Beckedorff, F.C., et al., *The intronic long noncoding RNA ANRASSF1 recruits PRC2 to the RASSF1A promoter, reducing the expression of RASSF1A and increasing cell proliferation.* PLoS Genet, 2013. **9**(8): p. e1003705.
139. Davidovich, C. and T.R. Cech, *The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2.* RNA, 2015. **21**(12): p. 2007-22.
140. Taverna, S.D., et al., *How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers.* Nat Struct Mol Biol, 2007. **14**(11): p. 1025-40.
141. Yun, M., et al., *Readers of histone modifications.* Cell Res, 2011. **21**(4): p. 564-78.
142. Patel, D.J. and Z. Wang, *Readout of epigenetic modifications.* Annu Rev Biochem, 2013. **82**: p. 81-118.
143. Smith, B.C. and J.M. Denu, *Chemical mechanisms of histone lysine and arginine modifications.* Biochim Biophys Acta, 2009. **1789**(1): p. 45-57.
144. Bedford, M.T. and S. Richard, *Arginine methylation an emerging regulator of protein function.* Molecular Cell, 2005. **18**(3): p. 263-72.
145. Rotili, D. and A. Mai, *Targeting Histone Demethylases: A New Avenue for the Fight against Cancer.* Genes Cancer, 2011. **2**(6): p. 663-79.
146. Labbe, R.M., A. Holowatyj, and Z.Q. Yang, *Histone lysine demethylase (KDM) subfamily 4: structures, functions and therapeutic potential.* Am J Transl Res, 2013. **6**(1): p. 1-15.
147. Bannister, A.J., et al., *Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes.* J Biol Chem, 2005. **280**(18): p. 17732-6.
148. Mikkelsen, T.S., et al., *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. **448**(7153): p. 553-60.
149. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-37.
150. Cao, R., et al., *Role of histone H3 lysine 27 methylation in Polycomb-group silencing.* Science, 2002. **298**(5595): p. 1039-43.
151. Kuzmichev, A., et al., *Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein.* Genes Dev, 2002. **16**(22): p. 2893-905.
152. Schuettengruber, B., et al., *Genome regulation by polycomb and trithorax proteins.* Cell, 2007. **128**(4): p. 735-45.
153. Schotta, G., et al., *A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin.* Genes Dev, 2004. **18**(11): p. 1251-62.
154. Muramatsu, D., et al., *Pericentric heterochromatin generated by HP1 protein interaction-defective histone methyltransferase Suv39h1.* J Biol Chem, 2013. **288**(35): p. 25285-96.

155. Jacobs, S.A. and S. Khorasanizadeh, *Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail.* Science, 2002. **295**(5562): p. 2080-3.

156. Struhl, K., *Histone acetylation and transcriptional regulatory mechanisms.* Genes Dev, 1998. **12**(5): p. 599-606.

157. Grunstein, M., *Histone acetylation in chromatin structure and transcription.* Nature, 1997. **389**(6649): p. 349-52.

158. Seto, E. and M. Yoshida, *Erasers of histone acetylation: the histone deacetylase enzymes.* Cold Spring Harb Perspect Biol, 2014. **6**(4): p. a018713.

159. Yang, X.J. and E. Seto, *The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men.* Nat Rev Mol Cell Biol, 2008. **9**(3): p. 206-18.

160. Wang, Z., et al., *Combinatorial patterns of histone acetylations and methylations in the human genome.* Nat Genet, 2008. **40**(7): p. 897-903.

161. Pasini, D., et al., *Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes.* Nucleic Acids Res, 2010. **38**(15): p. 4958-69.

162. Holmqvist, P.H. and M. Mannervik, *Genomic occupancy of the transcriptional co-activators p300 and CBP.* Transcription, 2013. **4**(1): p. 18-23.

163. Siddiqui, M.S., et al., *Persistent gammaH2AX: A promising molecular marker of DNA damage and aging.* Mutat Res Rev Mutat Res, 2015. **766**: p. 1-19.

164. Burma, S., et al., *ATM phosphorylates histone H2AX in response to DNA double-strand breaks.* J Biol Chem, 2001. **276**(45): p. 42462-7.

165. Yuan, J., R. Adamski, and J. Chen, *Focus on histone variant H2AX: to be or not to be.* FEBS Lett, 2010. **584**(17): p. 3717-24.

166. Badeaux, A.I. and Y. Shi, *Emerging roles for chromatin as a signal integration and storage platform.* Nat Rev Mol Cell Biol, 2013. **14**(4): p. 211-24.

167. Zentner, G.E. and S. Henikoff, *Regulation of nucleosome dynamics by histone modifications.* Nat Struct Mol Biol, 2013. **20**(3): p. 259-66.

168. Yuan, G.C., et al., *Genome-scale identification of nucleosome positions in S. cerevisiae.* Science, 2005. **309**(5734): p. 626-30.

169. Radman-Livaja, M. and O.J. Rando, *Nucleosome positioning: how is it established, and why does it matter?* Dev Biol, 2010. **339**(2): p. 258-66.

170. Hayes, J.J. and A.P. Wolffe, *The interaction of transcription factors with nucleosomal DNA.* Bioessays, 1992. **14**(9): p. 597-603.

171. Lieb, J.D. and N.D. Clarke, *Control of transcription through intragenic patterns of nucleosome composition.* Cell, 2005. **123**(7): p. 1187-90.

172. Rando, O.J. and K. Ahmad, *Rules and regulation in the primary structure of chromatin.* Curr Opin Cell Biol, 2007. **19**(3): p. 250-6.

173. Kassabov, S.R., et al., *SWI/SNF unwraps, slides, and rewraps the nucleosome.* Molecular Cell, 2003. **11**(2): p. 391-403.

174. Gangaraju, V.K. and B. Bartholomew, *Mechanisms of ATP dependent chromatin remodeling.* Mutat Res, 2007. **618**(1-2): p. 3-17.

175. Ramachandran, S. and S. Henikoff, *Nucleosome dynamics during chromatin remodeling in vivo.* Nucleus, 2016: p. 1-7.

176. Clapier, C.R. and B.R. Cairns, *The biology of chromatin remodeling complexes.* Annu Rev Biochem, 2009. **78**: p. 273-304.

177. Hargreaves, D.C. and G.R. Crabtree, *ATP-dependent chromatin remodeling: genetics, genomics and mechanisms.* Cell Res, 2011. **21**(3): p. 396-420.

178. Hota, S.K. and B. Bartholomew, *Diversity of operation in ATP-dependent chromatin remodelers.* Biochim Biophys Acta, 2011. **1809**(9): p. 476-87.

179. Segura-Totten, M. and K.L. Wilson, *BAF: roles in chromatin, nuclear structure and retrovirus integration.* Trends Cell Biol, 2004. **14**(5): p. 261-6.

180. Wu, J.I., *Diverse functions of ATP-dependent chromatin remodeling complexes in development and cancer.* Acta Biochim Biophys Sin (Shanghai), 2012. **44**(1): p. 54-69.

181. Stanley, F.K., S. Moore, and A.A. Goodarzi, *CHD chromatin remodelling enzymes and the DNA damage response.* Mutat Res, 2013. **750**(1-2): p. 31-44.

182. Murawska, M. and A. Brehm, *CHD chromatin remodelers and the transcription cycle.* Transcription, 2011. **2**(6): p. 244-53.

183. Marfella, C.G. and A.N. Imbalzano, *The Chd family of chromatin remodelers.* Mutat Res, 2007. **618**(1-2): p. 30-40.

184. Erdel, F. and K. Rippe, *Chromatin remodelling in mammalian cells by ISWI-type complexes--where, when and why?* FEBS J, 2011. **278**(19): p. 3608-18.

185. Gerhold, C.B., M.H. Hauer, and S.M. Gasser, *INO80-C and SWR-C: guardians of the genome.* J Mol Biol, 2015. **427**(3): p. 637-51.

186. Morrison, A.J. and X. Shen, *Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes.* Nat Rev Mol Cell Biol, 2009. **10**(6): p. 373-84.

187. Brownlee, P.M., C. Meisenberg, and J.A. Downs, *The SWI/SNF chromatin remodelling complex: Its role in maintaining genome stability and preventing tumourigenesis.* DNA Repair (Amst), 2015. **32**: p. 127-33.

188. Yodh, J., *ATP-Dependent Chromatin Remodeling.* Adv Exp Med Biol, 2013. **767**: p. 263-95.

189. Becker, P.B. and J.L. Workman, *Nucleosome remodeling and epigenetics.* Cold Spring Harb Perspect Biol, 2013. **5**(9).

190. Toto, M., G. D'Angelo, and D.F. Corona, *Regulation of ISWI chromatin remodelling activity.* Chromosoma, 2014. **123**(1-2): p. 91-102.

191. Li, M., et al., *Dynamic regulation of transcription factors by nucleosome remodeling.* Elife, 2015. **4**.

192. Mattick, J.S., *Non-coding RNAs: the architects of eukaryotic complexity.* EMBO Rep, 2001. **2**(11): p. 986-91.

193. Wang, K.C. and H.Y. Chang, *Molecular mechanisms of long noncoding RNAs.* Molecular Cell, 2011. **43**(6): p. 904-14.

194. Nagano, T. and P. Fraser, *No-nonsense functions for long noncoding RNAs.* Cell, 2011. **145**(2): p. 178-81.

195. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

196. Roy, A.L. and D.S. Singer, *Core promoters in transcription: old problem, new insights.* Trends Biochem Sci, 2015. **40**(3): p. 165-71.

197. Chen, J., et al., *Over 20% of human transcripts might form sense-antisense pairs.* Nucleic Acids Res, 2004. **32**(16): p. 4812-20.

198. Li, K. and R. Ramchandran, *Natural antisense transcript: a concomitant engagement with protein-coding transcript.* Oncotarget, 2010. **1**(6): p. 447-52.

199. Schmitz, S.U., P. Grote, and B.G. Herrmann, *Mechanisms of long noncoding RNA function in development and disease.* Cell Mol Life Sci, 2016.

200. Taft, R.J., M. Pheasant, and J.S. Mattick, *The relationship between non-protein-coding DNA and eukaryotic complexity.* Bioessays, 2007. **29**(3): p. 288-99.

201. Ravasi, T., et al., *Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome.* Genome Res, 2006. **16**(1): p. 11-9.

202. Cabili, M.N., et al., *Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses.* Genes Dev, 2011. **25**(18): p. 1915-27.

203. Derrien, T., et al., *The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression.* Genome Res, 2012. **22**(9): p. 1775-89.

204. Huang, Y., et al., *Molecular functions of small regulatory noncoding RNA.* Biochemistry (Mosc), 2013. **78**(3): p. 221-30.

205. Hangauer, M.J., I.W. Vaughn, and M.T. McManus, *Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs.* PLoS Genet, 2013. **9**(6): p. e1003569.

206. Taft, R.J., et al., *Non-coding RNAs: regulators of disease.* J Pathol, 2010. **220**(2): p. 126-39.

207. Wapinski, O. and H.Y. Chang, *Long noncoding RNAs and human disease*. Trends Cell Biol, 2011. **21**(6): p. 354-61.

208. Esteller, M., *Non-coding RNAs in human disease.* Nat Rev Genet, 2011. **12**(12): p. 861-74.

209. Elia, L. and G. Condorelli, *RNA (Epi)genetics in cardiovascular diseases.* J Mol Cell Cardiol, 2015. **89**(Pt A): p. 11-6.

210. Kapranov, P., et al., *RNA maps reveal new RNA classes and a possible function for pervasive transcription.* Science, 2007. **316**(5830): p. 1484-8.

211. Ma, L., V.B. Bajic, and Z. Zhang, *On the classification of long non-coding RNAs.* RNA Biol, 2013. **10**(6): p. 925-33.

212. Cao, J., *The functional role of long non-coding RNAs and epigenetics.* Biol Proced Online, 2014. **16**: p. 11.

213. Macfarlane, L.A. and P.R. Murphy, *MicroRNA: Biogenesis, Function and Role in Cancer.* Curr Genomics, 2010. **11**(7): p. 537-61.

214. Valinezhad Orang, A., R. Safaralizadeh, and M. Kazemzadeh-Bavili, *Mechanisms of miRNA-Mediated Gene Regulation from Common Downregulation to mRNA-Specific Upregulation.* Int J Genomics, 2014. **2014**: p. 970607.

215. Autuoro, J.M., S.P. Pirnie, and G.G. Carmichael, *Long noncoding RNAs in imprinting and X chromosome inactivation.* Biomolecules, 2014. **4**(1): p. 76-100.

216. Gendrel, A.V. and E. Heard, *Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation.* Annu Rev Cell Dev Biol, 2014. **30**: p. 561-80.

217. Kugel, J.F. and J.A. Goodrich, *Non-coding RNAs: key regulators of mammalian transcription.* Trends Biochem Sci, 2012. **37**(4): p. 144-51.

218. Vance, K.W. and C.P. Ponting, *Transcriptional regulatory functions of nuclear long noncoding RNAs.* Trends Genet, 2014. **30**(8): p. 348-55.

219. Khalil, A.M., et al., *Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression.* Proc Natl Acad Sci U S A, 2009. **106**(28): p. 11667-72.

220. Yap, K.L., et al., *Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a.* Molecular Cell, 2010. **38**(5): p. 662-74.

221. Rinn, J.L., et al., *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.* Cell, 2007. **129**(7): p. 1311-23.

222. Chu, C., et al., *Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions.* Molecular Cell, 2011. **44**(4): p. 667-78.

223. Wang, K.C., et al., *A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.* Nature, 2011. **472**(7341): p. 120-4.

224. Schmitz, K.M., et al., *Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes.* Genes Dev, 2010. **24**(20): p. 2264-9.

225. Wang, X., et al., *Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription.* Nature, 2008. **454**(7200): p. 126-30.

226. Tsai, M.C., et al., *Long noncoding RNA as modular scaffold of histone modification complexes.* Science, 2010. **329**(5992): p. 689-93.

227. Yakovchuk, P., J.A. Goodrich, and J.F. Kugel, *B2 RNA represses TFIIH phosphorylation of RNA polymerase II.* Transcription, 2011. **2**(1): p. 45-9.

228. Kino, T., et al., *Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor.* Sci Signal, 2010. **3**(107): p. ra8.

229. Guo, X., et al., *Advances in long noncoding RNAs: identification, structure prediction and function annotation.* Brief Funct Genomics, 2016. **15**(1): p. 38-46.

230. Koziol, M.J. and J.L. Rinn, *RNA traffic control of chromatin complexes.* Curr Opin Genet Dev, 2010. **20**(2): p. 142-8.

231. Dogini, D.B., et al., *The new world of RNAs.* Genet Mol Biol, 2014. **37**(1 Suppl): p. 285-93.

232. Wang, X. and C. He, *Dynamic RNA modifications in posttranscriptional regulation.* Molecular Cell, 2014. **56**(1): p. 5-12.

233. Romero, I.G., I. Ruvinsky, and Y. Gilad, *Comparative studies of gene expression and the evolution of gene regulation.* Nat Rev Genet, 2012. **13**(7): p. 505-16.

234. Riethoven, J.J., *Regulatory regions in DNA: promoters, enhancers, silencers, and insulators.* Methods Mol Biol, 2010. **674**: p. 33-42.

235. Banerji, J., S. Rusconi, and W. Schaffner, *Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences.* Cell, 1981. **27**(2 Pt 1): p. 299-308.

236. Akhtar, W., et al., *Chromatin position effects assayed by thousands of reporters integrated in parallel.* Cell, 2013. **154**(4): p. 914-27.

237. Stergachis, A.B., et al., *Developmental fate and cellular maturity encoded in human regulatory DNA landscapes.* Cell, 2013. **154**(4): p. 888-903.

238. Zhang, Y., et al., *Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations.* Nature, 2013. **504**(7479): p. 306-10.

239. Thakurela, S., et al., *Dynamics and function of distal regulatory elements during neurogenesis and neuroplasticity.* Genome Res, 2015. **25**(9): p. 1309-24.

240. Kieffer-Kwon, K.R., et al., *Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation.* Cell, 2013. **155**(7): p. 1507-20.

241. Ghisletti, S., et al., *Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages.* Immunity, 2010. **32**(3): p. 317-28.

242. Nord, A.S., et al., *Rapid and pervasive changes in genome-wide enhancer usage during mammalian development.* Cell, 2013. **155**(7): p. 1521-31.

243. Visel, A., et al., *A high-resolution enhancer atlas of the developing telencephalon.* Cell, 2013. **152**(4): p. 895-908.

244. Rada-Iglesias, A., et al., *A unique chromatin signature uncovers early developmental enhancers in humans.* Nature, 2011. **470**(7333): p. 279-83.

245. Joo, J.Y., et al., *Stimulus-specific combinatorial functionality of neuronal c-fos enhancers.* Nat Neurosci, 2016. **19**(1): p. 75-83.

246. Heinz, S., et al., *The selection and function of cell type-specific enhancers.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 144-54.

247. Shlyueva, D., G. Stampfel, and A. Stark, *Transcriptional enhancers: from properties to genome-wide predictions.* Nat Rev Genet, 2014. **15**(4): p. 272-86.

248. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Molecular Cell, 2010. **38**(4): p. 576-89.

249. Gualdi, R., et al., *Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control.* Genes Dev, 1996. **10**(13): p. 1670-82.

250. Lam, M.T., et al., *Enhancer RNAs and regulated transcriptional programs.* Trends Biochem Sci, 2014. **39**(4): p. 170-82.

251. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.

252. Kim, T.K., et al., *Widespread transcription at neuronal activity-regulated enhancers.* Nature, 2010. **465**(7295): p. 182-7.

253. Natoli, G. and J.C. Andrau, *Noncoding transcription at enhancers: general principles and functional models.* Annu Rev Genet, 2012. **46**: p. 1-19.

254. Li, W., M.T. Lam, and D. Notani, *Enhancer RNAs.* Cell Cycle, 2014. **13**(20): p. 3151-2.

255. Li, W., et al., *Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation.* Nature, 2013. **498**(7455): p. 516-20.

256. Hsieh, C.L., et al., *Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation.* Proc Natl Acad Sci U S A, 2014. **111**(20): p. 7319-24.

257. Yang, Y., et al., *Enhancer RNA-driven looping enhances the transcription of the long noncoding RNA DHRS4-AS1, a controller of the DHRS4 gene cluster.* Sci Rep, 2016. **6**: p. 20961.

258. Plank, J.L. and A. Dean, *Enhancer function: mechanistic and genome-wide insights come together.* Molecular Cell, 2014. **55**(1): p. 5-14.

259. Mousavi, K., et al., *eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci.* Molecular Cell, 2013. **51**(5): p. 606-17.

260. Sigova, A.A., et al., *Transcription factor trapping by RNA in gene regulatory elements.* Science, 2015. **350**(6263): p. 978-81.

261. Mueller-Storm, H.P., J.M. Sogo, and W. Schaffner, *An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge.* Cell, 1989. **58**(4): p. 767-77.

262. Levine, M., C. Cattoglio, and R. Tjian, *Looping back to leap forward: transcription enters a new era.* Cell, 2014. **157**(1): p. 13-25.

263. Matharu, N. and N. Ahituv, *Minor Loops in Major Folds: Enhancer-Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease.* PLoS Genet, 2015. **11**(12): p. e1005640.

264. Kagey, M.H., et al., *Mediator and cohesin connect gene expression and chromatin architecture.* Nature, 2010. **467**(7314): p. 430-5.

265. Stadhouders, R., et al., *Transcription regulation by distal enhancers: who's in the loop?* Transcription, 2012. **3**(4): p. 181-6.

266. Ren, X., et al., *Direct interactions of OCA-B and TFII-I regulate immunoglobulin heavy-chain gene transcription by facilitating enhancer-promoter communication.* Molecular Cell, 2011. **42**(3): p. 342-55.

267. Sawado, T., et al., *The beta -globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation.* Genes Dev, 2003. **17**(8): p. 1009-18.

268. Tolhuis, B., et al., *Looping and interaction between hypersensitive sites in the active beta-globin locus.* Molecular Cell, 2002. **10**(6): p. 1453-65.

269. Sanyal, A., et al., *The long-range interaction landscape of gene promoters.* Nature, 2012. **489**(7414): p. 109-13.

270. Thurman, R.E., et al., *The accessible chromatin landscape of the human genome.* Nature, 2012. **489**(7414): p. 75-82.

271. Nagano, T., et al., *Single-cell Hi-C reveals cell-to-cell variability in chromosome structure.* Nature, 2013. **502**(7469): p. 59-64.

272. Maston, G.A., et al., *Characterization of enhancer function from genome-wide analyses.* Annu Rev Genomics Hum Genet, 2012. **13**: p. 29-57.

273. Li, Q., et al., *Locus control regions.* Blood, 2002. **100**(9): p. 3077-86.

274. Grosveld, F., et al., *Position-independent, high-level expression of the human beta-globin gene in transgenic mice.* Cell, 1987. **51**(6): p. 975-85.

275. Whyte, W.A., et al., *Master transcription factors and mediator establish super-enhancers at key cell identity genes.* Cell, 2013. **153**(2): p. 307-19.

276. Pott, S. and J.D. Lieb, *What are super-enhancers?* Nat Genet, 2015. **47**(1): p. 8-12.

277. Hnisz, D., et al., *Super-enhancers in the control of cell identity and disease.* Cell, 2013. **155**(4): p. 934-47.

278. Hnisz, D., et al., *Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers.* Molecular Cell, 2015. **58**(2): p. 362-70.

279. Brown, J.D., et al., *NF-kappaB directs dynamic super enhancer formation in inflammation and atherogenesis.* Molecular Cell, 2014. **56**(2): p. 219-31.

280. Hah, N., et al., *Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs.* Proc Natl Acad Sci U S A, 2015. **112**(3): p. E297-302.

281. Schmidt, S.F., et al., *Acute TNF-induced repression of cell identity genes is mediated by NFkappaB-directed redistribution of cofactors from super-enhancers.* Genome Res, 2015. **25**(9): p. 1281-94.

282. Cremer, T. and C. Cremer, *Chromosome territories, nuclear architecture and gene regulation in mammalian cells.* Nat Rev Genet, 2001. **2**(4): p. 292-301.

283. Cremer, T. and M. Cremer, *Chromosome territories.* Cold Spring Harb Perspect Biol, 2010. **2**(3): p. a003889.

284. Zorn, C., et al., *Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus. Distribution in interphase and metaphase.* Exp Cell Res, 1979. **124**(1): p. 111-9.

285. Cremer, T., et al., *Role of chromosome territories in the functional compartmentalization of the cell nucleus.* Cold Spring Harb Symp Quant Biol, 1993. **58**: p. 777-92.

286. Oakes, M.L., et al., *Expression of rRNA genes and nucleolus formation at ectopic chromosomal sites in the yeast Saccharomyces cerevisiae.* Mol Cell Biol, 2006. **26**(16): p. 6223-38.

287. Weierich, C., et al., *Three-dimensional arrangements of centromeres and telomeres in nuclei of human and murine lymphocytes.* Chromosome Res, 2003. **11**(5): p. 485-502.

288. Derenzini, M., M. Thiry, and G. Goessens, *Ultrastructural cytochemistry of the mammalian cell nucleolus.* J Histochem Cytochem, 1990. **38**(9): p. 1237-56.

289. Guelen, L., et al., *Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions.* Nature, 2008. **453**(7197): p. 948-51.

290. Meuleman, W., et al., *Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence.* Genome Res, 2013. **23**(2): p. 270-80.

291. van Koningsbruggen, S., et al., *High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli.* Mol Biol Cell, 2010. **21**(21): p. 3735-48.

292. Nemeth, A., et al., *Initial genomics of the human nucleolus.* PLoS Genet, 2010. **6**(3): p. e1000889.

293. Kosak, S.T., et al., *Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development.* Science, 2002. **296**(5565): p. 158-62.

294. Schubeler, D., et al., *Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus.* Genes Dev, 2000. **14**(8): p. 940-50.

295. Peric-Hupkes, D., et al., *Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation.* Molecular Cell, 2010. **38**(4): p. 603-13.

296. Rieder, D., Z. Trajanoski, and J.G. McNally, *Transcription factories.* Front Genet, 2012. **3**: p. 221.

297. Papantonis, A. and P.R. Cook, *Transcription factories: genome organization and gene regulation.* Chem Rev, 2013. **113**(11): p. 8683-705.

298. Buckley, M.S. and J.T. Lis, *Imaging RNA Polymerase II transcription sites in living cells.* Curr Opin Genet Dev, 2014. **25**: p. 126-30.

299. Cremer, T., et al., *The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments.* FEBS Lett, 2015. **589**(20 Pt A): p. 2931-43.

300. Lieberman-Aiden, E., et al., *Comprehensive mapping of long-range interactions reveals folding principles of the human genome.* Science, 2009. **326**(5950): p. 289-93.

301. Yaffe, E. and A. Tanay, *Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture.* Nat Genet, 2011. **43**(11): p. 1059-65.

302. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-80.

303. Rao, S.S., et al., *A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.* Cell, 2014. **159**(7): p. 1665-80.

304. Bouwman, B.A. and W. de Laat, *Getting the genome in shape: the formation of loops, domains and compartments.* Genome Biol, 2015. **16**: p. 154.

305. Ea, V., et al., *Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization.* Genes (Basel), 2015. **6**(3): p. 734-50.

306. Rastogi, R.P., et al., *Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair.* J Nucleic Acids, 2010. **2010**: p. 592980.

307. De Bont, R. and N. van Larebeke, *Endogenous DNA damage in humans: a review of quantitative data.* Mutagenesis, 2004. **19**(3): p. 169-85.

308. Helleday, T., S. Eshtad, and S. Nik-Zainal, *Mechanisms underlying mutational signatures in human cancers.* Nat Rev Genet, 2014. **15**(9): p. 585-98.

309. Latonen, L. and M. Laiho, *Cellular UV damage responses--functions of tumor suppressor p53.* Biochim Biophys Acta, 2005. **1755**(2): p. 71-89.

310. Jackson, S.P. and J. Bartek, *The DNA-damage response in human biology and disease.* Nature, 2009. **461**(7267): p. 1071-8.

311. Marteijn, J.A., et al., *Understanding nucleotide excision repair and its roles in cancer and ageing.* Nat Rev Mol Cell Biol, 2014. **15**(7): p. 465-81.

312. Scharer, O.D., *Nucleotide excision repair in eukaryotes.* Cold Spring Harb Perspect Biol, 2013. **5**(10): p. a012609.

313. Shah, P. and Y.Y. He, *Molecular regulation of UV-induced DNA repair.* Photochem Photobiol, 2015. **91**(2): p. 254-64.

314. Spivak, G., *Nucleotide excision repair in humans.* DNA Repair (Amst), 2015. **36**: p. 13-8.

315. Mullenders, L., *DNA damage mediated transcription arrest: Step back to go forward.* DNA Repair (Amst), 2015.

316. Bohr, V.A., et al., *DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall.* Cell, 1985. **40**(2): p. 359-69.

317. Mellon, I., G. Spivak, and P.C. Hanawalt, *Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene.* Cell, 1987. **51**(2): p. 241-9.

318. Fousteri, M. and L.H. Mullenders, *Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects.* Cell Res, 2008. **18**(1): p. 73-84.

319. Laine, J.P. and J.M. Egly, *When transcription and repair meet: a complex system.* Trends Genet, 2006. **22**(8): p. 430-6.

320. Tornaletti, S., et al., *Nucleotide sequence context effect of a cyclobutane pyrimidine dimer upon RNA polymerase II transcription.* J Biol Chem, 1997. **272**(50): p. 31719-24.

321. Ljungman, M., et al., *Inhibition of RNA polymerase II as a trigger for the p53 response.* Oncogene, 1999. **18**(3): p. 583-92.

322. Ljungman, M. and D.P. Lane, *Transcription - guarding the genome by sensing DNA damage.* Nat Rev Cancer, 2004. **4**(9): p. 727-37.

323. Chapman, J.R., M.R. Taylor, and S.J. Boulton, *Playing the end game: DNA double-strand break repair pathway choice.* Molecular Cell, 2012. **47**(4): p. 497-510.

324. Ceccaldi, R., B. Rondinelli, and A.D. D'Andrea, *Repair Pathway Choices and Consequences at the Double-Strand Break.* Trends Cell Biol, 2016. **26**(1): p. 52-64.

325. Amaro-Ortiz, A., B. Yan, and J.A. D'Orazio, *Ultraviolet radiation, aging and the skin: prevention of damage by topical cAMP manipulation.* Molecules, 2014. **19**(5): p. 6202-19.

326. D'Orazio, J., et al., *UV radiation and the skin.* Int J Mol Sci, 2013. **14**(6): p. 12222-48.

327. de Gruijl, F.R., H.J. van Kranen, and L.H. Mullenders, *UV-induced DNA damage, repair, mutations and oncogenic pathways in skin cancer.* J Photochem Photobiol B, 2001. **63**(1-3): p. 19-27.

328. Bradford, P.T., et al., *Cancer and neurologic degeneration in xeroderma pigmentosum: long term follow-up characterises the role of DNA repair.* J Med Genet, 2011. **48**(3): p. 168-76.

329. Zhou, B.B. and S.J. Elledge, *The DNA damage response: putting checkpoints in perspective.* Nature, 2000. **408**(6811): p. 433-9.

330. Pustisek, N. and M. Situm, *UV-radiation, apoptosis and skin.* Coll Antropol, 2011. **35 Suppl 2**: p. 339-41.

331. Shiloh, Y., *ATM and related protein kinases: safeguarding genome integrity.* Nat Rev Cancer, 2003. **3**(3): p. 155-68.

332. Mei Kwei, J.S., et al., *Blockage of RNA polymerase II at a cyclobutane pyrimidine dimer and 6-4 photoproduct.* Biochem Biophys Res Commun, 2004. **320**(4): p. 1133-8.

333. Brueckner, F., et al., *CPD damage recognition by transcribing RNA polymerase II.* Science, 2007. **315**(5813): p. 859-62.

334. Marietta, C. and P.J. Brooks, *Transcriptional bypass of bulky DNA lesions causes new mutant RNA transcripts in human cells.* EMBO Rep, 2007. **8**(4): p. 388-93.

335. Saxowsky, T.T. and P.W. Doetsch, *RNA polymerase encounters with DNA damage: transcription-coupled repair or transcriptional mutagenesis?* Chem Rev, 2006. **106**(2): p. 474-88.

336. Walmacq, C., et al., *Mechanism of translesion transcription by RNA polymerase II and its role in cellular resistance to DNA damage.* Molecular Cell, 2012. **46**(1): p. 18-29.

337. Mone, M.J., et al., *Local UV-induced DNA damage in cell nuclei results in local transcription inhibition.* EMBO Rep, 2001. **2**(11): p. 1013-7.

338. Vichi, P., et al., *Cisplatin- and UV-damaged DNA lure the basal transcription factor TFIID/TBP.* EMBO J, 1997. **16**(24): p. 7444-56.

339. Coin, F., et al., *TATA binding protein discriminates between different lesions on DNA, resulting in a transcription decrease.* Mol Cell Biol, 1998. **18**(7): p. 3907-14.

340. Fowler, T., R. Sen, and A.L. Roy, *Regulation of primary response genes.* Molecular Cell, 2011. **44**(3): p. 348-60.

341. Tullai, J.W., et al., *Immediate-early and delayed primary response genes are distinct in function and genomic architecture.* J Biol Chem, 2007. **282**(33): p. 23981-95.

342. Buscher, M., et al., *Activation of the c-fos gene by UV and phorbol ester: different signal transduction pathways converge to the same enhancer element.* Oncogene, 1988. **3**(3): p. 301-11.

343. Devary, Y., et al., *Rapid and preferential activation of the c-jun gene during the mammalian UV response.* Mol Cell Biol, 1991. **11**(5): p. 2804-11.

344. Schreiber, M., et al., *Fos is an essential component of the mammalian UV response.* EMBO J, 1995. **14**(21): p. 5338-49.

345. Devary, Y., et al., *The mammalian ultraviolet response is triggered by activation of Src tyrosine kinases.* Cell, 1992. **71**(7): p. 1081-91.

346. Devary, Y., et al., *NF-kappa B activation by ultraviolet light not dependent on a nuclear signal.* Science, 1993. **261**(5127): p. 1442-5.

347. Sachsenmaier, C., et al., *Involvement of growth factor receptors in the mammalian UVC response.* Cell, 1994. **78**(6): p. 963-72.

348. Radler-Pohl, A., et al., *UV-induced activation of AP-1 involves obligatory extranuclear steps including Raf-1 kinase.* EMBO J, 1993. **12**(3): p. 1005-12.

349. Derijard, B., et al., *Independent human MAP-kinase signal transduction pathways defined by MEK and MKK isoforms.* Science, 1995. **267**(5198): p. 682-5.

350. Raingeaud, J., et al., *Pro-inflammatory cytokines and environmental stress cause p38 mitogen-activated protein kinase activation by dual phosphorylation on tyrosine and threonine.* J Biol Chem, 1995. **270**(13): p. 7420-6.

351. Muthusamy, V. and T.J. Piva, *The UV response of the skin: a review of the MAPK, NFkappaB and TNFalpha signal transduction pathways.* Arch Dermatol Res, 2010. **302**(1): p. 5-17.

352. Derijard, B., et al., *JNK1: a protein kinase stimulated by UV light and Ha-Ras that binds and phosphorylates the c-Jun activation domain.* Cell, 1994. **76**(6): p. 1025-37.

353. Stein, B., et al., *UV-induced DNA damage is an intermediate step in UV-induced expression of human immunodeficiency virus type 1, collagenase, c-fos, and metallothionein.* Mol Cell Biol, 1989. **9**(11): p. 5169-81.

354. Miralles, F., et al., *UV irradiation induces the murine urokinase-type plasminogen activator gene via the c-Jun N-terminal kinase signaling pathway: requirement of an AP1 enhancer element.* Mol Cell Biol, 1998. **18**(8): p. 4537-47.

355. Hibi, M., et al., *Identification of an oncoprotein- and UV-responsive protein kinase that binds and potentiates the c-Jun activation domain.* Genes Dev, 1993. **7**(11): p. 2135-48.

356. Sluss, H.K., et al., *Signal transduction by tumor necrosis factor mediated by JNK protein kinases.* Mol Cell Biol, 1994. **14**(12): p. 8376-84.

357. Kallunki, T., et al., *JNK2 contains a specificity-determining region responsible for efficient c-Jun binding and phosphorylation.* Genes Dev, 1994. **8**(24): p. 2996-3007.

358. Minden, A., et al., *c-Jun N-terminal phosphorylation correlates with activation of the JNK subgroup but not the ERK subgroup of mitogen-activated protein kinases.* Mol Cell Biol, 1994. **14**(10): p. 6683-8.

359. Kyriakis, J.M., et al., *The stress-activated protein kinase subfamily of c-Jun kinases.* Nature, 1994. **369**(6476): p. 156-60.

360. Adachi, M., et al., *Specificity in stress response: epidermal keratinocytes exhibit specialized UV-responsive signal transduction pathways.* DNA Cell Biol, 2003. **22**(10): p. 665-77.

361. Tanos, T., et al., *Phosphorylation of c-Fos by members of the p38 MAPK family. Role in the AP-1 response to UV light.* J Biol Chem, 2005. **280**(19): p. 18842-52.

362. Manzoor, Z. and Y.S. Koh, *Mitogen-activated Protein Kinases in Inflammation* J Bacteriol Virol., 2012. **42**(3): p. 189-195.

363. Milne, D.M., et al., *p53 is phosphorylated in vitro and in vivo by an ultraviolet radiation-induced protein kinase characteristic of the c-Jun kinase, JNK1.* J Biol Chem, 1995. **270**(10): p. 5511-8.

364. She, Q.B., N. Chen, and Z. Dong, *ERKs and p38 kinase phosphorylate p53 protein at serine 15 in response to UV radiation.* J Biol Chem, 2000. **275**(27): p. 20444-9.

365. Tibbetts, R.S., et al., *A role for ATR in the DNA damage-induced phosphorylation of p53.* Genes Dev, 1999. **13**(2): p. 152-7.

366. Lakin, N.D. and S.P. Jackson, *Regulation of p53 in response to DNA damage.* Oncogene, 1999. **18**(53): p. 7644-55.

367. Hall, P.A., et al., *High levels of p53 protein in UV-irradiated normal human skin.* Oncogene, 1993. **8**(1): p. 203-7.

368. Healy, E., et al., *Dissociation of erythema and p53 protein expression in human skin following UVB irradiation, and induction of p53 protein and mRNA following application of skin irritants.* J Invest Dermatol, 1994. **103**(4): p. 493-9.

369. Levine, A.J., J. Momand, and C.A. Finlay, *The p53 tumour suppressor gene.* Nature, 1991. **351**(6326): p. 453-6.

370. Kuerbitz, S.J., et al., *Wild-type p53 is a cell cycle checkpoint determinant following irradiation.* Proc Natl Acad Sci U S A, 1992. **89**(16): p. 7491-5.

371. Sionov, R.V. and Y. Haupt, *The cellular response to p53: the decision between life and death.* Oncogene, 1999. **18**(45): p. 6145-57.

372. Kastan, M.B., et al., *Participation of p53 protein in the cellular response to DNA damage.* Cancer Res, 1991. **51**(23 Pt 1): p. 6304-11.

373. Meek, D.W., *Regulation of the p53 response and its relationship to cancer.* Biochem J, 2015. **469**(3): p. 325-46.

374. Al-Mohanna, M.A., et al., *p53 is dispensable for UV-induced cell cycle arrest at late G(1) in mammalian cells.* Carcinogenesis, 2001. **22**(4): p. 573-8.

375. Eliyahu, D., et al., *Wild-type p53 can inhibit oncogene-mediated focus formation.* Proc Natl Acad Sci U S A, 1989. **86**(22): p. 8763-7.

376. Yonish-Rouach, E., et al., *Wild-type p53 induces apoptosis of myeloid leukaemic cells that is inhibited by interleukin-6.* Nature, 1991. **352**(6333): p. 345-7.

377. Shaw, P., et al., *Induction of apoptosis by wild-type p53 in a human colon tumor-derived cell line.* Proc Natl Acad Sci U S A, 1992. **89**(10): p. 4495-9.

378. Roos, W.P., A.D. Thomas, and B. Kaina, *DNA damage and the balance between survival and death in cancer biology.* Nat Rev Cancer, 2016. **16**(1): p. 20-33.

379. Benjamin, C.L. and H.N. Ananthaswamy, *p53 and the pathogenesis of skin cancer.* Toxicol Appl Pharmacol, 2007. **224**(3): p. 241-8.

380. Smith, M.L. and A.J. Fornace, Jr., *p53-mediated protective responses to UV irradiation.* Proc Natl Acad Sci U S A, 1997. **94**(23): p. 12255-7.

381. Giglia-Mari, G. and A. Sarasin, *TP53 mutations in human skin cancers.* Hum Mutat, 2003. **21**(3): p. 217-28.

382. Huarte, M., et al., *A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response.* Cell, 2010. **142**(3): p. 409-19.

383. Liu, Y. and X. Lu, *Non-coding RNAs in DNA damage response.* Am J Cancer Res, 2012. **2**(6): p. 658-75.

384. Wan, G., et al., *A novel non-coding RNA lncRNA-JADE connects DNA damage signalling to histone H4 acetylation.* EMBO J, 2013. **32**(21): p. 2833-47.

385. Negishi, M., et al., *A new lncRNA, APTR, associates with and represses the CDKN1A/p21 promoter by recruiting polycomb proteins.* PLoS One, 2014. **9**(4): p. e95216.

386. Rashi-Elkeles, S., et al., *Parallel profiling of the transcriptome, cistrome, and epigenome in the cellular response to ionizing radiation.* Sci Signal, 2014. **7**(325): p. rs3.

387. Younger, S.T., et al., *Integrative genomic analysis reveals widespread enhancer regulation by p53 in response to DNA damage.* Nucleic Acids Res, 2015. **43**(9): p. 4447-62.

388. Francia, S., *Non-Coding RNA: Sequence-Specific Guide for Chromatin Modification and DNA Damage Signaling.* Front Genet, 2015. **6**: p. 320.

389. Zhang, C. and G. Peng, *Non-coding RNAs: an emerging player in DNA damage response.* Mutat Res Rev Mutat Res, 2015. **763**: p. 202-11.

390. Zeng, Q., et al., *Analysis of lncRNAs expression in UVB-induced stress responses of melanocytes.* J Dermatol Sci, 2016. **81**(1): p. 53-60.

391. Smerdon, M.J. and M.W. Lieberman, *Nucleosome rearrangement in human chromatin during UV-induced DNA- reapir synthesis.* Proc Natl Acad Sci U S A, 1978. **75**(9): p. 4238-41.

392. Soria, G., S.E. Polo, and G. Almouzni, *Prime, repair, restore: the active role of chromatin in the DNA damage response.* Molecular Cell, 2012. **46**(6): p. 722-34.

393. Polo, S.E. and G. Almouzni, *Chromatin dynamics after DNA damage: The legacy of the access-repair-restore model.* DNA Repair (Amst), 2015. **36**: p. 114-21.

394. Luijsterburg, M.S., et al., *DDB2 promotes chromatin decondensation at UV-induced DNA damage.* J Cell Biol, 2012. **197**(2): p. 267-81.

395. Lans, H., J.A. Marteijn, and W. Vermeulen, *ATP-dependent chromatin remodeling in the DNA-damage response.* Epigenetics Chromatin, 2012. **5**: p. 4.

396. Jiang, Y., et al., *INO80 chromatin remodeling complex promotes the removal of UV lesions by the nucleotide excision repair pathway.* Proc Natl Acad Sci U S A, 2010. **107**(40): p. 17274-9.

397. Zhao, Q., et al., *Modulation of nucleotide excision repair by mammalian SWI/SNF chromatin-remodeling complex.* J Biol Chem, 2009. **284**(44): p. 30424-32.

398. Zhang, L., et al., *The chromatin remodeling factor BRG1 stimulates nucleotide excision repair by facilitating recruitment of XPC to sites of DNA damage.* Cell Cycle, 2009. **8**(23): p. 3953-9.

399. Park, J.H., et al., *Mammalian SWI/SNF chromatin remodeling complexes are required to prevent apoptosis after DNA damage.* DNA Repair (Amst), 2009. **8**(1): p. 29-39.

400. Adam, S., J. Dabin, and S.E. Polo, *Chromatin plasticity in response to DNA damage: The shape of things to come.* DNA Repair (Amst), 2015. **32**: p. 120-6.

401. Adam, S., S.E. Polo, and G. Almouzni, *Transcription recovery after DNA damage requires chromatin priming by the H3.3 histone chaperone HIRA.* Cell, 2013. **155**(1): p. 94-106.

402. Frey, A., et al., *Histone H3.3 is required to maintain replication fork progression after UV damage.* Curr Biol, 2014. **24**(18): p. 2195-201.

403. Polo, S.E., D. Roche, and G. Almouzni, *New histone incorporation marks sites of UV repair in human cells.* Cell, 2006. **127**(3): p. 481-93.

404. Green, C.M. and G. Almouzni, *Local action of the chromatin assembly factor CAF-1 at sites of nucleotide excision repair in vivo.* EMBO J, 2003. **22**(19): p. 5163-74.

405. Ikura, T., et al., *DNA damage-dependent acetylation and ubiquitination of H2AX enhances chromatin dynamics.* Mol Cell Biol, 2007. **27**(20): p. 7028-40.

406. Nishibuchi, I., et al., *Reorganization of damaged chromatin by the exchange of histone variant H2A.Z-2.* Int J Radiat Oncol Biol Phys, 2014. **89**(4): p. 736-44.

407. Dinant, C., et al., *Enhanced chromatin dynamics by FACT promotes transcriptional restart after UV-induced DNA damage.* Molecular Cell, 2013. **51**(4): p. 469-79.

408. Dantuma, N.P. and H. van Attikum, *Spatiotemporal regulation of posttranslational modifications in the DNA damage response.* EMBO J, 2016. **35**(1): p. 6-23.

409. Podhorecka, M., A. Skladanowski, and P. Bozko, *H2AX Phosphorylation: Its Role in DNA Damage Response and Cancer Therapy.* J Nucleic Acids, 2010. **2010**.

410. Chapman, J.R. and S.P. Jackson, *Phospho-dependent interactions between NBS1 and MDC1 mediate chromatin retention of the MRN complex at sites of DNA damage.* EMBO Rep, 2008. **9**(8): p. 795-801.

411. Spycher, C., et al., *Constitutive phosphorylation of MDC1 physically links the MRE11-RAD50-NBS1 complex to damaged chromatin.* J Cell Biol, 2008. **181**(2): p. 227-40.

412. Goldberg, M., et al., *MDC1 is required for the intra-S-phase DNA damage checkpoint.* Nature, 2003. **421**(6926): p. 952-6.

413. Stewart, G.S., et al., *MDC1 is a mediator of the mammalian DNA damage checkpoint.* Nature, 2003. **421**(6926): p. 961-6.

414. Stucki, M., et al., *MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks.* Cell, 2005. **123**(7): p. 1213-26.

415. Iacovoni, J.S., et al., *High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome.* EMBO J, 2010. **29**(8): p. 1446-57.

416. Hanasoge, S. and M. Ljungman, *H2AX phosphorylation after UV irradiation is triggered by DNA repair intermediates and is mediated by the ATR kinase.* Carcinogenesis, 2007. **28**(11): p. 2298-304.

417. Marti, T.M., et al., *H2AX phosphorylation within the G1 phase after UV irradiation depends on nucleotide excision repair and not DNA double-strand breaks.* Proc Natl Acad Sci U S A, 2006. **103**(26): p. 9891-6.

418. Halicka, H.D., et al., *Histone H2AX phosphorylation after cell irradiation with UV-B: relationship to cell cycle phase and induction of apoptosis.* Cell Cycle, 2005. **4**(2): p. 339-45.

419. Oh, K.S., et al., *UV-induced histone H2AX phosphorylation and DNA damage related proteins accumulate and persist in nucleotide excision repair-deficient XP-B cells.* DNA Repair (Amst), 2011. **10**(1): p. 5-15.

420. Tjeertes, J.V., K.M. Miller, and S.P. Jackson, *Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells.* EMBO J, 2009. **28**(13): p. 1878-89.

421. Corpet, A. and G. Almouzni, *A histone code for the DNA damage response in mammalian cells?* EMBO J, 2009. **28**(13): p. 1828-30.

422. Sawicka, A., et al., *H3S28 phosphorylation is a hallmark of the transcriptional response to cellular stress.* Genome Res, 2014. **24**(11): p. 1808-20.

423. Dinant, C., A.B. Houtsmuller, and W. Vermeulen, *Chromatin structure and DNA damage repair.* Epigenetics Chromatin, 2008. **1**(1): p. 9.

424. Heo, K., et al., *FACT-mediated exchange of histone variant H2AX regulated by phosphorylation of H2AX and ADP-ribosylation of Spt16.* Molecular Cell, 2008. **30**(1): p. 86-97.

425. Lee, H.S., et al., *A cooperative activation loop among SWI/SNF, gamma-H2AX and H3 acetylation for DNA double-strand break repair.* EMBO J, 2010. **29**(8): p. 1434-45.

426. Luijsterburg, M.S. and H. van Attikum, *Chromatin and the DNA damage response: the cancer connection.* Mol Oncol, 2011. **5**(4): p. 349-67.

427. Mehta, I.S., et al., *Chromosome territories reposition during DNA damage-repair response.* Genome Biol, 2013. **14**(12): p. R135.

428. Dion, V. and S.M. Gasser, *Chromatin movement in the maintenance of genome stability.* Cell, 2013. **152**(6): p. 1355-64.

429. Johnson, C.H., T. Mori, and Y. Xu, *A cyanobacterial circadian clockwork.* Curr Biol, 2008. **18**(17): p. R816-R825.

430. Baker, C.L., J.J. Loros, and J.C. Dunlap, *The circadian clock of Neurospora crassa.* FEMS Microbiol Rev, 2012. **36**(1): p. 95-110.

431. Hsu, P.Y. and S.L. Harmer, *Wheels within wheels: the plant circadian system.* Trends Plant Sci, 2014. **19**(4): p. 240-9.

432. Yu, W. and P.E. Hardin, *Circadian oscillators of Drosophila and mammals.* J Cell Sci, 2006. **119**(Pt 23): p. 4793-5.

433. Kondratova, A.A., et al., *Circadian clock proteins control adaptation to novel environment and memory formation.* Aging (Albany NY), 2010. **2**(5): p. 285-97.

434. Nakahata, Y., et al., *Signaling to the circadian clock: plasticity by chromatin remodeling.* Curr Opin Cell Biol, 2007. **19**(2): p. 230-7.

435. Richards, J. and M.L. Gumz, *Mechanism of the circadian clock in physiology.* Am J Physiol Regul Integr Comp Physiol, 2013. **304**(12): p. R1053-64.

436. Feillet, C., et al., *Coupling between the Circadian Clock and Cell Cycle Oscillators: Implication for Healthy Cells and Malignant Growth.* Front Neurol, 2015. **6**: p. 96.

437. Sancar, A., et al., *Circadian clock control of the cellular response to DNA damage.* FEBS Lett, 2010. **584**(12): p. 2618-25.

438. Antoch, M.P. and R.V. Kondratov, *Circadian proteins and genotoxic stress response.* Circ Res, 2010. **106**(1): p. 68-78.

439. Uth, K. and R. Sleigh, *Deregulation of the circadian clock constitutes a significant factor in tumorigenesis: a clockwork cancer. Part II. studies.* Biotechnol Biotechnol Equip, 2014. **28**(3): p. 379-386.

440. Sancar, A., et al., *Circadian clock, cancer, and chemotherapy.* Biochemistry, 2015. **54**(2): p. 110-23.

441. Forni, D., et al., *Genetic adaptation of the human circadian clock to day-length latitudinal variations and relevance for affective disorders.* Genome Biol, 2014. **15**(10): p. 499.

442. Merbitz-Zahradnik, T. and E. Wolf, *How is the inner circadian clock controlled by interactive clock proteins?: Structural analysis of clock proteins elucidates their physiological role.* FEBS Lett, 2015. **589**(14): p. 1516-29.

443. Rosenwasser, A.M. and F.W. Turek, *Neurobiology of Circadian Rhythm Regulation.* Sleep Med Clin, 2015. **10**(4): p. 403-12.

444. Dibner, C., U. Schibler, and U. Albrecht, *The mammalian circadian timing system: organization and coordination of central and peripheral clocks.* Annu Rev Physiol, 2010. **72**: p. 517-49.

445. Morf, J. and U. Schibler, *Body temperature cycles: gatekeepers of circadian clocks.* Cell Cycle, 2013. **12**(4): p. 539-40.

446. Patton, D.F. and R.E. Mistlberger, *Circadian adaptations to meal timing: neuroendocrine mechanisms.* Front Neurosci, 2013. **7**: p. 185.

447. Asher, G. and P. Sassone-Corsi, *Time for food: the intimate interplay between nutrition, metabolism, and the circadian clock.* Cell, 2015. **161**(1): p. 84-92.

448. Ki, Y., et al., *Warming Up Your Tick-Tock: Temperature-Dependent Regulation of Circadian Clocks.* Neuroscientist, 2015. **21**(5): p. 503-18.

449. Hickie, I.B., et al., *Manipulating the sleep-wake cycle and circadian rhythms to improve clinical management of major depression.* BMC Med, 2013. **11**: p. 79.

450. Andreani, T.S., et al., *Genetics of Circadian Rhythms.* Sleep Med Clin, 2015. **10**(4): p. 413-21.

451. King, D.P., et al., *Positional cloning of the mouse circadian clock gene.* Cell, 1997. **89**(4): p. 641-53.

452. Gekakis, N., et al., *Role of the CLOCK protein in the mammalian circadian mechanism.* Science, 1998. **280**(5369): p. 1564-9.

453. Bunger, M.K., et al., *Mop3 is an essential component of the master circadian pacemaker in mammals.* Cell, 2000. **103**(7): p. 1009-17.

454. Jin, X., et al., *A molecular mechanism regulating rhythmic output from the suprachiasmatic circadian clock.* Cell, 1999. **96**(1): p. 57-68.

455. Zheng, B., et al., *The mPer2 gene encodes a functional component of the mammalian circadian clock.* Nature, 1999. **400**(6740): p. 169-73.

456. Zheng, B., et al., *Nonredundant roles of the mPer1 and mPer2 genes in the mammalian circadian clock.* Cell, 2001. **105**(5): p. 683-94.

457. Yoo, S.H., et al., *A noncanonical E-box enhancer drives mouse Period2 circadian oscillations in vivo.* Proc Natl Acad Sci U S A, 2005. **102**(7): p. 2608-13.

458. Kume, K., et al., *mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop.* Cell, 1999. **98**(2): p. 193-205.

459. Griffin, E.A., Jr., D. Staknis, and C.J. Weitz, *Light-independent role of CRY1 and CRY2 in the mammalian circadian clock.* Science, 1999. **286**(5440): p. 768-71.

460. Mohawk, J.A., C.B. Green, and J.S. Takahashi, *Central and peripheral circadian clocks in mammals.* Annu Rev Neurosci, 2012. **35**: p. 445-62.

461. Shirogane, T., et al., *SCFbeta-TRCP controls clock-dependent transcription via casein kinase 1-dependent degradation of the mammalian period-1 (Per1) protein.* J Biol Chem, 2005. **280**(29): p. 26863-72.

462. Reischl, S., et al., *Beta-TrCP1-mediated degradation of PERIOD2 is essential for circadian dynamics.* J Biol Rhythms, 2007. **22**(5): p. 375-86.

463. Busino, L., et al., *SCFFbxl3 controls the oscillation of the circadian clock by directing the degradation of cryptochrome proteins.* Science, 2007. **316**(5826): p. 900-4.

464. Siepka, S.M., et al., *Circadian mutant Overtime reveals F-box protein FBXL3 regulation of cryptochrome and period gene expression.* Cell, 2007. **129**(5): p. 1011-23.

465. Lee, C., et al., *Posttranslational mechanisms regulate the mammalian circadian clock.* Cell, 2001. **107**(7): p. 855-67.

466. Lamia, K.A., et al., *AMPK regulates the circadian clock by cryptochrome phosphorylation and degradation.* Science, 2009. **326**(5951): p. 437-40.

467. Asher, G., et al., *SIRT1 regulates circadian clock gene expression through PER2 deacetylation.* Cell, 2008. **134**(2): p. 317-28.

468. Yoo, S.H., et al., *Competing E3 ubiquitin ligases govern circadian periodicity by degradation of CRY in nucleus and cytoplasm.* Cell, 2013. **152**(5): p. 1091-105.

469. St John, P.C., et al., *Spatiotemporal separation of PER and CRY posttranslational regulation in the mammalian circadian clock.* Proc Natl Acad Sci U S A, 2014. **111**(5): p. 2040-5.

470. Stratmann, M., et al., *Circadian Dbp transcription relies on highly dynamic BMAL1-CLOCK interaction with E boxes and requires the proteasome.* Molecular Cell, 2012. **48**(2): p. 277-87.

471. Preitner, N., et al., *The orphan nuclear receptor REV-ERBalpha controls circadian transcription within the positive limb of the mammalian circadian oscillator.* Cell, 2002. **110**(2): p. 251-60.

472. Sato, T.K., et al., *A functional genomics strategy reveals Rora as a component of the mammalian circadian clock.* Neuron, 2004. **43**(4): p. 527-37.

473. Nakajima, Y., et al., *Bidirectional role of orphan nuclear receptor RORalpha in clock gene transcriptions demonstrated by a novel reporter assay system.* FEBS Lett, 2004. **565**(1-3): p. 122-6.

474. Guillaumond, F., et al., *Differential control of Bmal1 circadian transcription by REV-ERB and ROR nuclear receptors.* J Biol Rhythms, 2005. **20**(5): p. 391-403.

475. Liu, A.C., et al., *Redundant function of REV-ERBalpha and beta and non-essential role for Bmal1 cycling in transcriptional regulation of intracellular circadian rhythms.* PLoS Genet, 2008. **4**(2): p. e1000023.

476. Takeda, Y., et al., *RORgamma directly regulates the circadian expression of clock genes and downstream targets in vivo.* Nucleic Acids Res, 2012. **40**(17): p. 8519-35.

477. Gachon, F., et al., *The circadian PAR-domain basic leucine zipper transcription factors DBP, TEF, and HLF modulate basal and inducible xenobiotic detoxification.* Cell Metab, 2006. **4**(1): p. 25-36.

478. Brown, S.A., *Circadian clock-mediated control of stem cell division and differentiation: beyond night and day.* Development, 2014. **141**(16): p. 3105-11.

479. Koike, N., et al., *Transcriptional architecture and chromatin landscape of the core circadian clock in mammals.* Science, 2012. **338**(6105): p. 349-54.

480. Ueda, H.R., et al., *A transcription factor response element for gene expression during circadian night.* Nature, 2002. **418**(6897): p. 534-9.

481. Fang, B., et al., *Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo.* Cell, 2014. **159**(5): p. 1140-52.

482. Zhang, R., et al., *A circadian gene expression atlas in mammals: implications for biology and medicine.* Proc Natl Acad Sci U S A, 2014. **111**(45): p. 16219-24.

483. Storch, K.F., et al., *Extensive and divergent circadian gene expression in liver and heart.* Nature, 2002. **417**(6884): p. 78-83.

484. Panda, S., et al., *Coordinated transcription of key pathways in the mouse by the circadian clock.* Cell, 2002. **109**(3): p. 307-20.

485. Hughes, M.E., et al., *Harmonics of circadian gene transcription in mammals.* PLoS Genet, 2009. **5**(4): p. e1000442.

486. Le Martelot, G., et al., *Genome-wide RNA polymerase II profiles and RNA accumulation reveal kinetics of transcription and associated epigenetic changes during diurnal cycles.* PLoS Biol, 2012. **10**(11): p. e1001442.

487. Staiger, D. and T. Koster, *Spotlight on post-transcriptional control in the circadian system.* Cell Mol Life Sci, 2011. **68**(1): p. 71-83.

488. Kojima, S., D.L. Shingle, and C.B. Green, *Post-transcriptional control of circadian rhythms.* J Cell Sci, 2011. **124**(Pt 3): p. 311-20.

489. Lim, C. and R. Allada, *Emerging roles for post-transcriptional regulation in circadian clocks.* Nat Neurosci, 2013. **16**(11): p. 1544-50.

490. Partch, C.L., C.B. Green, and J.S. Takahashi, *Molecular architecture of the mammalian circadian clock.* Trends Cell Biol, 2014. **24**(2): p. 90-9.

491. O'Neill, J.S., et al., *Circadian rhythms persist without transcription in a eukaryote.* Nature, 2011. **469**(7331): p. 554-8.

492. O'Neill, J.S. and A.B. Reddy, *Circadian clocks in human red blood cells.* Nature, 2011. **469**(7331): p. 498-503.

493. Menet, J.S., S. Pescatore, and M. Rosbash, *CLOCK:BMAL1 is a pioneer-like transcription factor.* Genes Dev, 2014. **28**(1): p. 8-13.

494. Valekunja, U.K., et al., *Histone methyltransferase MLL3 contributes to genome-scale circadian transcription.* Proc Natl Acad Sci U S A, 2013. **110**(4): p. 1554-9.

495. Katada, S. and P. Sassone-Corsi, *The histone methyltransferase MLL1 permits the oscillation of circadian gene expression.* Nat Struct Mol Biol, 2010. **17**(12): p. 1414-21.

496. Doi, M., J. Hirayama, and P. Sassone-Corsi, *Circadian regulator CLOCK is a histone acetyltransferase.* Cell, 2006. **125**(3): p. 497-508.

497. Aguilar-Arnal, L., et al., *NAD(+)-SIRT1 control of H3K4 trimethylation through circadian deacetylation of MLL1.* Nat Struct Mol Biol, 2015. **22**(4): p. 312-8.

498. Nakahata, Y., et al., *The NAD+-dependent deacetylase SIRT1 modulates CLOCK-mediated chromatin remodeling and circadian control.* Cell, 2008. **134**(2): p. 329-40.

499. Duong, H.A., et al., *A molecular mechanism for circadian clock negative feedback.* Science, 2011. **332**(6036): p. 1436-9.

500. Tamayo, A.G., et al., *Histone monoubiquitination by Clock-Bmal1 complex marks Per1 and Per2 genes for circadian feedback.* Nat Struct Mol Biol, 2015. **22**(10): p. 759-66.

501. Etchegaray, J.P., et al., *Rhythmic histone acetylation underlies transcription in the mammalian circadian clock.* Nature, 2003. **421**(6919): p. 177-82.

502. Hosoda, H., et al., *CBP/p300 is a cell type-specific modulator of CLOCK/BMAL1-mediated transcription.* Mol Brain, 2009. **2**: p. 34.

503. Sun, Z., et al., *Circadian epigenomic remodeling and hepatic lipogenesis: lessons from HDAC3.* Cold Spring Harb Symp Quant Biol, 2011. **76**: p. 49-55.

504. Shi, G., et al., *Distinct Roles of HDAC3 in the Core Circadian Negative Feedback Loop Are Critical for Clock Function.* Cell Rep, 2016. **14**(4): p. 823-34.

505. Ripperger, J.A., et al., *CLOCK, an essential pacemaker component, controls expression of the circadian transcription factor DBP.* Genes Dev, 2000. **14**(6): p. 679-89.

506. Ripperger, J.A. and U. Schibler, *Rhythmic CLOCK-BMAL1 binding to multiple E-box motifs drives circadian Dbp transcription and chromatin transitions.* Nat Genet, 2006. **38**(3): p. 369-74.

507. Aguilar-Arnal, L., et al., *Cycles in spatial and temporal chromosomal organization driven by the circadian clock.* Nat Struct Mol Biol, 2013. **20**(10): p. 1206-13.

508. Lavery, D.J., et al., *Circadian expression of the steroid 15 alpha-hydroxylase (Cyp2a4) and coumarin 7-hydroxylase (Cyp2a5) genes in mouse liver is regulated by the PAR leucine zipper transcription factor DBP.* Mol Cell Biol, 1999. **19**(10): p. 6488-99.

509. Lopez-Molina, L., et al., *The DBP gene is expressed according to a circadian rhythm in the suprachiasmatic nucleus and influences circadian behavior.* EMBO J, 1997. **16**(22): p. 6762-71.

510. Wuarin, J., et al., *The role of the transcriptional activator protein DBP in circadian liver gene expression.* J Cell Sci Suppl, 1992. **16**: p. 123-7.

511. Mehta, N. and H.Y. Cheng, *Micro-managing the circadian clock: The role of microRNAs in biological timekeeping.* J Mol Biol, 2013. **425**(19): p. 3609-24.

512. Du, N.H., et al., *MicroRNAs shape circadian hepatic gene expression on a transcriptome-wide scale.* Elife, 2014. **3**: p. e02510.

513. Shende, V.R., et al., *MicroRNAs function as cis- and trans-acting modulators of peripheral circadian clocks.* FEBS Lett, 2014. **588**(17): p. 3015-22.

514. Vollmers, C., et al., *Circadian oscillations of protein-coding and regulatory RNAs in a highly dynamic mammalian liver epigenome.* Cell Metab, 2012. **16**(6): p. 833-45.

515. Coon, S.L., et al., *Circadian changes in long noncoding RNAs in the pineal gland.* Proc Natl Acad Sci U S A, 2012. **109**(33): p. 13319-24.

516. Li, N., et al., *The frequency natural antisense transcript first promotes, then represses, frequency gene expression via facultative heterochromatin.* Proc Natl Acad Sci U S A, 2015. **112**(14): p. 4357-62.

517. Aronson, B.D., et al., *Negative feedback defining a circadian clock: autoregulation of the clock gene frequency.* Science, 1994. **263**(5153): p. 1578-84.

518. Kramer, C., et al., *Role for antisense RNA in regulating circadian clock function in Neurospora crassa.* Nature, 2003. **421**(6926): p. 948-52.

519. Xue, Z., et al., *Transcriptional interference by antisense RNA is required for circadian clock function.* Nature, 2014. **514**(7524): p. 650-3.

520. Azzi, A., et al., *Circadian behavior is light-reprogrammed by plastic DNA methylation.* Nat Neurosci, 2014. **17**(3): p. 377-82.

521. Solnica-Krezel, L., *Conserved patterns of cell movements during vertebrate gastrulation.* Curr Biol, 2005. **15**(6): p. R213-28.

522. Colas, J.F. and G.C. Schoenwolf, *Towards a cellular and molecular understanding of neurulation.* Dev Dyn, 2001. **221**(2): p. 117-45.

523. Ishikawa, Y., et al., *The primary brain vesicles revisited: are the three primary vesicles (forebrain/midbrain/hindbrain) universal in vertebrates?* Brain Behav Evol, 2012. **79**(2): p. 75-83.

524. Greig, L.C., et al., *Molecular logic of neocortical projection neuron specification, development and diversity.* Nat Rev Neurosci, 2013. **14**(11): p. 755-69.

525. Tan, S.S., et al., *Separate progenitors for radial and tangential cell dispersion during development of the cerebral neocortex.* Neuron, 1998. **21**(2): p. 295-304.

526. Parnavelas, J.G., *The origin and migration of cortical neurones: new vistas.* Trends Neurosci, 2000. **23**(3): p. 126-31.

527. Anderson, S.A., et al., *Distinct origins of neocortical projection neurons and interneurons in vivo.* Cereb Cortex, 2002. **12**(7): p. 702-9.

528. Wichterle, H., et al., *In utero fate mapping reveals distinct migratory pathways and fates of neurons born in the mammalian basal forebrain.* Development, 2001. **128**(19): p. 3759-71.

529. Wonders, C.P. and S.A. Anderson, *The origin and specification of cortical interneurons.* Nat Rev Neurosci, 2006. **7**(9): p. 687-96.

530. Chu, J. and S.A. Anderson, *Development of cortical interneurons.* Neuropsychopharmacology, 2015. **40**(1): p. 16-23.

531. Gorski, J.A., et al., *Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage.* J Neurosci, 2002. **22**(15): p. 6309-14.

532. Molyneaux, B.J., et al., *Neuronal subtype specification in the cerebral cortex.* Nat Rev Neurosci, 2007. **8**(6): p. 427-37.

533. Angevine, J.B., Jr. and R.L. Sidman, *Autoradiographic study of cell migration during histogenesis of cerebral cortex in the mouse.* Nature, 1961. **192**: p. 766-8.

534. Gotz, M. and W.B. Huttner, *The cell biology of neurogenesis.* Nat Rev Mol Cell Biol, 2005. **6**(10): p. 777-88.

535. Kriegstein, A.R. and M. Gotz, *Radial glia diversity: a matter of cell fate.* Glia, 2003. **43**(1): p. 37-43.

536. Malatesta, P., E. Hartfuss, and M. Gotz, *Isolation of radial glial cells by fluorescent-activated cell sorting reveals a neuronal lineage.* Development, 2000. **127**(24): p. 5253-63.

537. Noctor, S.C., et al., *Neurons derived from radial glial cells establish radial units in neocortex.* Nature, 2001. **409**(6821): p. 714-20.

538. Miyata, T., et al., *Asymmetric inheritance of radial glial fibers by cortical neurons.* Neuron, 2001. **31**(5): p. 727-41.

539. Noctor, S.C., et al., *Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases.* Nat Neurosci, 2004. **7**(2): p. 136-44.

540. Haubensak, W., et al., *Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: a major site of neurogenesis.* Proc Natl Acad Sci U S A, 2004. **101**(9): p. 3196-201.

541. Sessa, A., et al., *Tbr2 directs conversion of radial glia into basal precursors and guides neuronal amplification by indirect neurogenesis in the developing neocortex.* Neuron, 2008. **60**(1): p. 56-69.

542. Hansen, D.V., et al., *Neurogenic radial glia in the outer subventricular zone of human neocortex.* Nature, 2010. **464**(7288): p. 554-561.

543. Wang, X., et al., *A new subtype of progenitor cell in the mouse embryonic neocortex.* Nat Neurosci, 2011. **14**(5): p. 555-61.

544. Florio, M. and W.B. Huttner, *Neural progenitors, neurogenesis and the evolution of the neocortex.* Development, 2014. **141**(11): p. 2182-94.

545. Fietz, S.A., et al., *OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling.* Nat Neurosci, 2010. **13**(6): p. 690-9.

546. Shitamukai, A., D. Konno, and F. Matsuzaki, *Oblique radial glial divisions in the developing mouse neocortex induce self-renewing progenitors outside the germinal zone that resemble primate outer subventricular zone progenitors.* J Neurosci, 2011. **31**(10): p. 3683-95.

547. Miyata, T., et al., *Asymmetric production of surface-dividing and non-surface-dividing cortical progenitor cells.* Development, 2004. **131**(13): p. 3133-45.

548. Shibata, M., F.O. Gulden, and N. Sestan, *From trans to cis: transcriptional regulatory networks in neocortical development.* Trends Genet, 2015. **31**(2): p. 77-87.

549. Kwan, K.Y., N. Sestan, and E.S. Anton, *Transcriptional co-regulation of neuronal migration and laminar identity in the neocortex.* Development, 2012. **139**(9): p. 1535-46.

550. Paridaen, J.T. and W.B. Huttner, *Neurogenesis during development of the vertebrate central nervous system.* EMBO Rep, 2014. **15**(4): p. 351-64.

551. Vieira, C., et al., *Molecular mechanisms controlling brain development: an overview of neuroepithelial secondary organizers.* Int J Dev Biol, 2010. **54**(1): p. 7-20.

552. Azzarelli, R., L.J. Hardwick, and A. Philpott, *Emergence of neuronal diversity from patterning of telencephalic progenitors.* Wiley Interdiscip Rev Dev Biol, 2015. **4**(3): p. 197-214.

553. Siddiqi, F., et al., *Fate mapping by piggyBac transposase reveals that neocortical GLAST+ progenitors generate more astrocytes than Nestin+ progenitors in rat neocortex.* Cereb Cortex, 2014. **24**(2): p. 508-20.

554. Tabata, H., *Diverse subtypes of astrocytes and their development during corticogenesis.* Front Neurosci, 2015. **9**: p. 114.

555. Hirabayashi, Y. and Y. Gotoh, *Epigenetic control of neural precursor cell fate during development.* Nat Rev Neurosci, 2010. **11**(6): p. 377-88.

556. Ypsilanti, A.R. and J.L. Rubenstein, *Transcriptional and epigenetic mechanisms of early cortical development: An examination of how Pax6 coordinates cortical development.* J Comp Neurol, 2016. **524**(3): p. 609-29.

557. Osumi, N., et al., *Concise review: Pax6 transcription factor contributes to both embryonic and adult neurogenesis as a multifunctional regulator.* Stem Cells, 2008. **26**(7): p. 1663-72.

558. Scardigli, R., et al., *Direct and concentration-dependent regulation of the proneural gene Neurogenin2 by Pax6.* Development, 2003. **130**(14): p. 3269-81.

559. Sansom, S.N., et al., *The level of the transcription factor Pax6 is essential for controlling the balance between neural stem cell self-renewal and neurogenesis.* PLoS Genet, 2009. **5**(6): p. e1000511.

560. Walcher, T., et al., *Functional dissection of the paired domain of Pax6 reveals molecular mechanisms of coordinating neurogenesis and proliferation.* Development, 2013. **140**(5): p. 1123-36.

561. Porter, F.D., et al., *Lhx2, a LIM homeobox gene, is required for eye, forebrain, and definitive erythrocyte development.* Development, 1997. **124**(15): p. 2935-44.

562. Warren, N., et al., *The transcription factor, Pax6, is required for cell proliferation and differentiation in the developing cerebral cortex.* Cereb Cortex, 1999. **9**(6): p. 627-35.

563. Estivill-Torrus, G., et al., *Pax6 is required to regulate the cell cycle and the rate of progression from symmetrical to asymmetrical division in mammalian cortical progenitors.* Development, 2002. **129**(2): p. 455-66.

564. Hanashima, C., et al., *Brain factor-1 controls the proliferation and differentiation of neocortical progenitor cells through independent mechanisms.* J Neurosci, 2002. **22**(15): p. 6526-36.

565. Roy, K., et al., *The Tlx gene regulates the timing of neurogenesis in the cortex.* J Neurosci, 2004. **24**(38): p. 8333-45.

566. Friocourt, G., et al., *Cell-autonomous roles of ARX in cell proliferation and neuronal migration during corticogenesis.* J Neurosci, 2008. **28**(22): p. 5794-805.

567. Martynoga, B., D. Drechsel, and F. Guillemot, *Molecular control of neurogenesis: a view from the mammalian cerebral cortex.* Cold Spring Harb Perspect Biol, 2012. **4**(10).

568. Siegenthaler, J.A., B.A. Tremper-Wells, and M.W. Miller, *Foxg1 haploinsufficiency reduces the population of cortical intermediate progenitor cells: effect of increased p21 expression.* Cereb Cortex, 2008. **18**(8): p. 1865-75.

569. Yoshida, M., et al., *Emx1 and Emx2 functions in development of dorsal telencephalon.* Development, 1997. **124**(1): p. 101-11.

570. Bertrand, N., D.S. Castro, and F. Guillemot, *Proneural genes and the specification of neural cell types.* Nat Rev Neurosci, 2002. **3**(7): p. 517-30.

571. Lee, J.E., *Basic helix-loop-helix genes in neural development.* Curr Opin Neurobiol, 1997. **7**(1): p. 13-20.

572. Nieto, M., et al., *Neural bHLH genes control the neuronal versus glial fate decision in cortical progenitors.* Neuron, 2001. **29**(2): p. 401-13.

573. Imayoshi, I. and R. Kageyama, *bHLH factors in self-renewal, multipotency, and fate choice of neural progenitor cells.* Neuron, 2014. **82**(1): p. 9-23.

574. Ma, Q., C. Kintner, and D.J. Anderson, *Identification of neurogenin, a vertebrate neuronal determination gene.* Cell, 1996. **87**(1): p. 43-52.

575. Nakada, Y., et al., *Distinct domains within Mash1 and Math1 are required for function in neuronal differentiation versus neuronal cell-type specification.* Development, 2004. **131**(6): p. 1319-30.

576. Berninger, B., F. Guillemot, and M. Gotz, *Directing neurotransmitter identity of neurones derived from expanded adult neural stem cells.* Eur J Neurosci, 2007. **25**(9): p. 2581-90.

577. Zhang, Y., et al., *Rapid single-step induction of functional neurons from human pluripotent stem cells.* Neuron, 2013. **78**(5): p. 785-98.

578. Heinrich, C., et al., *Directing astroglia from the cerebral cortex into subtype specific functional neurons.* PLoS Biol, 2010. **8**(5): p. e1000373.

579. Vierbuchen, T., et al., *Direct conversion of fibroblasts to functional neurons by defined factors.* Nature, 2010. **463**(7284): p. 1035-41.

580. Thoma, E.C., et al., *Ectopic expression of neurogenin 2 alone is sufficient to induce differentiation of embryonic stem cells into mature neurons.* PLoS One, 2012. **7**(6): p. e38651.

581. Casarosa, S., C. Fode, and F. Guillemot, *Mash1 regulates neurogenesis in the ventral telencephalon.* Development, 1999. **126**(3): p. 525-34.

582. Mattar, P., et al., *Basic helix-loop-helix transcription factors cooperate to specify a cortical projection neuron identity.* Mol Cell Biol, 2008. **28**(5): p. 1456-69.

583. Kovach, C., et al., *Neurog2 simultaneously activates and represses alternative gene expression programs in the developing neocortex.* Cereb Cortex, 2013. **23**(8): p. 1884-900.

584. Kageyama, R., et al., *Dynamic Notch signaling in neural progenitor cells and a revised view of lateral inhibition.* Nat Neurosci, 2008. **11**(11): p. 1247-51.

585. Shimojo, H., T. Ohtsuka, and R. Kageyama, *Oscillations in notch signaling regulate maintenance of neural progenitors.* Neuron, 2008. **58**(1): p. 52-64.

586. Imayoshi, I., et al., *Oscillatory control of factors determining multipotency and fate in mouse neural progenitors.* Science, 2013. **342**(6163): p. 1203-8.

587. Castro, D.S., et al., *A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets.* Genes Dev, 2011. **25**(9): p. 930-45.

588. Ochiai, W., et al., *Periventricular notch activation and asymmetric Ngn2 and Tbr2 expression in pair-generated neocortical daughter cells.* Mol Cell Neurosci, 2009. **40**(2): p. 225-33.

589. Englund, C., et al., *Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex.* J Neurosci, 2005. **25**(1): p. 247-51.

590. Arnold, S.J., et al., *The T-box transcription factor Eomes/Tbr2 regulates neurogenesis in the cortical subventricular zone.* Genes Dev, 2008. **22**(18): p. 2479-84.

591. Farkas, L.M., et al., *Insulinoma-associated 1 has a panneurogenic role and promotes the generation and expansion of basal progenitors in the developing mouse neocortex.* Neuron, 2008. **60**(1): p. 40-55.

592. Pinto, L., et al., *AP2gamma regulates basal progenitor fate in a region- and layer-specific manner in the developing cortex.* Nat Neurosci, 2009. **12**(10): p. 1229-37.

593. Schuurmans, C., et al., *Sequential phases of cortical specification involve Neurogenin-dependent and -independent pathways.* EMBO J, 2004. **23**(14): p. 2892-902.

594. Mattar, P., et al., *A screen for downstream effectors of Neurogenin2 in the embryonic neocortex.* Dev Biol, 2004. **273**(2): p. 373-89.

595. Bulfone, A., et al., *T-brain-1: a homolog of Brachyury whose expression defines molecularly distinct domains within the cerebral cortex.* Neuron, 1995. **15**(1): p. 63-78.

596. Hevner, R.F., et al., *Tbr1 regulates differentiation of the preplate and layer 6.* Neuron, 2001. **29**(2): p. 353-66.

597. Ross, S.E., M.E. Greenberg, and C.D. Stiles, *Basic helix-loop-helix factors in cortical development.* Neuron, 2003. **39**(1): p. 13-25.

598. Borello, U. and A. Pierani, *Patterning the cerebral cortex: traveling with morphogens.* Curr Opin Genet Dev, 2010. **20**(4): p. 408-15.

599. Arai, Y. and A. Pierani, *Development and evolution of cortical fields.* Neurosci Res, 2014. **86**: p. 66-76.

600. Ishibashi, M., et al., *Targeted disruption of mammalian hairy and Enhancer of split homolog-1 (HES-1) leads to up-regulation of neural helix-loop-helix factors, premature neurogenesis, and severe neural tube defects.* Genes Dev, 1995. **9**(24): p. 3136-48.

601. de la Pompa, J.L., et al., *Conservation of the Notch signalling pathway in mammalian neurogenesis.* Development, 1997. **124**(6): p. 1139-48.

602. Ohtsuka, T., et al., *Hes1 and Hes5 as notch effectors in mammalian neuronal differentiation.* EMBO J, 1999. **18**(8): p. 2196-207.

603. Ohtsuka, T., et al., *Roles of the basic helix-loop-helix genes Hes1 and Hes5 in expansion of neural stem cells of the developing brain.* J Biol Chem, 2001. **276**(32): p. 30467-74.

604. Hitoshi, S., et al., *Notch pathway molecules are essential for the maintenance, but not the generation, of mammalian neural stem cells.* Genes Dev, 2002. **16**(7): p. 846-58.

605. Lutolf, S., et al., *Notch1 is required for neuronal and glial differentiation in the cerebellum.* Development, 2002. **129**(2): p. 373-85.

606. Hatakeyama, J., et al., *Hes genes regulate size, shape and histogenesis of the nervous system by control of the timing of neural stem cell differentiation.* Development, 2004. **131**(22): p. 5539-50.

607. Yoon, K. and N. Gaiano, *Notch signaling in the mammalian central nervous system: insights from mouse mutants.* Nat Neurosci, 2005. **8**(6): p. 709-15.

608. Mizutani, K., et al., *Differential Notch signalling distinguishes neural stem cells from intermediate progenitors.* Nature, 2007. **449**(7160): p. 351-5.

609. Imayoshi, I., et al., *Hes genes and neurogenin regulate non-neural versus neural fate specification in the dorsal telencephalic midline.* Development, 2008. **135**(15): p. 2531-41.

610. Kageyama, R., H. Shimojo, and I. Imayoshi, *Dynamic expression and roles of Hes factors in neural development.* Cell Tissue Res, 2015. **359**(1): p. 125-33.

611. Hirata, H., et al., *Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop.* Science, 2002. **298**(5594): p. 840-3.

612. Dong, Z., et al., *Intralineage directional Notch signaling regulates self-renewal and differentiation of asymmetrically dividing radial glia.* Neuron, 2012. **74**(1): p. 65-78.

613. Vasconcelos, F.F. and D.S. Castro, *Transcriptional control of vertebrate neurogenesis by the proneural factor Ascl1.* Front Cell Neurosci, 2014. **8**: p. 412.

614. Ables, J.L., et al., *Not(ch) just development: Notch signalling in the adult brain.* Nat Rev Neurosci, 2011. **12**(5): p. 269-83.

615. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.

616. Azuara, V., et al., *Chromatin signatures of pluripotent cell lines.* Nat Cell Biol, 2006. **8**(5): p. 532-8.

617. Mohn, F., et al., *Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors.* Molecular Cell, 2008. **30**(6): p. 755-66.

618. Shimomura A., H.E., *Epigenetic Regulation of Neural Differentiation from Embryonic Stem Cells.* Trends in Cell Signaling Pathways in Neuronal Fate Decision, 2013. **ISBN: 978-953-51-1059-0, InTech**.

619. Corley, M. and K.L. Kroll, *The roles and regulation of Polycomb complexes in neural development.* Cell Tissue Res, 2015. **359**(1): p. 65-85.

620. Pereira, J.D., et al., *Ezh2, the histone methyltransferase of PRC2, regulates the balance between self-renewal and differentiation in the cerebral cortex.* Proc Natl Acad Sci U S A, 2010. **107**(36): p. 15957-62.

621. Jepsen, K., et al., *SMRT-mediated repression of an H3K27 demethylase in progression from neural stem cell to neuron.* Nature, 2007. **450**(7168): p. 415-9.

622. Burgold, T., et al., *The histone H3 lysine 27-specific demethylase Jmjd3 is required for neural commitment.* PLoS One, 2008. **3**(8): p. e3034.

623. Golebiewska, A., et al., *Epigenetic landscaping during hESC differentiation to neural cells.* Stem Cells, 2009. **27**(6): p. 1298-308.

624. Wen, B., et al., *Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells.* Nat Genet, 2009. **41**(2): p. 246-50.

625. Oda, M., et al., *DNA methylation restricts lineage-specific functions of transcription factor Gata4 during embryonic stem cell differentiation.* PLoS Genet, 2013. **9**(6): p. e1003574.

626. Xie, W., et al., *Epigenomic analysis of multilineage differentiation of human embryonic stem cells.* Cell, 2013. **153**(5): p. 1134-48.

627. Auclair, G., et al., *Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse.* Genome Biol, 2014. **15**(12): p. 545.

628. Fan, G., et al., *DNA methylation controls the timing of astrogliogenesis through regulation of JAK-STAT signaling.* Development, 2005. **132**(15): p. 3345-56.

629. Wu, Z., et al., *Dnmt3a regulates both proliferation and differentiation of mouse neural stem cells.* J Neurosci Res, 2012. **90**(10): p. 1883-91.

630. Watanabe, D., K. Uchiyama, and K. Hanaoka, *Transition of mouse de novo methyltransferases expression from Dnmt3b to Dnmt3a during neural progenitor cell development.* Neuroscience, 2006. **142**(3): p. 727-37.

631. Feng, J., et al., *Dynamic expression of de novo DNA methyltransferases Dnmt3a and Dnmt3b in the central nervous system.* J Neurosci Res, 2005. **79**(6): p. 734-46.

632. Martins-Taylor, K., et al., *Role of DNMT3B in the regulation of early neural and neural crest specifiers.* Epigenetics, 2012. **7**(1): p. 71-82.

633. Hitoshi, S., et al., *Mammalian Gcm genes induce Hes5 expression by active DNA demethylation and induce neural stem cells.* Nat Neurosci, 2011. **14**(8): p. 957-64.

634. Lv, X., et al., *MicroRNA-15b promotes neurogenesis and inhibits neural progenitor proliferation by directly repressing TET3 during early neocortical development.* EMBO Rep, 2014. **15**(12): p. 1305-14.

635. Li, T., et al., *Critical role of Tet3 in neural progenitor cell maintenance and terminal differentiation.* Mol Neurobiol, 2015. **51**(1): p. 142-54.

636. Aprea, J., et al., *Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment.* EMBO J, 2013. **32**(24): p. 3145-60.

637. Cao, X., et al., *Noncoding RNAs in the mammalian central nervous system.* Annu Rev Neurosci, 2006. **29**: p. 77-103.

638. Ng, S.Y., R. Johnson, and L.W. Stanton, *Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors.* EMBO J, 2012. **31**(3): p. 522-33.

639. Hu, W., J.R. Alvarez-Dominguez, and H.F. Lodish, *Regulation of mammalian cell differentiation by long non-coding RNAs.* EMBO Rep, 2012. **13**(11): p. 971-83.

640. Clark, B.S. and S. Blackshaw, *Long non-coding RNA-dependent transcriptional regulation in neuronal development and disease.* Front Genet, 2014. **5**: p. 164.

641. Nowakowski, T.J., et al., *MicroRNA-92b regulates the development of intermediate cortical progenitors in embryonic mouse brain.* Proc Natl Acad Sci U S A, 2013. **110**(17): p. 7056-61.

642. Shibata, M., et al., *MicroRNA-9 regulates neurogenesis in mouse telencephalon by targeting multiple transcription factors.* J Neurosci, 2011. **31**(9): p. 3407-22.

643. Mehler, M.F. and J.S. Mattick, *Non-coding RNAs in the nervous system.* J Physiol, 2006. **575**(Pt 2): p. 333-41.

644. Mehler, M.F. and J.S. Mattick, *Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease.* Physiol Rev, 2007. **87**(3): p. 799-823.

645. Kosik, K.S., *The neuronal microRNA system.* Nat Rev Neurosci, 2006. **7**(12): p. 911-20.

646. Szafranski, K., K.J. Abraham, and K. Mekhail, *Non-coding RNA in neural function, disease, and aging.* Front Genet, 2015. **6**: p. 87.

647. Lv, J., et al., *Identification and characterization of long non-coding RNAs related to mouse embryonic brain development from available transcriptomic data.* PLoS One, 2013. **8**(8): p. e71152.

648. Cimadamore, F., et al., *SOX2-LIN28/let-7 pathway regulates proliferation and neurogenesis in neural precursors.* Proc Natl Acad Sci U S A, 2013. **110**(32): p. E3017-26.

649. Patterson, M., et al., *let-7 miRNAs can act through notch to regulate human gliogenesis.* Stem Cell Reports, 2014. **3**(5): p. 758-73.

650. Rehfeld, F., et al., *Lin28 and let-7: ancient milestones on the road from pluripotency to neurogenesis.* Cell Tissue Res, 2015. **359**(1): p. 145-60.

651. Schwamborn, J.C., E. Berezikov, and J.A. Knoblich, *The TRIM-NHL protein TRIM32 activates microRNAs and prevents self-renewal in mouse neural progenitors.* Cell, 2009. **136**(5): p. 913-25.

652. Ramos, A.D., et al., *The long noncoding RNA Pnky regulates neuronal differentiation of embryonic and postnatal neural stem cells.* Cell Stem Cell, 2015. **16**(4): p. 439-47.

653. Fraser, J., et al., *Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation.* Mol Syst Biol, 2015. **11**(12): p. 852.

654. Alexander, J.M. and S. Lomvardas, *Nuclear architecture as an epigenetic regulator of neural development and function.* Neuroscience, 2014. **264**: p. 39-50.

655. Shchuka, V.M., et al., *Chromatin Dynamics in Lineage Commitment and Cellular Reprogramming.* Genes (Basel), 2015. **6**(3): p. 641-61.

656. Raposo, A.A., et al., *Ascl1 Coordinately Regulates Gene Expression and the Chromatin Landscape during Neurogenesis.* Cell Rep, 2015.

657. Lessard, J., et al., *An essential switch in subunit composition of a chromatin remodeling complex during neural development.* Neuron, 2007. **55**(2): p. 201-15.

658. Yan, Z., et al., *BAF250B-associated SWI/SNF chromatin-remodeling complex is required to maintain undifferentiated mouse embryonic stem cells.* Stem Cells, 2008. **26**(5): p. 1155-65.

659. Ho, L., et al., *An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network.* Proc Natl Acad Sci U S A, 2009. **106**(13): p. 5187-91.

660. Ho, L., et al., *An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency.* Proc Natl Acad Sci U S A, 2009. **106**(13): p. 5181-6.

661. Yoo, A.S. and G.R. Crabtree, *ATP-dependent chromatin remodeling in neural development.* Curr Opin Neurobiol, 2009. **19**(2): p. 120-6.

662. Takeuchi, J.K., et al., *Baf60c is a nuclear Notch signaling component required for the establishment of left-right asymmetry.* Proc Natl Acad Sci U S A, 2007. **104**(3): p. 846-51.

663. Lamba, D.A., et al., *Baf60c is a component of the neural progenitor-specific BAF complex in developing retina.* Dev Dyn, 2008. **237**(10): p. 3016-23.

664. Tuoc, T.C., et al., *Chromatin regulation by BAF170 controls cerebral cortical size and thickness.* Dev Cell, 2013. **25**(3): p. 256-69.

665. Seo, S., G.A. Richardson, and K.L. Kroll, *The SWI/SNF chromatin remodeling protein Brg1 is required for vertebrate neurogenesis and mediates transactivation of Ngn and NeuroD.* Development, 2005. **132**(1): p. 105-15.

666. Yoo, A.S., et al., *MicroRNA-mediated switching of chromatin-remodelling complexes in neural development (vol 460, pg 642, 2009).* Nature, 2009. **461**(7261): p. 296-296.

667. Ballas, N., et al., *REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis.* Cell, 2005. **121**(4): p. 645-657.

668. Ballas, N. and G. Mandel, *The many faces of REST oversee epigenetic programming of neuronal genes.* Curr Opin Neurobiol, 2005. **15**(5): p. 500-506.

669. Chong, J.A., et al., *REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons.* Cell, 1995. **80**(6): p. 949-57.

670. Schoenherr, C.J. and D.J. Anderson, *The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes.* Science, 1995. **267**(5202): p. 1360-3.

671. Lunyak, V.V., et al., *Corepressor-dependent silencing of chromosomal regions encoding neuronal genes.* Science, 2002. **298**(5599): p. 1747-52.

672. Vance, K.W., et al., *The long non-coding RNA Paupar regulates the expression of both local and distal genes.* EMBO J, 2014. **33**(4): p. 296-311.

673. You, J.S., et al., *SNF5 is an essential executor of epigenetic regulation during differentiation.* PLoS Genet, 2013. **9**(4): p. e1003459.

674. Pataskar, A., et al., *NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program.* EMBO J, 2016. **35**(1): p. 24-45.

# 2.1

# Dynamics of chromatin accessibility and epigenetic state in response to UV damage

Sandra Schick[1, #], David Fournier[1, #], Sudhir Thakurela[1], Sanjeeb Kumar Sahu[1], Angela Garding[1], Vijay K. Tiwari[1,*]

1. Institute of Molecular Biology (IMB), Mainz, Germany

# These authors contributed equally to this work.

* Author for correspondence (v.tiwari@imb-mainz.de)

Running title: Chromatin response to UV damage.

**Published in the Journal of Cell Science (JCS):**

J Cell Sci. 2015 Dec 1; 128(23): 4380-94. doi: 10.1242/jcs.173633. Epub 2015 Oct 7.

## ABSTRACT

Epigenetic mechanisms determine the access of regulatory factors to DNA during events such as transcription and the DNA damage response. However, the global response of histone modifications and chromatin accessibility to UV exposure remains poorly understood. Here, we report that UV exposure results in a genome-wide reduction in chromatin accessibility, while the distribution of the active regulatory mark H3K27ac undergoes massive reorganization. Genomic loci subjected to epigenetic reprogramming upon UV exposure represent target sites for sequence-specific transcription factors. Most of these are distal regulatory regions, highlighting their importance in the cellular response to UV exposure. Furthermore, UV exposure results in an extensive reorganization of super-enhancers, accompanied by expression changes of associated genes, which may in part contribute to the stress response. Taken together, our study provides the first comprehensive resource for genome-wide chromatin changes upon UV irradiation in relation to gene expression and elucidates new aspects of this relationship.

# INTRODUCTION

Maintenance of genome integrity is essential for cell survival and reproduction as DNA is continuously challenged, for example, by environmental factors like solar UV light. UV light primarily induces DNA lesions, such as cyclobutane pyrimidine dimers (CPDs), pyrimidine 6-4 pyrimidone photoproducts (6-4PPs) and their Dewar isomers, but it can also cause DNA double-strand breaks [1]. Many DNA repair pathways have evolved to correct these diverse types of damages in order to prevent DNA lesions causing pathologies such as skin cancer [2, 3]. The DNA damage response involves recognition of damaged sites and transduction of the signal to effector molecules, which further regulate transcriptional changes, DNA repair, cell-cycle arrest or, if the damages are too severe, cell death [4].

Consequences of DNA lesions, either as a result of UV irradiation or other genotoxic agents, to cell physiology are widely governed by changes in gene expression regulated on various levels [5-8]. UVB has been shown to inhibit initiation of transcription by impeding the binding of RNA polymerase II to the promoters of many transcribed genes [9]. UV can further modulate the action of polymerases through activation of transcription factors such as the tumor suppressor protein TP53 and activator protein 1 (AP1) resulting in gene expression changes [10, 11]. Recent studies have revealed that epigenetic mechanisms also play a role in the transcriptional response following UV irradiation. For example, the components of the ATP-dependent SWI-SNF chromatin remodeling complex BRG1 and BRM (also known as SMARCA4 and SMARCA2, respectively) have been implicated in the transcriptional regulation of UV-induced genes [12, 13]. Furthermore, many posttranslational histone modifications have been shown to be modulated during the cellular response to DNA damage [14-17]. For example, following UV exposure, the histone acetyltransferase p300 (also known as EP300) is recruited close to the MMP-1 promoter, regulating MMP-1 transcription by its catalytic activity [18]. UV exposure induces phosphorylation of histone H3 serine 28 (H3S28p) at promoters of stress-response genes and causes dissociation of histone deacetylase (HDAC) co-repressor complexes that further accompanies enhanced levels of histone acetylation and transcriptional induction of these genes [19, 20]. H3S28 phosphorylation is furthermore involved in the regulation of RNA-polymerase-III-dependent transcription [21]. Given that many long non-coding RNAs (lncRNAs) are

deregulated upon X-ray irradiation and doxorubicin treatment in mammalian fibroblasts, and several lncRNAs have been implicated in the DNA damage response [22-27], it is likely that lncRNAs also play a crucial role in modulating gene expression upon UV irradiation.

In addition to promoters, the importance of distal regulatory regions in regulating gene expression is increasingly being appreciated. Enhancers are among such distal regulatory regions that function to augment the transcription of associated genes [28]. Their active state is characterized by high levels of histone 3 acetylated at lysine 27 (H3K27ac) and chromatin accessibility [29-31]. Recent findings suggest that enhancer regions might also play an essential role in the DNA damage response by recruiting factors such as TP53 [23]. Recently, clusters of putative enhancers (named 'super-enhancers') have been shown to regulate the expression of genes defining cell identity [32]. They have also been found to be deregulated in diseases such as cancer and inflammation [33-38], but have not yet been studied in context of UV exposure.

Chromatin dynamics constitute a critical part of the DNA damage response and many studies have shown recruitment of several ATP-dependent chromatin-remodeling complexes as well as histone-modifying enzymes to damage sites [14, 39, 40]. An early and transient nucleosome destabilization at the site of damage has been shown [41-44], which allows binding of the repair machinery to DNA lesions before chromatin architecture is restored after the repair (the 'prime-repair-restore model') [45]. In this direction, a recent study has shown that, upon UV irradiation, chromatin transiently undergoes de-condensation to allow binding of several repressive factors such as the polycomb complex component (BMI1), the nucleosome-remodeling deacetylase complex (NuRD) as well as the heterochromatin proteins HP1 and HP1-binding protein 3 (HP1BP3) [46]. Another study proposes that during the response to UV irradiation, recruitment of repressive complexes induces condensation of chromatin to protect DNA from further potential damage [47]. Until now, the function and the resultant chromatin changes following recruitment of these complexes remains controversial [48]. Therefore, there is a need to perform a comprehensive investigation into the dynamics of chromatin state following UV irradiation.

Here, we have generated genome-wide datasets to study transcriptome and epigenome changes in response to UV irradiation. We assessed genome-wide changes

in the transcriptome by performing RNA-seq, chromatin accessibility by performing formaldehyde-assisted isolation of regulatory elements (FAIRE)-seq and epigenetic landscape by performing H3K27ac chromatin immunoprecipitation (ChIP)-seq at 6 h after UV irradiation of murine fibroblast cells (NIH3T3). Strikingly, we observed a genome-wide loss of chromatin accessibility affecting all genomic regions. The active histone mark H3K27ac also showed massive reorganization throughout the genome. Many potential regulatory regions harboring open chromatin and H3K27ac undergo changes upon UV exposure and serve as binding sites for sequence-specific transcription factors. Importantly, we also observed a dramatic remodeling of super-enhancers that often accompanied expression changes of associated genes. In general, a large fraction of the UV-induced gene expression changes could be explained by the observed chromatin changes. Our observations also reveal that the chromatin status prior to UV damage might influence the expression and epigenetic changes following UV exposure. Overall, these findings provide the first comprehensive resource revealing global changes in the epigenetic state as well as chromatin accessibility following UV damage and their relation to the UV-induced expression changes.

# RESULTS

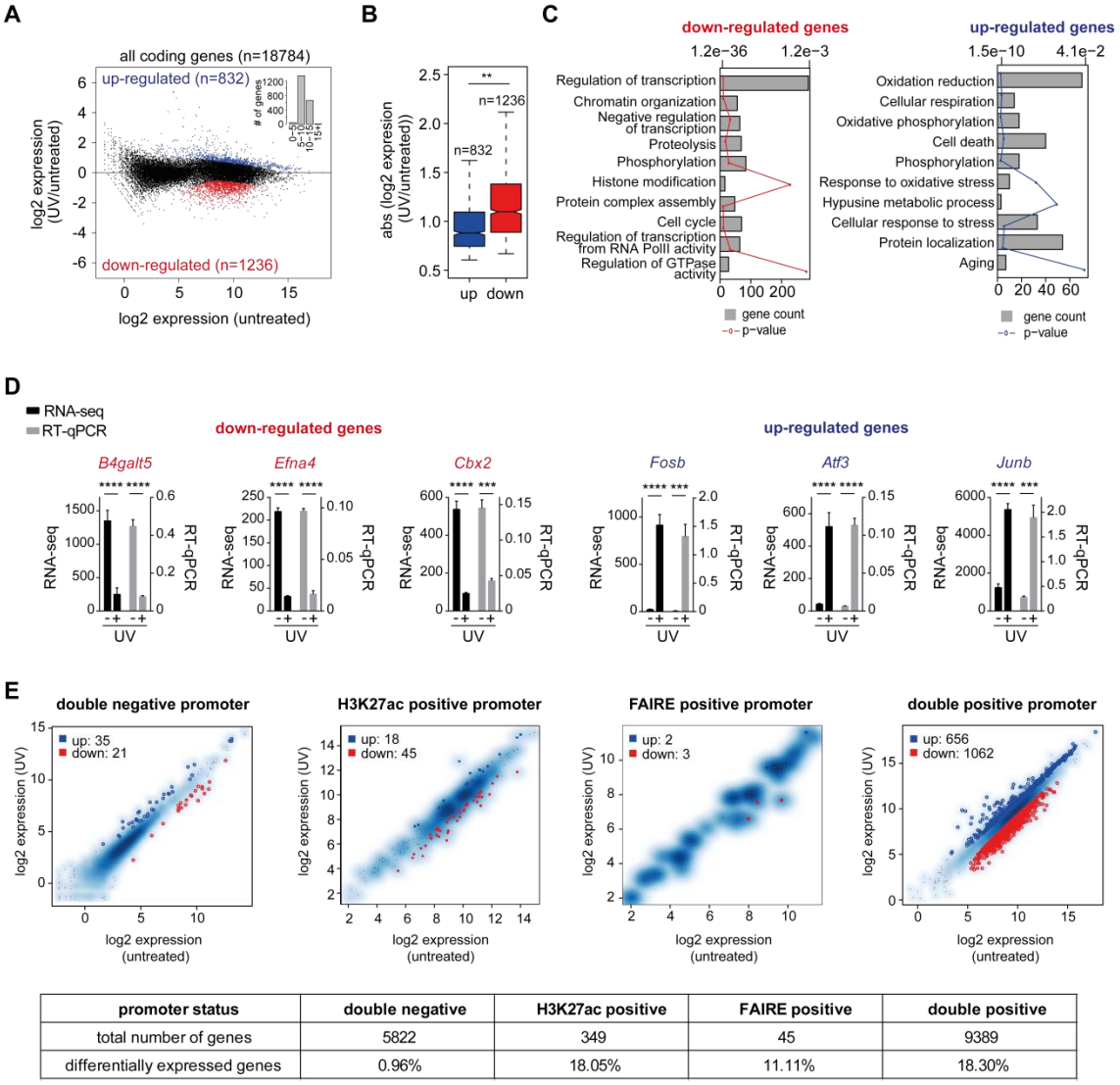
### UV-induced gene expression changes primarily occur at expressed genes exhibiting active chromatin

To identify global transcriptome changes in response to UV, we irradiated murine fibroblast cells (NIH3T3) with UVC and performed genome-wide expression profiling (RNA-seq) on RNA collected after 6 h (Fig. S1A,B). Computational analysis revealed 832 and 1236 genes that were significantly up- or down-regulated in response to UV, respectively (Fig. 1A). Interestingly, both the up- and down-regulated genes were expressed in the untreated condition indicating that active genes are prime responders to UV irradiation (Fig. 1A, inset, Fig. S1C). Importantly, not only the count of down-regulated genes, but also the magnitude of down-regulation was significantly greater compared to that of the up-regulation (Fig. 1B). Gene ontology (GO) enrichment analysis of the down-regulated genes showed a strong enrichment for terms such as regulation of transcription (including genes encoding general transcription factors, mediators and zinc finger proteins) and chromatin organization (Fig. 1C; Table S1). In contrast, the up-regulated genes were involved, for example, in oxidation reduction, cell death and the stress response (Fig. 1C; Table S1). Furthermore, genes in pathways related to translation were found to be up-regulated as exemplified by the up-regulation of ribosomal proteins (e.g. Rpl18 and Rpl17), translation initiation factors (e.g. Eif5b and Eif3g) and elongation factors (e.g. Eef1g and Eef1b2). To further consolidate our findings we also performed Gene Set Enrichment Analysis (GSEA) [49] and cellular component enrichment analysis for up- and down-regulated genes. These results supported the indication of an enhanced translation process in response to UV irradiation (Fig. S1D,E; Table S1). Although we did not observe any increase in the total protein amount 6 h after UV (Fig. S1F), the potential increase in translation of certain proteins might until that time be employed to compensate for the observed transcriptional down-regulation. A number of genes from the AP1 complex (e.g. *Atf3, Fosb* and *Junb*), which is involved in cellular stress response, were found to be strongly up-regulated in response to UV irradiation, whereas, for example, *Cbx2*, a component of the polycomb multiprotein complex involved in chromatin organization, was down-regulated (Fig. 1D) [50-52].

Although long intergenic non-coding RNAs (lincRNA) are crucial regulators of gene expression, to our knowledge there have not been any genome-wide studies exploring the role of lincRNAs in response to UV. Therefore, we decided to explore differentially expressed lincRNAs from our RNA-seq data. Surprisingly, in contrast to the coding genes, differentially expressed lincRNAs were predominantly up-regulated (*n*=86) and only a small proportion showed down-regulation (*n*=12) (Fig. S1G). Interestingly, unlike coding genes, up-regulated lincRNAs were mostly either not expressed or only expressed at low levels in the untreated condition. Previous studies have linked lincRNAs to the transcriptional regulation of nearby genes [53-55]. Correlating the expression changes of lincRNAs to the expression change of the closest gene revealed that genes near to down-regulated lincRNAs were almost exclusively reduced in expression, whereas genes next to up-regulated lincRNAs were also often differentially expressed, but were either up- or down-regulated (Fig. S1H). Interestingly, some up-regulated lincRNAs occurred next to established stress response genes such as *Btg2, Egr3, Ier2* and *Ier3* (Fig. S1I) [56-59].

We next investigated whether the chromatin state of gene promoters and the distal regulatory elements could predict the transcriptional behavior of genes in response to UV irradiation. H3K27ac is an established mark of active promoters and enhancers. Therefore, we performed ChIP assays in NIH3T3 cells using H3K27ac-specific antibody followed by next-generation sequencing (ChIP-seq) of recovered genomic DNA. To further assess whether chromatin accessibility predicts changes in gene expression, we also performed a FAIRE assay and sequenced the isolated genomic DNA (FAIRE-seq) [60]. Computational analysis revealed that the genes that changed expression after UV exposure were mainly expressed genes whose promoters harbored H3K27ac and/or that were accessible [denoted double-positive (promoters enriched for both H3K27ac and FAIRE), or as H3K27ac-positive or FAIRE-positive] compared to genes whose promoters had none of these features (denoted double-negative) (Fig. 1E). In general, as expected, genes that showed neither H3K27ac nor were accessible were mostly not expressed; however, genes in this category who changed their expression upon UV were expressed in the untreated condition (Fig. 1E, left panel). This finding is consistent with our above observation that active genes are primarily differentially expressed upon UV irradiation (Fig. 1A). Taken together, these observations suggest that gene

expression status and the prior chromatin state at promoters determine the UV-induced gene expression response.




| promoter status | double negative | H3K27ac positive | FAIRE positive | double positive |
|---|---|---|---|---|
| total number of genes | 5822 | 349 | 45 | 9389 |
| differentially expressed genes | 0.96% | 18.05% | 11.11% | 18.30% |

***Figure 1: Active genes are prime responders to UV exposure.***

(A) MA plot showing expression changes of protein-coding genes 6 h after UV treatment of NIH3T3 cells revealed by RNA-seq analysis ($n$=3). The inset shows the number of differentially expressed genes ($y$-axis) in four bins of log2 expression ($x$-axis). (B) Box plots showing distribution of the absolute values of the log2 expression ratio of differentially expressed genes. The box represents the 25–75th percentiles, and the median is indicated. The whiskers show 1.5 times the interquartile range (IQR) added to the 75th percentile (upper whisker) or subtracted from the 25th percentile (lower whisker). The notches represent median±1.57×IQR/($n^{0.5}$). **$P$<2.2×$10^{-6}$ (Wilcoxon rank-sum test). (C) Bar plot showing enrichment of biological processes for up-regulated (right panel) and down-regulated (left panel) genes. Bar length represents the number of genes (primary $x$-axis), and the $P$-value is shown by line graph (upper $x$-axis). (D) Independent validation of detected gene expression changes 6 h after UV treatment in NIH3T3 cells by RT-qPCR analysis. Mean±s.e.m. expression levels detected by RNA sequencing are plotted as normalized read counts (left $y$-axis; $n$=3) and mean±s.e.m. mRNA abundance measured by RT-qPCR are plotted normalized to *Ctcf* (ΔCT) (right $y$-axis; $n$=4). ***$P$<0.001; ****$P$<0.0001 (unpaired $t$-test for RT-qPCR results and as determined by DESeq for the RNA-seq data). (E) Scatterplots showing gene expression changes for genes harboring different epigenetic features at their promoters in the untreated condition with highlighted differential expressed genes identified by DESeq (blue, up-regulated; red, down-regulated). Genes whose promoters show no enrichment for FAIRE or H3K27ac (double negative, first panel), only enrichment for H3K27ac (H3K27ac positive, second panel), only enrichment for FAIRE (FAIRE positive, third panel) and enrichment for FAIRE and H3K27ac (double positive, fourth panel) are displayed. The table provides the total number of genes and the percentage of differentially expressed genes in each category.

**UV irradiation results in a genome-wide reduction of chromatin accessibility**

We next investigated whether UV treatment influences chromatin accessibility. Towards this, we performed FAIRE-seq 6 h after UV treatment in biological duplicates and compared it to FAIRE-seq performed on untreated cells to reveal genome-wide changes in chromatin accessibility following UV irradiation. Quality control analyses revealed a good correlation between replicates and also confirmed that analysis with individual replicates showed the same biological interpretations (Fig. S2A-H).

Analyzing the total accessible genome showed an approximately one-third reduction upon UV irradiation (Fig. 2A). The distribution of accessible sites in the genome revealed a loss in accessibility following UV treatment at all genomic features; however, the reduction was much stronger in non-promoter regions (Fig. 2B; Fig. S2I). Furthermore, enrichments normalized to the genomic feature size revealed that promoters were more enriched for accessible sites compared to other genomic regions (Fig. S2J,K). Our analysis revealed that 16,782 genomic regions lost accessibility following UV treatment (denoted untreated unique), whereas only 2206 acquired open chromatin (denoted UV unique) (Fig. 2C). To further investigate the chromatin compaction, we next analyzed the FAIRE peak width in both conditions, as this may reflect the size of the accessible DNA available for the binding of proteins such as transcription factors. Interestingly, regions that gained accessibility were of relatively smaller width compared to accessible sites that were lost following UV treatment, with the exception of exonic sites (Fig. 2C,D). We found ~40-fold more regions losing accessibility compared to regions gaining accessibility, reflecting the finding that chromatin compaction was much more pronounced upon UV exposure (Fig. 2E, left panel). Similar observations were made when analyzing promoters and intergenic regions separately (Fig. 2E, right panels). In order to validate FAIRE-seq detected chromatin accessibility changes at promoters as well as intergenic regions, we used independent methods to assess chromatin accessibility changes. First, accessibility changes measured by FAIRE-quantitative PCR (FAIRE-qPCR) at earlier time points after UV irradiation revealed dynamic changes of chromatin accessibility over time arguing against any technical bias due to UV induced crosslinking or that DNA damage is preventing effective isolation of open regions (Fig. S2L). Furthermore, we also employed an assay for transposase-accessible chromatin (ATAC) on untreated and UV-

irradiated cells as an independent measure of chromatin accessibility, revealing that the FAIRE-seq results were largely reproducible (Fig. S2M) [61-63]. To further augment these results, nucleosome occupancy was investigated by assessing total H3 enrichments at such sites by ChIP-qPCR. As expected, the results showed the opposite dynamics to the accessibility changes at the same loci (Fig. 2F; Fig. S2L). This argues that indeed the observed global loss of accessibility upon UV treatment led to chromatin condensation at these loci with an increased amount of nucleosomes.

Intrigued by the largely reproducible chromatin accessibility response upon UV treatment, we next asked whether these changes were caused by the DNA-damage response occurring at these loci. Because it has been reported that UV damage induces γH2AX at damaged sites within 6 h in human fibroblasts [64] and we also observed a global gain of γH2AX 6 h following UV treatment (Fig. S2N), we monitored damaged DNA regions by performing γH2AX ChIP in untreated and UV irradiated NIH3T3 cells. We measured γH2AX enrichments at regions that were non-accessible or accessible in the untreated condition and either changed or did not change their accessibility at 6 h after UV exposure. These results indicate that UV-induced γH2AX occurrence is independent of the accessibility status of the regions before UV treatment and, furthermore, chromatin accessibility changes seem to be independent from the γH2AX occurrence (Fig. S2O,P). We also investigated whether the sites responding or non-responding in accessibility upon UV irradiation had different chromatin features prior to the treatment by using published ChIP-seq datasets for different histone modifications, histone variants and phosphorylated RNA polymerase generated from ChIP experiments performed on untreated NIH3T3 cells. The analyses revealed that certain chromatin features are significantly different between these classes indicating that the chromatin status prior to UV irradiation impacts on the chromatin response to UV with respect to changes in accessibility (Fig. S2Q).

The observation of a major chromatin accessibility loss is further consistent with our findings that UV-induced transcriptional down-regulation is much more pronounced compared to up-regulation (Fig. 1B). Prompted by these findings, we next investigated the correlation of chromatin accessibility at promoters with the expression level of the genes in the untreated and UV conditions and observed a good positive correlation (Fig. S2R). We then asked whether changes in the accessibility of promoters and distal

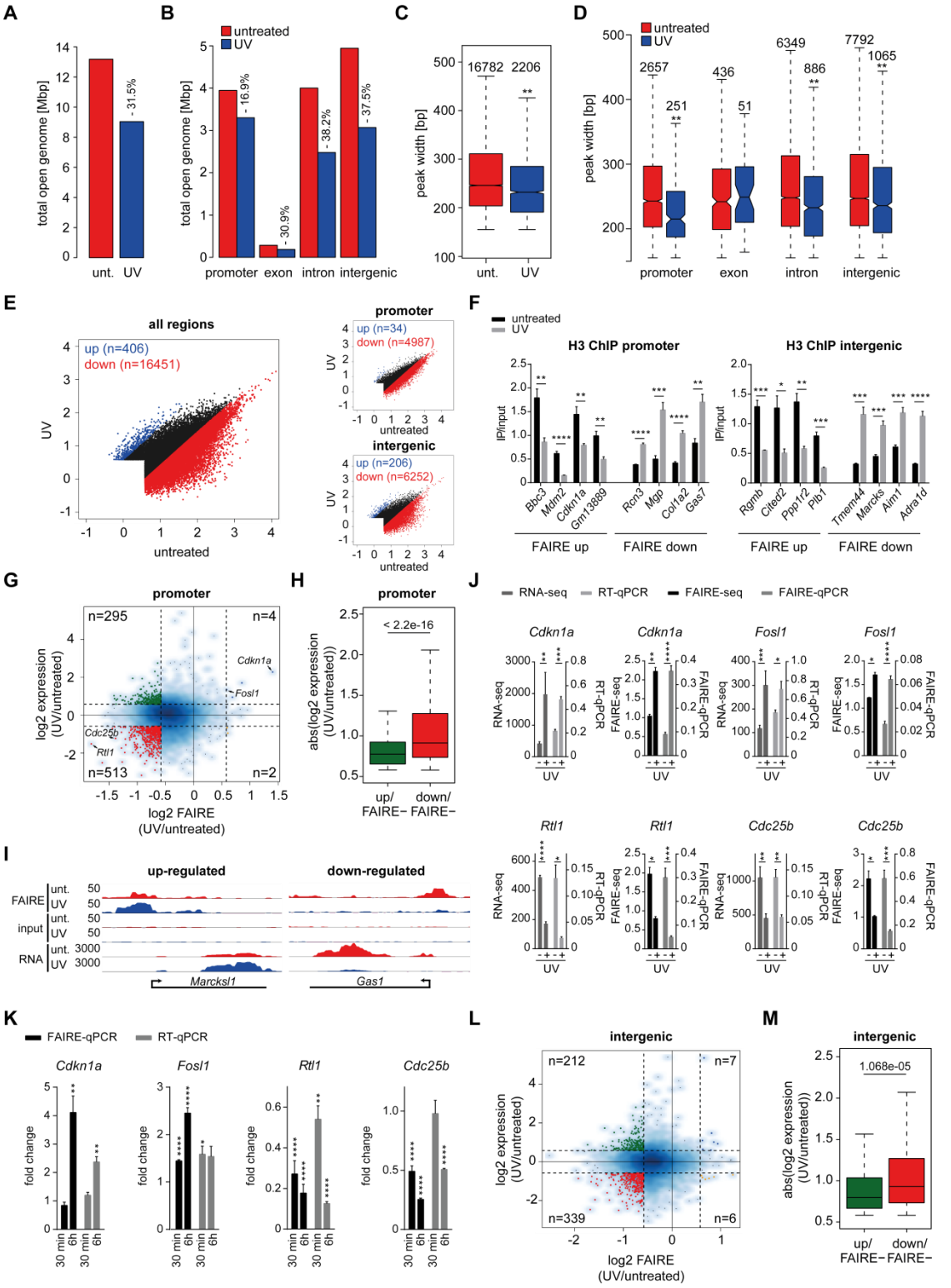



***Figure 2: UV irradiation results in a genome-wide chromatin compaction.***

(A) Bar plot showing total open genome size in Mbp of untreated and UV-treated NIH3T3 cells. Total open genome size was calculated as the sum of the widths of all enriched FAIRE peaks. The percentage displays the loss of accessible chromatin upon UV. (B) Same as in A, but for individual genome features. (C) Box plots showing distribution of peak width for unique open regions in untreated and UV-treated cells. The box represents the 25–75th percentiles, and the median is indicated. The whiskers show 1.5 times the interquartile range (IQR) added to the 75th percentile (upper whisker) or subtracted from the 25th percentile (lower whisker). The notches represent median±1.57×IQR/($n^{0.5}$). The numbers above each box plot display the counts of unique peaks for each condition. $^{**}P<2.2\times10^{-16}$ (between peak width in untreated and UV condition; Wilcoxon rank-sum test with a cut-off of 0.01). (D) Same as in C, but for different genomic regions. Significant changes upon UV are indicated above the blue box plots. $^{**}P<0.01$ (promoters, $1.581\times10^{-9}$; exons, 0.774; introns, $1.762\times10^{-15}$; intergenic regions, $8.188\times10^{-7}$). (E) Scatter plot showing the dynamics of FAIRE enrichment changes between untreated and UV conditions for all genomic regions together (left panel), promoter regions (right, upper panel) and intergenic regions

(right, lower panel). Peaks with at least 1.5-fold enrichment change following UV irradiation are highlighted (gain in blue, loss in red). The numbers of up- and down-regulated peaks are displayed in the plots. (F) H3 ChIP-qPCR results for promoter (left) and intergenic (right) regions that showed gain or loss in the FAIRE-seq analysis (indicated below). The results are plotted as immunoprecipitated DNA/input DNA (IP/input) ($n$=4, mean±s.e.m.). *$P$<0.05; **$P$<0.01; ***$P$<0.001; ****$P$<0.0001 (unpaired $t$-test). (G) Scatter plot showing relationship between chromatin accessibility at promoters and expression of the corresponding genes. Quadrants indicated by dashed lines show changes in expression ($y$-axis) and accessibility ($x$-axis) greater than 1.5-fold and numbers indicate the amount of genes falling in each quadrant. (H) Boxplots summarizing the absolute values of expression changes for genes belonging to the green and red categories in G. Significance was determined using a Wilcoxon rank-sum test for non-normal distributions with a cut-off of 0.01. Up and down indicate up- or down-regulation of the gene expression 6 h following UV exposure, FAIRE− indicates loss of FAIRE-seq enrichment 6 h after UV treatment. (I) Genome browser plots for genes showing either concomitant increase or decrease in expression and FAIRE enrichment at their promoters. Each track is shown from 0 to the indicated height. (J) Validation of RNA-seq and FAIRE-seq results by independent RT-qPCR and FAIRE-qPCR for genes selected from the red and blue categories in G (indicated by arrows). Mean±s.e.m. expression measured by RNA-seq analysis is plotted as normalized read counts on the left $y$-axis ($n$=3) and the mean±s.e.m. mRNA abundance determined by RT-qPCR is plotted as ΔCT on the right $y$-axis, normalized to *Ctcf* ($n$=3). As determined by qPCR, FAIRE enrichment above input is plotted normalized to FAIRE enrichment at the *Ctcf* promoter at the right $y$-axis ($n$=4, error bars represent s.e.m.), whereas FAIRE-seq enrichment normalized to input is plotted at the left $y$-axis ($n$=2). *$P$<0.05; **$P$<0.01; ***$P$<0.001; ****$P$<0.0001 (unpaired $t$-test for RT-qPCR, FAIRE-qPCR and FAIRE-seq, and as determined by DESeq for RNA-seq data). (K) Relative expression determined by RT-qPCR is plotted as the fold change for NIH3T3 cells 30 min and 6 h after UV treatment relative to the untreated condition (ΔΔCT normalized to *Ctcf*, $n$=3, error bars represent s.e.m.). FAIRE enrichment above input determined by FAIRE-qPCR at same time points was normalized to FAIRE enrichment at the *Ctcf* promoter and then plotted as fold change to untreated condition ($n$=4, error bars represent s.e.m.). *$P$<0.05; **$P$<0.01; ***$P$<0.001; ****$P$<0.0001 (unpaired $t$-test). (L) Same plot as in G, but FAIRE enrichments are shown for peaks falling in intergenic regions and plotted against the expression changes of the closest gene. (M) Same plot as in H, but for the red and green quadrants in L.

regions are in accordance with alterations in expression of the associated genes following UV treatment. This analysis showed that the majority of genes whose promoter lost accessibility were also down-regulated ($n$=513, red) (Fig. 2G-I). These genes were predominantly those involved in regulation of transcription and chromatin organization (Fig. S2S; Table S1). We observed another set of genes that lost accessibility at promoters but slightly gained transcription *(n*=295, green) (Fig. 2G,H). There were four genes found to gain accessibility as well as expression (*Cdkn1a, Fosl1, Marcksl1* and *Olfr1226*). Among them were typical DNA-damage response genes like the proliferation inhibitor *Cdkn1a* and the AP1 component *Fosl1*, whose changes in expression and promoter accessibility were validated by independent reverse transcription quantitative real-time PCR (RT-qPCR) and FAIRE-qPCR (Fig. 2G,J). Moreover, analysis of expression and accessibility changes that had occurred by 30 min after UV irradiation revealed that accessibility changes either precede expression changes or occur at the same time (Fig. 2K). A comparison of accessibility changes at potential distal regulatory regions with the expression changes of the closest gene revealed patterns comparable to those seen for promoters. Most genes whose distal regions lost accessibility were down-regulated ($n$=339, red) following UV treatment (Fig. 2L,M). Another set lost accessibility at intergenic regions but showed a minor gain in expression ($n$=212, green) (Fig. 2L,M).

Overall, these findings suggest that UV exposure causes a genome-wide loss of chromatin accessibility that largely accompanies transcriptional down-regulation of the

associated genes. However, despite a genome-wide reduction in transcriptional competence, distinct genomic loci associated with the stress response are selectively kept accessible to allow their expression.

**UV exposure causes a global reprogramming of the H3K27ac mark**

Following our observations of a global change in chromatin accessibility, we next investigated the effect of UV on a specific histone modification that defines distinct transcriptional states. To this end, we selected H3K27ac, an established marker of active promoters and enhancers, and performed ChIP-seq for H3K27ac in untreated and UV treated NIH3T3 cells. Quality control analyses with respect to antibody specificity, chromatin shearing, alignment statistics and replicate correlation were performed (Fig. S3A-I). Whereas the total fraction of the genome and of various genomic regions enriched for H3K27ac stayed largely unchanged following UV treatment (Fig. 3A,B; Fig. S3J), the enrichment for H3K27ac at these genomic loci underwent a dramatic reorganization. Sites that gained H3K27ac (*n*=12,241) showed a significantly higher peak width compared to regions that lost H3K27ac following UV irradiation (*n*=11,441) (Fig. 3C). Furthermore, this observation pertains also to regions distributed throughout various genomic locations (Fig. S3K). The enrichment for H3K27ac at different genomic regions mainly showed an increase in H3K27ac enrichment after UV (Fig. 3D; Fig. S3L). We observed massive changes with respect to H3K27ac enrichment after UV treatment with nearly equal numbers of up- and down-regulated sites in response to UV irradiation (22,869 up and 20,103 down) (Fig. 3E, left panel). This observation further holds true when looking at the enrichment level changes either at promoters or at peaks falling in intergenic regions (Fig. 3E, right panels).

Given that H3K27ac enrichment levels at promoters showed a high correlation to the expression status of a gene in untreated as well as in UV irradiated cells (Fig. S3M), we next compared UV-induced alterations in H3K27ac levels at promoters to changes in gene expression. This revealed that most promoters losing H3K27ac also showed transcriptional down-regulation (*n*=390, red) (Fig. 3F-H). These genes were mainly enriched for processes such as regulation of transcription, chromatin organization or regulation of kinase activity (Fig. S3N; Table S1). Furthermore, a large number of genes whose promoters gained H3K27ac were also associated with increased transcription

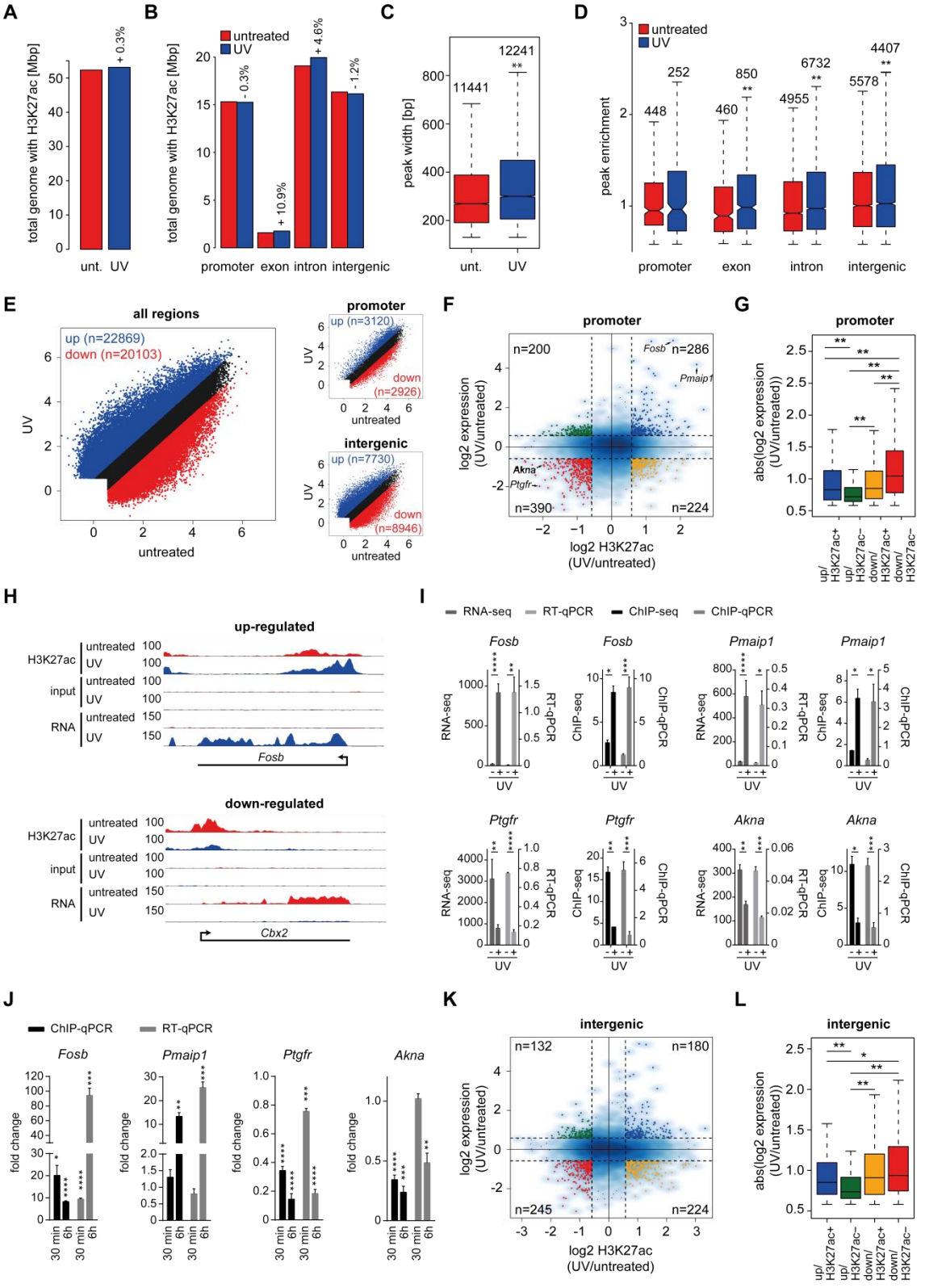



***Figure 3: UV exposure causes massive remodeling of H3K27ac landscape.***

(A) Bar plot showing total genome size in Mbp for the genome occupied by H3K27ac in untreated and UV-treated NIH3T3 cells calculated as the sum of the widths of all enriched H3K27ac peaks. The percentage displays the change of total genome size occupied by H3K27ac between UV-irradiated and untreated cells. (B) Same as in A, but for different genomic regions. (C) Box plot showing distribution of peak width for unique H3K27ac regions in untreated and UV-treated cells. The numbers of unique peaks for each condition are indicated on top. The box represents the 25–75th percentiles, and the median is indicated. The whiskers show 1.5 times the interquartile range (IQR) added to the 75th percentile (upper whisker) or subtracted from the 25th percentile (lower whisker). The notches represent median±1.57×IQR/($n^{0.5}$). $^{**}P<2.2\times10^{-16}$ (Wilcoxon rank-sum test for non-normal distributions). (D) Distribution of H3K27ac peak enrichment in four genomic regions for untreated and UV-treated cells at unique sites (i.e. H3K27ac is at least 1.5-fold enriched above

background either in the untreated or UV condition). Numbers above box plots indicate counts of unique peaks for this condition. $^{**}P<0.01$ (Wilcoxon rank-sum tests for non-normal distribution; promoters, 0.7776; exons, $1.281\times10^{-6}$; introns, $<2.2\times10^{-16}$; intergenic regions, $<2.2\times10^{-16}$). (E) Scatter plot showing dynamics of H3K27ac enrichment changes between untreated and UV conditions for all genomic regions together (left panel), for promoter regions (right, upper panel) and for intergenic regions (right, lower panel). Peaks with at least 1.5-fold enrichment change upon UV treatment are highlighted (gain in blue, loss in red). The numbers of peaks falling in each of these two categories are displayed on the plots. (F) Scatter plot showing relationship between H3K27ac enrichment changes at promoters (*x*-axis) and gene expression changes (*y*-axis). Quadrants indicated by dashed lines show changes in expression and H3K27ac enrichment greater than 1.5-fold, and numbers indicate the counts of genes falling in each category. (G) Boxplots summarizing the absolute values of expression changes for genes highlighted with the respective color in the four categories in F. Up/down indicates up-/down-regulation of the gene expression, H3K27ac−/+ indicates H3K27ac enrichment loss/ gain 6 h following UV treatment. Significance was determined using a Wilcoxon rank-sum test for non-normal distributions with a cut-off of 0.01. $^{**}P<0.01$ (blue versus green, $6.477\times10^{-10}$; blue versus red, $4.718\times10^{-12}$; green versus orange, $1.261\times10^{-11}$; green versus red, $<2.2\times10^{-16}$; orange versus red, $1.059\times10^{-13}$). (H) Genome browser plots of genes showing either concomitant increase or decrease in expression as well as H3K27ac enrichment at their promoters. Each track is shown from 0 to the indicated height. (I) Validation of RNA-seq and H3K27ac ChIP-seq results by independent RT-qPCR and ChIP-qPCR for genes selected from the red and blue categories of F (indicated by arrows). Mean±s.e.m. expression from RNA-seq is plotted as normalized read counts ($n=3$) on the left *y*-axis and mean±s.e.m. mRNA abundance determined by RT-qPCR is plotted as ΔCT, normalized to *Ctcf* on the right *y*-axis ($n=3$). ChIP-seq enrichments normalized to the input are plotted on the left *y*-axis ($n=2$), whereas ChIP-qPCR enrichment is plotted as immunoprecipitated DNA/input DNA (IP/input) on the right *y*-axis ($n=4$). $^{*}P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$; $^{****}P<0.0001$ (unpaired *t*-test for RT-qPCR, ChIP-qPCR and ChIP-seq, and as determined by DESeq for RNA-seq data). (J) Relative expression determined by RT-qPCR is plotted as fold change for NIH3T3 cells 30 min and 6 h after UV treatment relative to the untreated condition (ΔCT normalized to *Ctcf*, $n=3$, error bars represent s.e.m.). H3K27ac enrichments for the promoter of these genes were determined at the same stages by ChIP-qPCR. These results are normalized to input and then plotted as fold change to untreated condition ($n=3$, error bars represent s.e.m.). $^{*}P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$; $^{****}P<0.0001$ (unpaired *t*-test). (K) Same as in F, but H3K27ac enrichments at intergenic sites are plotted against the expression changes of the closest gene. (L) Boxplots summarize the absolute values of changes in expression associated with each of the four categories highlighted in K (similar to G). Significance was determined using a Wilcoxon rank-sum test for non-normal distributions with a cut-off of 0.01. $^{*}P<0.05$; $^{**}P<0.01$ (blue versus green, $1.4\times10^{-4}$; blue versus red, $3.6\times10^{-2}$; green versus orange, $2.464\times10^{-6}$; green versus red, $1.942\times10^{-9}$).

($n=286$, blue) (Fig. 3F-H) and they had GO terms such as oxidation reduction, cell death, regulation of transcription and response to DNA damage stimulus (Fig. S3O; Table S1). For the genes that showed gain of H3K27ac but loss of expression and vice versa, the magnitudes of expression changes were less or similar to the other two classes of genes (Fig. 3G). We further validated several of these changes by independent RT-qPCRs and ChIP-qPCRs (Fig. 3F,I). Similar analysis at an earlier time point indicates that changes in H3K27ac enrichment either precede or correlate with the timing of expression changes (Fig. 3J).

Distal regions marked with H3K27ac are known to often represent active enhancers [29]. To investigate whether UV-induced transcriptional changes also involve reprogramming of the enhancer landscape, we performed a comparative analysis of changes in distal H3K27ac sites with changes in expression for the nearest gene following UV treatment. A large fraction of the distal regions that showed loss of H3K27ac was associated with significant transcriptional down-regulation of the nearest gene ($n=245$, red) (Fig. 3K,L). This class was enriched for genes, for example, involved in regulation of transcription, chromatin organization and phosphorylation (Fig. S3P; Table S1). Similarly, in many distal regions, gain of H3K27ac was associated with an increase in expression of the nearest gene ($n=180$, blue) (Fig. 3K,L). This fraction contains many genes associated with the cellular response to stress, regulation of cell cycle, circulatory

system process and regulation of cell death (Fig. S3Q; Table S1) and includes well-known UV-induced genes like *Fosb* or *Egr1*. For a minor fraction of distal regions, loss of H3K27ac was also associated with an increase in expression (*n*=132, green) (Fig. 3K,L). Surprisingly, we also observed a fraction that showed gain of H3K27ac in intergenic regions but down-regulation in the expression of the nearest gene (*n*=224, yellow) (Fig. 3K,L). These contradicting associations of expression and H3K27ac might be due to a limitation in assigning enhancers to their target genes. This goes along with previous suggestions that enhancer-nearest gene pairing only holds true for ~40% of genes [65-67]. However, as such contradictory findings were also observed when correlating H3K27ac enrichment changes at promoters to gene expression changes, it is likely that other chromatin features and regulatory mechanisms might be more important in determining the transcriptional state of these genes that we are unable to capture here or that the expression status of these genes is influenced by post-transcriptional mechanisms.

Overall, our observations suggest that UV exposure causes a genome-wide reorganization of the H3K27ac mark at regulatory elements such as promoters and enhancers, which underlie expression changes of crucial genes during the stress response.

**Accessible distal enhancers are remodeled following UV treatment**

We next investigated the dynamics of H3K27ac-enriched sites that also exhibited accessible chromatin (i.e. double-positive for FAIRE and H3K27ac), assuming that such regions represent sites of intense gene-regulatory activity and harbor a plethora of regulatory elements including transcription-factor-binding sites. The total count of these regions was drastically reduced following UV treatment (Fig. 4A; Fig. S4A). Overlap of untreated and UV double-positive sites revealed that 12,414 of such regions were lost following UV irradiation, whereas only 1912 were gained (Fig. 4B). This substantial loss of double-positive regulatory elements is in line with the loss of chromatin accessibility (Fig. 2A-E). We observed that these sites showed overall loss of accessibility while acquiring H3K27ac following UV treatment (Fig. S4B). The loss of FAIRE enrichment is observed at the uniquely occurring (i.e. enriched only in the untreated sample or the UV-treated sample) as well as the double-positive sites shared

in both the untreated and UV-treated cells ('common' sites) upon UV treatment (Fig. S4C), whereas the overall gain of H3K27ac enrichment after UV arose from the double-positive sites common between the untreated and UV-treated cells (Fig. S4D). It is interesting to note that although the common sites lost FAIRE enrichment, they displayed a substantial gain in H3K27ac enrichment. Untreated-only double-positive sites mostly showed loss of H3K27ac enrichment following UV exposure, whereas UV-only double-positive sites showed a nearly exclusive gain of H3K27ac enrichment (Fig. S4E,F). Such loss and gain was even more prominent for peaks falling in double-positive intergenic regions (Fig. S4G,H). The untreated-only, double-positive peaks did not gain accessibility, whereas a majority of UV-only double-positive peaks gained chromatin openness (Fig. S4I-L).

Analysis of the genomic distribution of double-positive sites revealed, interestingly, that most of the common double-positive sites occurred at promoters, whereas the unique sites were mainly enriched at introns or intergenic regions (Fig. 4C). Inspection of the genes nearest to common, untreated-only and UV-only peaks revealed a substantial number of genes that uniquely gained or lost double-positive sites (Fig. S4M). Genes next to common peaks were mainly enriched in GO terms for housekeeping functions like RNA processing, protein localization, cell cycle regulation, chromatin organization, ribosome biogenesis, regulation of transcription and DNA repair (Fig. 4D; Table S1). Genes harboring double-positive sites only in the untreated condition include those involved in transport of metabolites and chromatin organization (Fig. 4E; Fig. S4N; Table S1). Genes close to double-positive UV-only peaks were enriched for GO term categories like vitamin metabolic processes, metal ion or protein transport, phosphorylation and negative regulation of transcription (Fig. 4F; Table S1). These findings are in line with the results of our transcriptome analysis indicating that cells in general reduce transcription and attempt to utilize available resources following UV exposure.

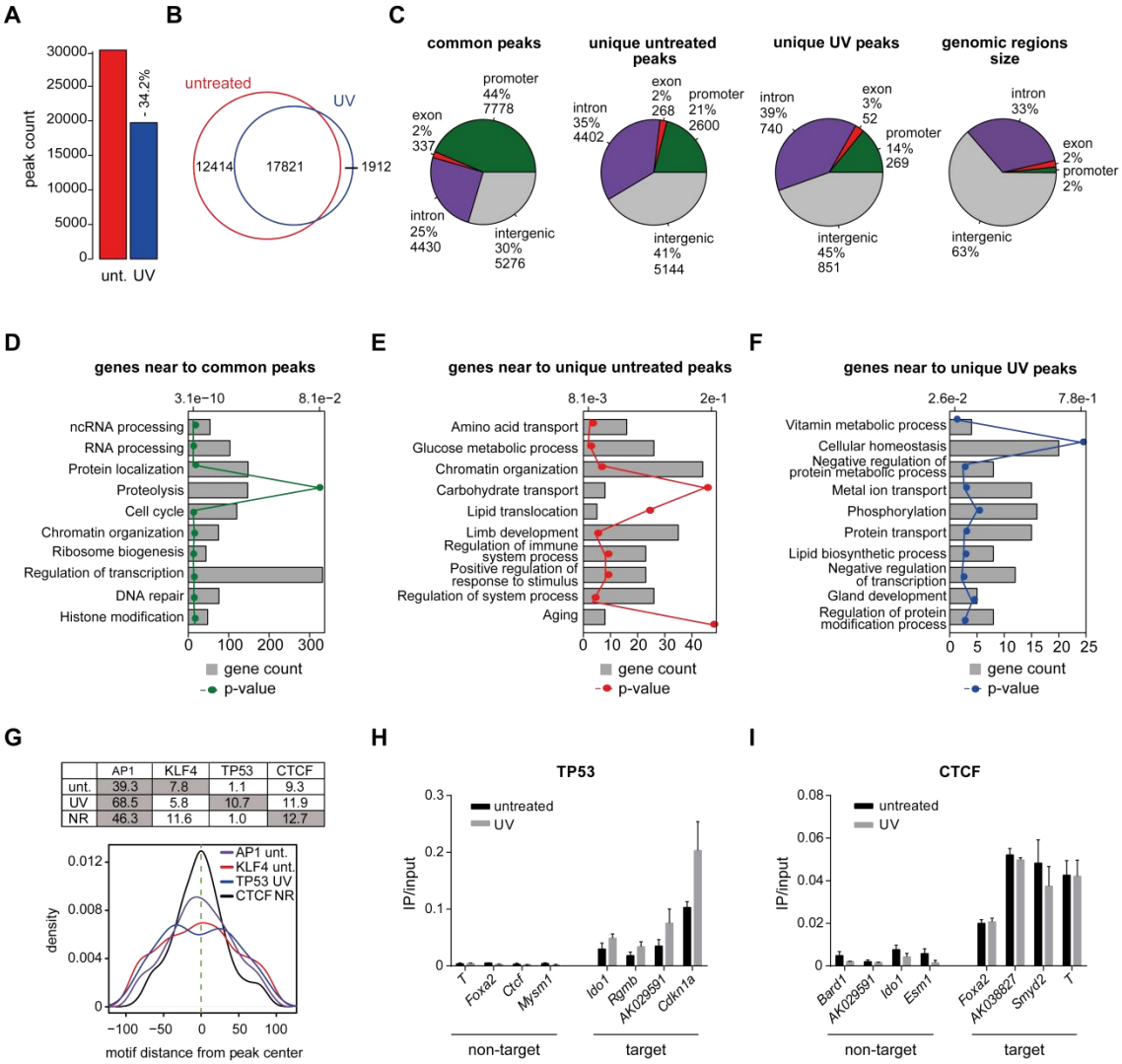



***Figure 4: Accessible distal enhancers are remodeled following UV exposure.***

(A) Bar plot showing the number of H3K27ac and FAIRE double-positive sites in untreated and UV treated cells. The FAIRE peak was considered to be overlapping with the H3K27ac peak if there was an H3K27ac peak within a 500-bp distance from the FAIRE peak summit. (B) Venn diagram showing overlap of double-positive peaks in untreated and UV conditions. (C) Pie charts showing the genomic distribution of peaks that were found in both conditions (common peaks), uniquely in untreated cells (unique untreated peaks) or only in UV-treated cells (unique UV peaks). For comparison, a pie chart is shown displaying the partitioning of the NIH3T3 genome. (D) Bar plot showing enrichment of biological processes for genes near to common peaks. The bars length represent the number of genes (lower *x*-axis), whereas the *P*-value is shown as a line graph with respect to the upper *x*-axis. (E) Same as in D but for genes near to unique untreated peaks. (F). Same as D, but for genes near to unique UV peaks. (G) The table shows the percentage of double-positive regions found in each set of sites (unt., unique untreated peaks; UV, unique UV peaks; NR, not responding, double-positive peaks at least 1.5-fold enriched in untreated and UV samples and not changing more than 1.25-fold after UV exposure) containing the AP1, KLF4, TP53 or CTCF motif (upper panel). The conditions in which the motif was predicted to be enriched by Pscan with respect to the local and global background are highlighted in gray. The density plot shows the average distribution of the motif position relative to the center of FAIRE peaks at the double-positive sites (lower panel). (H) ChIP-qPCR for TP53 in untreated and NIH3T3 cells 6 h after UV treatment for non-target genes as well as potential target genes identified in the UV unique double-positive class. The results are plotted as immunoprecipitated DNA/input DNA (IP/input) ($n$=3; error bars represent s.e.m.). (I) ChIP-qPCR for CTCF in untreated and NIH3T3 cells 6 h after UV exposure for non-target genes as well as potential target genes identified in the common double-positive class (NR). The results are plotted as IP/input ($n$=3; error bars represent s.e.m.).

We next investigated whether untreated-only and UV-only double-positive regions were enriched for transcription-factor-binding motifs. Together with these two sets, we also used a control set of double-positive regions that are enriched in H3K27ac and accessibility in untreated and UV-irradiated cells. Motif analysis led to two levels of motif enrichment information: motifs enriched with respect to the local background (local enrichment, mainly reflecting motifs enriched near peak centers) and motifs enriched with respect to the global background (global enrichment, representing motifs found in surrounding regions). Interestingly, the motif enrichment analysis revealed sites for the same set of factors (FOSL2, FOS, JUNB; all AP1 components), which were found to be up-regulated in their expression upon UV treatment, upstream of the peak center in all three investigated sets of peaks (Fig. 4G; Table S2). However, there were small positional differences with respect to the peak center of the double-positive sites (Table S2). Regions downstream of the untreated-only peaks contained motifs for factors such as TFAP2A, KLF4, SP2 and EGR1 (Table S2). These factors are related to histone deacetylation (SP2) [68], repression of TP53 (KLF4) [69], cell growth, apoptosis and DNA damage (SP1) and repression of apoptosis (EGR1) [70-72]. The regions downstream of the UV-unique motifs showed a high enrichment for the TP53 motif (Table S2). TP53 has been shown to mediate UV-induced global chromatin changes [44]. These UV-unique double-positive regions harboring TP53 motifs were further validated using ChIP-qPCR (Fig. 4H). Interestingly, these results showed that these sites might already be targeted by TP53 in untreated conditions. These findings are in line with a recent genome-wide analysis of TP53 binding, revealing TP53 to be found predominantly within enhancers and showing that these sites are already poised in the untreated conditions by binding of TP53, which then becomes functional upon phosphorylation in response to DNA damage signals [23]. The control set of non-changing peaks showed high local enrichment for CTCF and such CTCF binding at target sites in the non-changing class was validated by ChIP-qPCRs (Fig. 4I). Interestingly, these CTCF target sites did not show any changes in CTCF enrichment following UV, suggesting that binding of CTCF at these sites might protect them from UV-induced chromatin changes. Taken together, these results suggest that there is epigenetic remodeling of regulatory elements upon UV exposure, possibly involving the action of sequence-specific transcription factors, to either mediate or prevent changes following UV treatment.

**UV exposure results in a dramatic reorganization of super-enhancers**

Prompted by our observations that UV treatment results in genome-wide epigenetic reprogramming of regulatory elements, we next investigated whether the UV response also involves the reorganization of H3K27ac clusters called super-enhancers. Using our H3K27ac ChIP-seq data, we first identified such clusters of super-enhancers and then, using a publically available ChIP-seq dataset generated from ChIP performed on untreated NIH3T3 cells for histone 3 mono-methylated at lysine 4 (H3K4me1), confirmed their enrichment for H3K4me1 (Fig. S4O-Q) [73]. Interestingly, whereas the total number of super-enhancers hardly changed, UV treatment led to dramatic reorganization of these elements in which 360 genes lost and 337 genes gained super-enhancers (Fig. 5A,B). Furthermore, the gained super-enhancers were located closer to transcription start sites than those that were lost following UV treatment (Fig. 5C). Super-enhancers were found more frequently in introns and less frequently in intergenic regions after UV treatment (Fig. 5D,E). GO analysis of the nearest genes to untreated-only super-enhancers revealed that these genes were mainly related to cell morphogenesis, cell motion and cell proliferation (Fig. 5F; Table S1). Super-enhancers occurring uniquely in UV-treated cells were in vicinity of genes involved in cell cycle, apoptosis and macromolecule catabolic processes (Fig. 5G; Table S1). Furthermore, they were linked to kinase activity (Fig. 5G) as well as to various signaling pathways (Fig. 5H; Table S1). Further analyses revealed that unique genes near to untreated-unique super-enhancers significantly lost their expression upon UV irradiation, indicating that these super-enhancers contribute towards their higher expression in the untreated condition (Fig. 5I, left panel; Fig. S4R, left panel). Genes only occurring near UV-unique super-enhancers showed a slight trend towards gain in expression after UV irradiation (Fig. 5I, middle panel; Fig. S4R, middle panel). Moreover, the genes that retained super-enhancers, on average, did not show any changes in expression (Fig. 5I, right panel; Fig. S4R, right panel). These findings suggest that UV exposure results in a dramatic reorganization of the super-enhancer landscape, which possibly determines gene expression changes likely to be important for the UV-induced stress response.

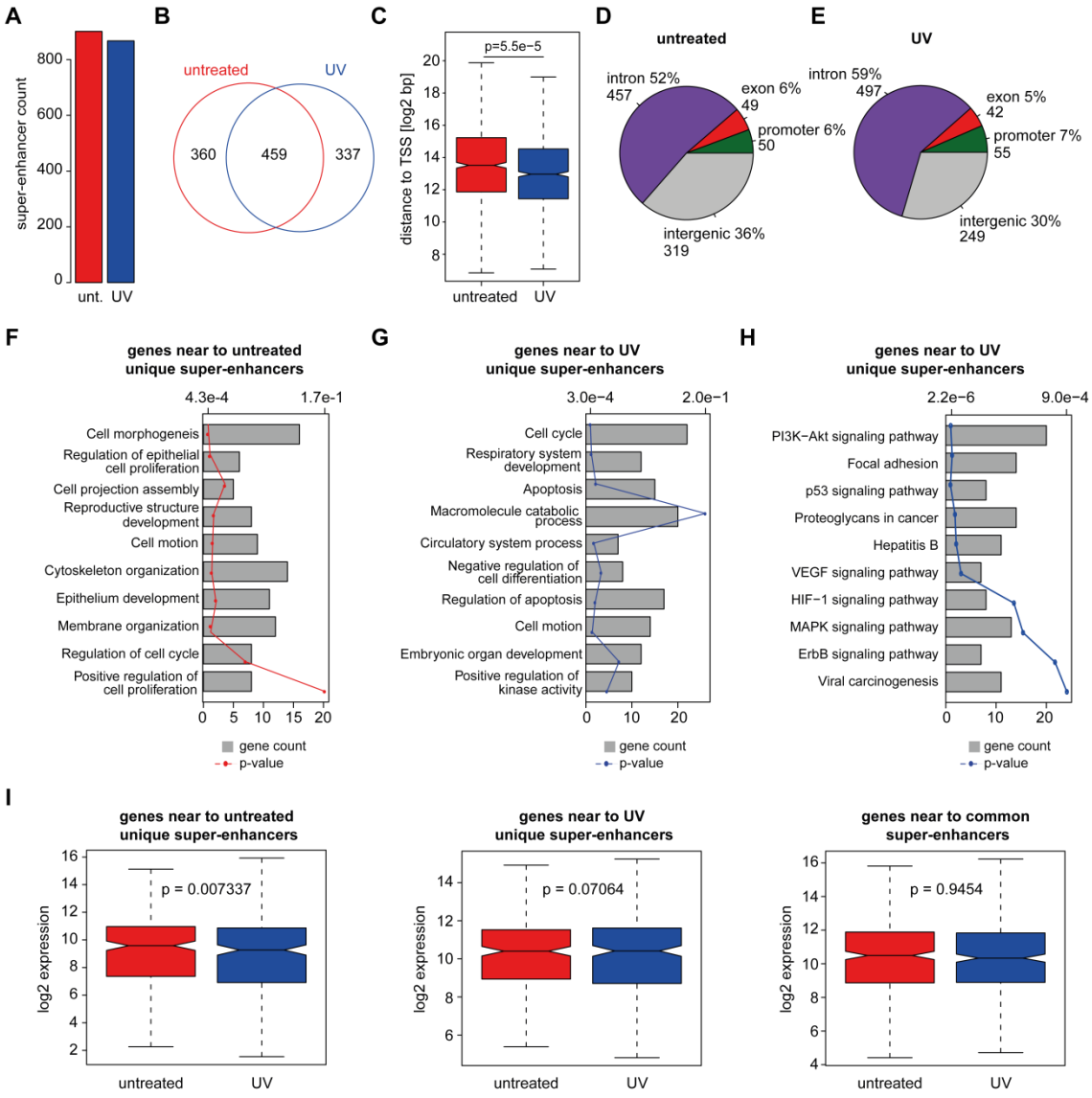



***Figure 5: UV exposure causes a dramatic reorganization of super-enhancers.***

(A) Bar plot showing total count of super-enhancers in untreated and UV conditions. (B) Venn diagram showing overlap of genes near to super-enhancers in untreated and UV conditions. (C) Box plot showing the absolute distance of super-enhancers to nearest transcription start site (TSS). The box represents the 25–75th percentiles, and the median is indicated. The whiskers show 1.5 times the interquartile range (IQR) added to the 75th percentile (upper whisker) or subtracted from the 25th percentile (lower whisker). The notches represent median±1.57×IQR/($n^{0.5}$). *P*-values are calculated using a Wilcoxon test. (D,E) Pie charts showing genomic distribution of super-enhancers detected in untreated NIH3T3 cells (D) or UV-irradiated cells (E). (F,G) Bar plot showing GO term enrichment for biological processes of genes near to super-enhancers unique to the untreated condition (F) or unique to the UV condition (G). Bar lengths represent the count of genes (primary *x*-axis), whereas the *P*-value is shown as a line graph with respect to the upper *x*-axis. (H). Same as in G, but GO term enrichment is calculated for pathways. (I) Boxplots showing expression of genes near to untreated unique super-enhancers (left), UV unique super-enhancers (middle), and common super-enhancers (right) in untreated (red) and UV conditions (blue). *P*-values are calculated using Wilcoxon signed-rank test.

## DISCUSSION

It has been known for a long time that there is nucleosome rearrangement in response to UV [41] and several studies have followed to investigate how chromatin changes are involved in the UV response, mainly by applying microscopy approaches, biochemical assays or single loci studies [74-77]. Recent evidence has suggested that chromatin plays a major role in the UV-induced cellular response including regulation of gene expression changes [5, 6, 8, 12, 13, 15, 46, 78, 79]. However, a detailed genome-wide analysis to investigate the relationship of chromatin and gene expression changes was still missing. It is well established that distal regulatory elements play a crucial role in regulating gene expression; however, until now, no studies had described the changes in the distal regulatory landscape and its relationship to gene expression changes upon UV irradiation [29, 30]. Therefore, for the first time, in this study we comprehensively delineate the genome-wide interplay between chromatin accessibility, H3K27 acetylation and gene expression in mouse fibroblasts 6 h after UV irradiation using several high-throughput sequencing methods. In-depth computational analyses revealed a global chromatin compaction in response to UV, which was accompanied by a massive reorganization of the active histone mark H3K27ac at promoters as well as distal sites. Our data therefore highlight a hitherto unknown role of distal regulatory elements in modulating gene expression in response to UV irradiation. Furthermore, our results show that these chromatin changes often reflect expression changes occurring following UV treatment. Overall, our combinatorial analysis of these comprehensive datasets not only provides new insights into UV-induced chromatin remodeling regulating the expression response but also serves as a highly refined resource for the scientific community.

Our results revealed a preferential down-regulation of gene expression accompanied by an up-regulation of translation machinery components. These results might indicate a stress response mechanism in which cells employ alternative mechanisms to keep the relative protein levels constant by avoiding transcription of damaged DNA that might give rise to aberrant transcripts and non-functional proteins. It is also in line with recent findings, that transcription initiation of a majority of transcribed gene promoters is inhibited following UVB exposure [9]. In the recent past, lincRNAs have been shown to play a role in gene regulation and in a plethora of other biological processes [80-82].

However, until now only a few lincRNAs have been implicated in the DNA damage response [22-24]. Our transcriptome analysis enabled us to reveal many lincRNAs that are differentially expressed upon UV irradiation, which might be crucial regulators of the stress response and be interesting candidates for further mechanistic analysis. These lincRNAs might act in a similar manner to lncRNAs upstream of the *CCND1* promoter, which have been reported to respond to DNA damage signals and to regulate the transcription of *CCND1* through recruitment of the histone acetyltransferase (HAT) inhibitor TLS (also known as FUS) to the *CCND1* promoter [83].

Previous studies have shown a global chromatin relaxation followed by either restoration or further compaction upon UV irradiation [44, 45, 47]. Our findings show a genome-wide loss of chromatin accessibility at 6 h after UV exposure, irrespective of the genomic location, which consolidates our previous observation indicating a potential cellular mechanism to protect genome and transcriptome integrity. The loss of accessibility is accompanied by a gain of H3, indicating that the progressively occurring chromatin compaction is achieved either by active chromatin remodeling or incorporation of new nucleosomes. Although previous studies have suggested that heterochromatic regions are less prone to DNA damage in response to ionizing irradiation, γ-rays or chemical agents [84-87], our and other published data indicate that UV-induced DNA damage might occur both in eu- and hetero-chromatic regions [86, 88].

Interestingly, loss of accessibility was less pronounced at promoters compared to other genomic regions. It is possible that some promoters are actively kept more accessible to maintain transcription or to be able to rapidly reactivate transcription following restoration of a normal cellular state. It is further conceivable that the compaction of promoters requires more energy and time given their occupancy by regulatory proteins as well as transcription machinery and their highly active chromatin status [89]. Despite a genome-wide reduction in chromatin accessibility, a few promoters, including those for genes implicated in the DNA damage response, gain accessibility to allow transcription of the respective genes.

H3K27ac is a histone modification known to mark active promoters and enhancers, and found in our analysis to be drastically reorganized genome-wide following UV exposure [28-31]. Furthermore, time-dependent acquisition of active chromatin status along with

following expression changes argues that UV irradiation modulates chromatin in order to achieve required transcriptome changes. These observations are in line with previous studies that showed changes in active chromatin status along with transcriptional modulation [90, 91]. Strikingly, our data also hint that the epigenetic status prior to UV exposure influences expression as well as chromatin accessibility changes upon UV irradiation, indicating that a repertoire of epigenetic players cooperate to establish the cellular response to UV exposure.

In line with the global loss of chromatin accessibility, we also observed a dramatic loss of accessible (as determined by FAIRE) and H3K27ac-marked (double-positive) regulatory elements. Motif analysis at untreated-unique double-positive sites indicates that, for example, TFAP2A, KLF4, SP2 and EGR1 might be targeted to these sites to support normal cellular functions. Identification of TP53 at UV-unique double-positive sites is further in line with previous observations showing that TP53 acts as a chromatin accessibility regulating factor mediating UV-induced chromatin relaxation as well as a factor that directly binds to enhancers [23, 44]. Moreover, the presence of CTCF, which has previously been shown to protect cells from UV-induced apoptosis [92], at the non-changing double-positive sites might confer protection against UV-induced alterations of the chromatin status.

Among regulatory regions, we also found that super-enhancers underwent a dramatic reorganization following UV treatment. They further seem to play an important role in transcriptional regulation of nearby genes, following the observation made previously that super-enhancers induce expression of proximal genes [32]. Many genes near to UV-induced super-enhancers have previously been shown to be important for the stress response, such as cell cycle, apoptosis and cell survival genes (e.g. *Zfp36l1, Tgif, Ctgf, Cyr61*) [93]. The super-enhancers might be gained next to them to mainly prevent their down-regulation upon UV exposure or even lead to an up-regulation of their expression. Moreover, UV-induced super-enhancers might be involved in the regulation of many signaling pathways of which some are shown to be activated in response to UV [10, 94, 95]. These findings further suggest that in response to stress, cells rearrange their regulatory landscape to allow desired gene expression programs. This expands our knowledge of the gene regulatory networks induced by UV irradiation.

Taken together, our study serves as a comprehensive resource of how chromatin remodeling of promoters and distal regulatory regions relate to expression changes upon UV exposure and provides new insights into the epigenetic responses to UV damage. Further work should involve investigating the functional role of new target genes, including lincRNAs, in UV-induced DNA damage response. In addition, given our findings showing that the genomic regions undergoing epigenetic reprogramming might serve as target sites for sequence-specific transcription factors, future work should aim to unravel their functional impact at these sites in response to UV irradiation.

# METHODS

### Cell culture and UV treatment

NIH3T3 cells (NIH/3T3; ATCC® CRL-1658™, ATCC) were cultured at 37°C under 7% $CO_2$ and 88% relative humidity. The culture medium contained Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, 2 mM L-glutamine and 1× non-essential amino acids. At 2 days before the experiment, cells were seeded at a density of 3000 cells/$cm^2$. Then, the culture medium was removed and the cells were washed twice with Dulbecco's phosphate-buffered saline (D-PBS). Any residual solution was carefully removed, and the uncovered cells were irradiated with 80 J/$m^2$ UVC (254 nm) in a CL-1000 Ultraviolet Crosslinker (UVP). Fresh culture medium containing DMSO (1:2000) was added to the cells, which were cultured again for 6 h or as indicated.

### ChIP assay and FAIRE assays

ChIP and FAIRE assay was performed as previously described with small adaptations [96].

### ATAC

50,000 untreated and UV treated NIH3T3 cells were applied to the ATAC assay according to Buenrostro et al., 2015 [62]. Transposed DNA was amplified for 11 cycles and purified using AMPure XP beads according to the manufacturer's protocol. The DNA was eluted in 15 µl elution buffer (EB, 10 mM Tris-Cl, pH 8.5) and 0.01 µl was used per qPCR measurement.

### RNA isolation, cDNA synthesis and quantitative RT-qPCR

Total RNA was prepared using Trizol and reverse transcribed with the First Strand cDNA Synthesis Kit (Thermo Scientific). Transcripts were quantified by PCR using SYBR Green PCR MasterMix on a ViiA7 PCR machine (Life Technologies). The sequences of all primers used in this study are provided in Table S3.

### Histone isolation and western blot analysis

Histones were isolated according to Abcam's protocol (Abcam Inc., Cambridge, MA). 10 µg of proteins were run on 15% polyacrylamide gels, transferred to a PVDF membrane and probed with respective antibodies.

**Dot blot**

Human H3.3 peptide containing H3K27ac and H3S28p modification (JPT, SP-His_0679; aa1.42) were spotted on a 0.45 µM Amersham™ Protran™ nitrocellulose membrane, blocked in 5% BSA in TBST (0.1% Tween) and then probed with respective antibodies.

**Reagents and antibodies**

Please refer to Table S4.

**RNA-seq**

RNA was isolated in biological triplicates with Purelink RNA Mini Kit and ribosomal RNA was removed using Ribo-Zero rRNA removal kit. 50-bp reads and single-end sequencing of the RNA-seq libraries, prepared with a TruSeq RNA Library Prep kit, were performed on an Illumina Hi-Seq 2000 platform. Fastq files generated from sequencing were processed using TopHat (version 2.0.9) with default parameters for alignment to mouse genome mm9 (available from UCSC) resulting in $14x10^6$-$22x10^6$ reads per sample [97]. FastQC was used for quality control and HTSeq (version 0.5.4p1) to calculate number of mapped reads on protein-coding genes ($n$=19,069) or lincRNAs ($n$=15,694). The normalized expression and differentially expressed genes between the two conditions were identified using DESeq with FDR of 0.1 [98]. Values showing no tag counts in either condition were excluded from the analysis (resulted in 18,784 protein-coding genes and 8857 lincRNAs).

**ChIP-seq and FAIRE-seq**

ChIP- and FAIRE-seq experiments were performed in biological duplicates. Fastq files generated by the sequencing platform were processed using Bowtie (version 0.12.7) with default parameters for alignment to mouse genome mm9 (available from UCSC) [99]. A master bam file was generated by merging the bam files from the replicates in both the untreated and UV conditions. MACS2 was used to call the peaks from the merged bam file without input [100]. Peak enrichments for any peak given by MACS2 or for specific genomic regions (promoters, exons, introns or intergenic regions) were then calculated using the QuasR package [101]. Peaks that were enriched at least 1.5-fold above input were considered for further analysis. Genomic regions (promoters, introns, exons and intergenic regions) were defined using information from UCSC by

applying a hierarchical approach. Promoters were defined first at -800 to +200 bp around the transcription start site (TSS), then exons and intron information was extracted from the UCSC gene annotation file and all remaining regions were considered as intergenic. To build the correlation plots, an overlap of 20% was required for two peaks to be at the same site.

Public datasets from the NCBI GEO database used were: SRR1014989, SRR1015026, SRR1015028, SRR1015029, SRR1015032, SRR1187052, SRR1187056, SRR1187061, SRR118706-4-5, SRR1187064, SRR350001.

**Gene Ontology analysis**

DAVID was used to perform functional annotation clustering using biological process and cellular compartments [102]. From each DAVID cluster, the sub-categories showing the highest number of genes and corresponding *P*-value were chosen as the representing values. Pathway analysis was performed using ToppGene [103]. Gene Set Enrichment analysis on differentially expressed genes was performed using GSEA package from GenePattern [49].

**Motif analysis**

Motif analysis was performed using the Pscan-ChIP tool applying mouse (mm9) genome assembly while using a mixed background set against JASPAR [104]. To identify the exact genomic locations of the motifs predicted by Pscan, we applied findMotifs.pl of Homer package [105].

**Super-enhancer identification**

Super-enhancers were identified using the Homer package [105], which applies a similar approach as described in the first paper reporting super-enhancers [32]. In brief, peak calling was performed using default options and then peaks within 6 kb were stitched together. A score for super-enhancers was calculated using the total normalized read count in the ChIP sample compared to the input sample. These regions were then sorted and normalized based on the highest super-enhancer score.

**Accession numbers**

All the next-generation sequencing datasets used in this study have been submitted to GEO and will be publically available under accession number GSE66286.

## ACKNOWLEDGEMENTS

We thank members of the Tiwari laboratory for cooperation and critical feedback during this project. Support by the Core Facilities of the Institute of Molecular Biology (IMB), Mainz, is gratefully acknowledged, especially the genomics and bioinformatics core facilities.

## COMPETING INTERESTS

The authors declare no competing or financial interests.

## AUTHOR CONTRIBUTIONS

S.S. performed experiments (all experiments except the generation of the RNA-seq samples) and wrote the manuscript. D.F. analyzed data and wrote the manuscript. S. T. analyzed data. S.K.S. performed experiments (generation of the RNA-seq samples). A.G. helped with the progress of the experiments during revision. V.K.T. designed the study, analyzed data and wrote the manuscript.

## FUNDING

Research in the laboratory of V.K.T. is supported by Wilhelm Sander Stiftung [grant number 2012.009.2]; the EpiGeneSys RISE1 program; Marie Curie CIG [grant number 322210]; and Deutsche Forschungsgemeinschaft (DFG) [grant number TI 799/1-1].

## SUPPLEMENTARY INFORMATION

Supplementary information available online at

http://jcs.biologists.org/lookup/suppl/doi:10.1242/jcs.173633/-/DC1

# REFERENCES

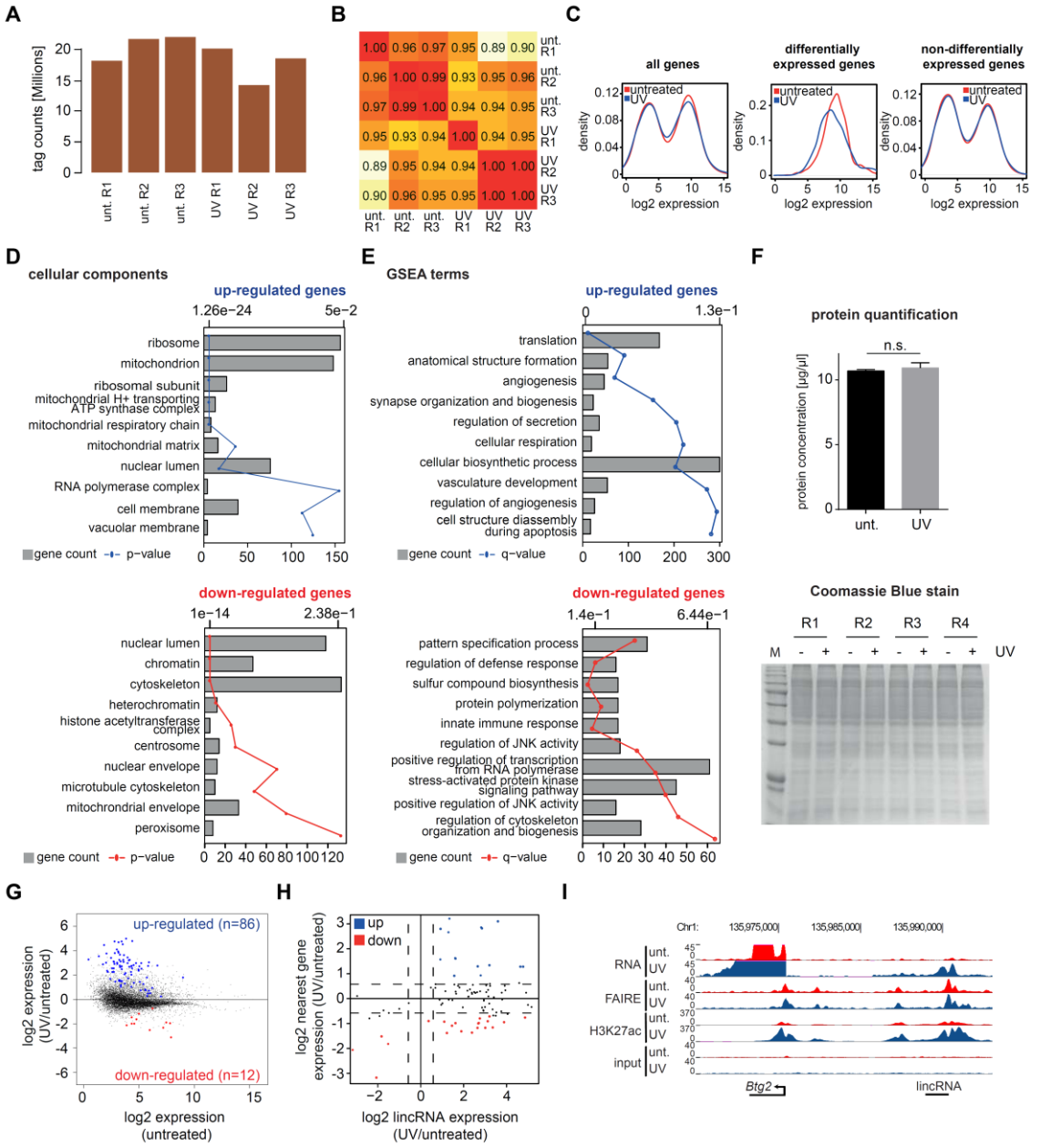

1. Rastogi, R.P., et al., *Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair.* J Nucleic Acids, 2010. **2010**: p. 592980.
2. Iyama, T. and D.M. Wilson, 3rd, *DNA repair mechanisms in dividing and non-dividing cells.* DNA Repair (Amst), 2013. **12**(8): p. 620-36.
3. Lo, J.A. and D.E. Fisher, *The melanoma revolution: from UV carcinogenesis to a new era in therapeutics.* Science, 2014. **346**(6212): p. 945-9.
4. Zhou, B.B. and S.J. Elledge, *The DNA damage response: putting checkpoints in perspective.* Nature, 2000. **408**(6811): p. 433-9.
5. Sesto, A., et al., *Analysis of the ultraviolet B response in primary human keratinocytes using oligonucleotide microarrays.* Proc Natl Acad Sci U S A, 2002. **99**(5): p. 2965-70.
6. Dawes, J.M., et al., *Genome-wide transcriptional profiling of skin and dorsal root ganglia after ultraviolet-B-induced inflammation.* PLoS One, 2014. **9**(4): p. e93338.
7. Bowden, N.A., et al., *Understanding Xeroderma Pigmentosum Complementation Groups Using Gene Expression Profiling after UV-Light Exposure.* Int J Mol Sci, 2015. **16**(7): p. 15985-96.
8. Andrade-Lima, L.C., et al., *DNA repair and recovery of RNA synthesis following exposure to ultraviolet light are delayed in long genes.* Nucleic Acids Res, 2015. **43**(5): p. 2744-56.
9. Gyenis, A., et al., *UVB induces a genome-wide acting negative regulatory mechanism that operates at the level of transcription initiation in human cells.* PLoS Genet, 2014. **10**(7): p. e1004483.
10. Engelberg, D., et al., *The UV response involving the Ras signaling pathway and AP-1 transcription factors is conserved between yeast and mammals.* Cell, 1994. **77**(3): p. 381-90.
11. Laptenko, O. and C. Prives, *Transcriptional regulation by p53: one protein, many possibilities.* Cell Death Differ, 2006. **13**(6): p. 951-61.
12. Hassan, N.M., et al., *Brm inhibits the proliferative response of keratinocytes and corneal epithelial cells to ultraviolet radiation-induced damage.* PLoS One, 2014. **9**(9): p. e107931.
13. Zhang, L., et al., *Whole genome expression profiling shows that BRG1 transcriptionally regulates UV inducible genes and other novel targets in human cells.* PLoS One, 2014. **9**(8): p. e105764.
14. Corpet, A. and G. Almouzni, *A histone code for the DNA damage response in mammalian cells?* EMBO J, 2009. **28**(13): p. 1828-30.
15. Li, Z., et al., *Dynamics of polycomb proteins-mediated histone modifications during UV irradiation-induced DNA damage.* Insect Biochem Mol Biol, 2014. **55**: p. 9-18.
16. Kumar, R., et al., *Chromatin modifications and the DNA damage response to ionizing radiation.* Front Oncol, 2012. **2**: p. 214.
17. Tjeertes, J.V., K.M. Miller, and S.P. Jackson, *Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells.* EMBO J, 2009. **28**(13): p. 1878-89.
18. Kim, M.K., et al., *The role of p300 histone acetyltransferase in UV-induced histone modifications and MMP-1 gene transcription.* PLoS One, 2009. **4**(3): p. e4864.
19. Keum, Y.S., et al., *UVB-induced COX-2 expression requires histone H3 phosphorylation at Ser10 and Ser28.* Oncogene, 2013. **32**(4): p. 444-52.
20. Sawicka, A., et al., *H3S28 phosphorylation is a hallmark of the transcriptional response to cellular stress.* Genome Res, 2014. **24**(11): p. 1808-20.
21. Zhang, Q., et al., *Phosphorylation of histone H3 serine 28 modulates RNA polymerase III-dependent transcription.* Oncogene, 2011. **30**(37): p. 3943-52.
22. Rashi-Elkeles, S., et al., *Parallel profiling of the transcriptome, cistrome, and epigenome in the cellular response to ionizing radiation.* Sci Signal, 2014. **7**(325): p. rs3.
23. Younger, S.T., et al., *Integrative genomic analysis reveals widespread enhancer regulation by p53 in response to DNA damage.* Nucleic Acids Res, 2015. **43**(9): p. 4447-62.
24. Liu, Y. and X. Lu, *Non-coding RNAs in DNA damage response.* Am J Cancer Res, 2012. **2**(6): p. 658-75.
25. Wan, G., et al., *A novel non-coding RNA lncRNA-JADE connects DNA damage signalling to histone H4 acetylation.* EMBO J, 2013. **32**(21): p. 2833-47.
26. Negishi, M., et al., *A new lncRNA, APTR, associates with and represses the CDKN1A/p21 promoter by recruiting polycomb proteins.* PLoS One, 2014. **9**(4): p. e95216.
27. Huarte, M., et al., *A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response.* Cell, 2010. **142**(3): p. 409-19.
28. Shlyueva, D., G. Stampfel, and A. Stark, *Transcriptional enhancers: from properties to genome-wide predictions.* Nat Rev Genet, 2014. **15**(4): p. 272-86.

29. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.
30. Rada-Iglesias, A., et al., *A unique chromatin signature uncovers early developmental enhancers in humans.* Nature, 2011. **470**(7333): p. 279-83.
31. Wang, Z., et al., *Combinatorial patterns of histone acetylations and methylations in the human genome.* Nat Genet, 2008. **40**(7): p. 897-903.
32. Whyte, W.A., et al., *Master transcription factors and mediator establish super-enhancers at key cell identity genes.* Cell, 2013. **153**(2): p. 307-19.
33. Hnisz, D., et al., *Super-enhancers in the control of cell identity and disease.* Cell, 2013. **155**(4): p. 934-47.
34. Brown, J.D., et al., *NF-kappaB directs dynamic super enhancer formation in inflammation and atherogenesis.* Molecular Cell, 2014. **56**(2): p. 219-31.
35. Chapuy, B., et al., *Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma.* Cancer Cell, 2013. **24**(6): p. 777-90.
36. Schmidt, S.F., et al., *Acute TNF-induced repression of cell identity genes is mediated by NFkappaB-directed redistribution of cofactors from super-enhancers.* Genome Res, 2015.
37. Vahedi, G., et al., *Super-enhancers delineate disease-associated regulatory nodes in T cells.* Nature, 2015. **520**(7548): p. 558-62.
38. Mansour, M.R., et al., *Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element.* Science, 2014. **346**(6215): p. 1373-7.
39. Lans, H., J.A. Marteijn, and W. Vermeulen, *ATP-dependent chromatin remodeling in the DNA-damage response.* Epigenetics Chromatin, 2012. **5**: p. 4.
40. Price, B.D. and A.D. D'Andrea, *Chromatin remodeling at DNA double-strand breaks.* Cell, 2013. **152**(6): p. 1344-54.
41. Smerdon, M.J. and M.W. Lieberman, *Nucleosome rearrangement in human chromatin during UV-induced DNA- reapir synthesis.* Proc Natl Acad Sci U S A, 1978. **75**(9): p. 4238-41.
42. Smerdon, M.J. and M.W. Lieberman, *Distribution within chromatin of deoxyribonucleic acid repair synthesis occurring at different times after ultraviolet radiation.* Biochemistry, 1980. **19**(13): p. 2992-3000.
43. Smerdon, M.J., J.F. Watkins, and M.W. Lieberman, *Effect of histone H1 removal on the distribution of ultraviolet-induced deoxyribonucleic acid repair synthesis within chromatin.* Biochemistry, 1982. **21**(16): p. 3879-85.
44. Rubbi, C.P. and J. Milner, *p53 is a chromatin accessibility factor for nucleotide excision repair of DNA damage.* EMBO J, 2003. **22**(4): p. 975-86.
45. Soria, G., S.E. Polo, and G. Almouzni, *Prime, repair, restore: the active role of chromatin in the DNA damage response.* Molecular Cell, 2012. **46**(6): p. 722-34.
46. Izhar, L., et al., *A Systematic Analysis of Factors Localized to Damaged Chromatin Reveals PARP-Dependent Recruitment of Transcription Factors.* Cell Rep, 2015. **11**(9): p. 1486-500.
47. Burgess, R.C., et al., *Activation of DNA damage response signaling by condensed chromatin.* Cell Rep, 2014. **9**(5): p. 1703-17.
48. Papamichos-Chronakis, M. and C.L. Peterson, *Chromatin and the genome integrity network.* Nat Rev Genet, 2013. **14**(1): p. 62-75.
49. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
50. Tanos, T., et al., *Phosphorylation of c-Fos by members of the p38 MAPK family. Role in the AP-1 response to UV light.* J Biol Chem, 2005. **280**(19): p. 18842-52.
51. Darwiche, N., et al., *Regulation of ultraviolet B radiation-mediated activation of AP1 signaling by retinoids in primary keratinocytes.* Radiat Res, 2005. **163**(3): p. 296-306.
52. Chaum, E., et al., *Quantitative AP-1 gene regulation by oxidative stress in the human retinal pigment epithelium.* J Cell Biochem, 2009. **108**(6): p. 1280-91.
53. Luo, H., et al., *Comprehensive characterization of 10,571 mouse large intergenic noncoding RNAs from whole transcriptome sequencing.* PLoS One, 2013. **8**(8): p. e70835.
54. Kornienko, A.E., et al., *Gene regulation by the act of long non-coding RNA transcription.* BMC Biol, 2013. **11**: p. 59.
55. Vance, K.W. and C.P. Ponting, *Transcriptional regulatory functions of nuclear long noncoding RNAs.* Trends Genet, 2014. **30**(8): p. 348-55.
56. Gallitano-Mendel, A., et al., *Mice lacking the immediate early gene Egr3 respond to the anti-aggressive effects of clozapine yet are relatively resistant to its sedating effects.* Neuropsychopharmacology, 2008. **33**(6): p. 1266-75.

57. Imran, M. and I.K. Lim, *Regulation of Btg2(/TIS21/PC3) expression via reactive oxygen species-protein kinase C-NuFkappaBeta pathway under stress conditions.* Cell Signal, 2013. **25**(12): p. 2400-12.
58. Arlt, A. and H. Schafer, *Role of the immediate early response 3 (IER3) gene in cellular stress response, inflammation and tumorigenesis.* Eur J Cell Biol, 2011. **90**(6-7): p. 545-52.
59. Landau, G., et al., *Expression profiling and biochemical analysis suggest stress response as a potential mechanism inhibiting proliferation of polyamine-depleted cells.* J Biol Chem, 2012. **287**(43): p. 35825-37.
60. Giresi, P.G., et al., *FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin.* Genome Res, 2007. **17**(6): p. 877-85.
61. Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nat Methods, 2013. **10**(12): p. 1213-8.
62. Buenrostro, J.D., et al., *ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide.* Curr Protoc Mol Biol, 2015. **109**: p. 21 29 1-9.
63. Davie, K., et al., *Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling.* PLoS Genet, 2015. **11**(2): p. e1004994.
64. Oh, K.S., et al., *UV-induced histone H2AX phosphorylation and DNA damage related proteins accumulate and persist in nucleotide excision repair-deficient XP-B cells.* DNA Repair (Amst), 2011. **10**(1): p. 5-15.
65. Doyle, B., et al., *Chromatin loops as allosteric modulators of enhancer-promoter interactions.* PLoS Comput Biol, 2014. **10**(10): p. e1003867.
66. Samee, M.A. and S. Sinha, *Quantitative modeling of a gene's expression from its intergenic sequence.* PLoS Comput Biol, 2014. **10**(3): p. e1003467.
67. Andersson, R., et al., *An atlas of active enhancers across human cell types and tissues.* Nature, 2014. **507**(7493): p. 455-61.
68. Phan, D., et al., *Identification of Sp2 as a transcriptional repressor of carcinoembryonic antigen-related cell adhesion molecule 1 in tumorigenesis.* Cancer Res, 2004. **64**(9): p. 3072-8.
69. Rowland, B.D., R. Bernards, and D.S. Peeper, *The KLF4 tumour suppressor is a transcriptional repressor of p53 that acts as a context-dependent oncogene.* Nat Cell Biol, 2005. **7**(11): p. 1074-82.
70. Huang, R.P., et al., *Egr-1 inhibits apoptosis during the UV response: correlation of cell survival with Egr-1 phosphorylation.* Cell Death Differ, 1998. **5**(1): p. 96-106.
71. de Belle, I., et al., *p53 and Egr-1 additively suppress transformed growth in HT1080 cells but Egr-1 counteracts p53-dependent apoptosis.* Oncogene, 1999. **18**(24): p. 3633-42.
72. Zhang, W. and S. Chen, *EGR-1, a UV-inducible gene in p53(-/-) mouse cells.* Exp Cell Res, 2001. **266**(1): p. 21-30.
73. Zhu, Y., D. van Essen, and S. Saccani, *Cell-type-specific control of enhancer activity by H3K9 trimethylation.* Molecular Cell, 2012. **46**(4): p. 408-23.
74. Polo, S.E., *Reshaping Chromatin after DNA Damage: The Choreography of Histone Proteins.* J Mol Biol, 2014.
75. Yu, Y., et al., *UV irradiation stimulates histone acetylation and chromatin remodeling at a repressed yeast locus.* Proc Natl Acad Sci U S A, 2005. **102**(24): p. 8650-5.
76. Yu, S., et al., *How chromatin is remodelled during DNA repair of UV-induced DNA damage in Saccharomyces cerevisiae.* PLoS Genet, 2011. **7**(6): p. e1002124.
77. Kruhlak, M.J., et al., *Changes in chromatin structure and mobility in living cells at sites of DNA double-strand breaks.* J Cell Biol, 2006. **172**(6): p. 823-34.
78. Duan, M.R. and M.J. Smerdon, *Histone H3 lysine 14 (H3K14) acetylation facilitates DNA repair in a positioned nucleosome by stabilizing the binding of the chromatin Remodeler RSC (Remodels Structure of Chromatin).* J Biol Chem, 2014. **289**(12): p. 8353-63.
79. Dinant, C., et al., *Enhanced chromatin dynamics by FACT promotes transcriptional restart after UV-induced DNA damage.* Molecular Cell, 2013. **51**(4): p. 469-79.
80. Cheetham, S.W., et al., *Long noncoding RNAs and the genetics of cancer.* Br J Cancer, 2013. **108**(12): p. 2419-25.
81. Dinger, M.E., et al., *Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation.* Genome Res, 2008. **18**(9): p. 1433-45.
82. Xu, D., et al., *Long noncoding RNAs associated with liver regeneration 1 accelerates hepatocyte proliferation during liver regeneration by activating Wnt/beta-catenin signaling.* Hepatology, 2013. **58**(2): p. 739-51.
83. Wang, X., et al., *Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription.* Nature, 2008. **454**(7200): p. 126-30.
84. Seo, J., et al., *Genome-wide profiles of H2AX and gamma-H2AX differentiate endogenous and exogenous DNA damage hotspots in human cells.* Nucleic Acids Res, 2012. **40**(13): p. 5965-74.
85. Takata, H., et al., *Chromatin compaction protects genomic DNA from radiation damage.* PLoS One, 2013. **8**(10): p. e75622.

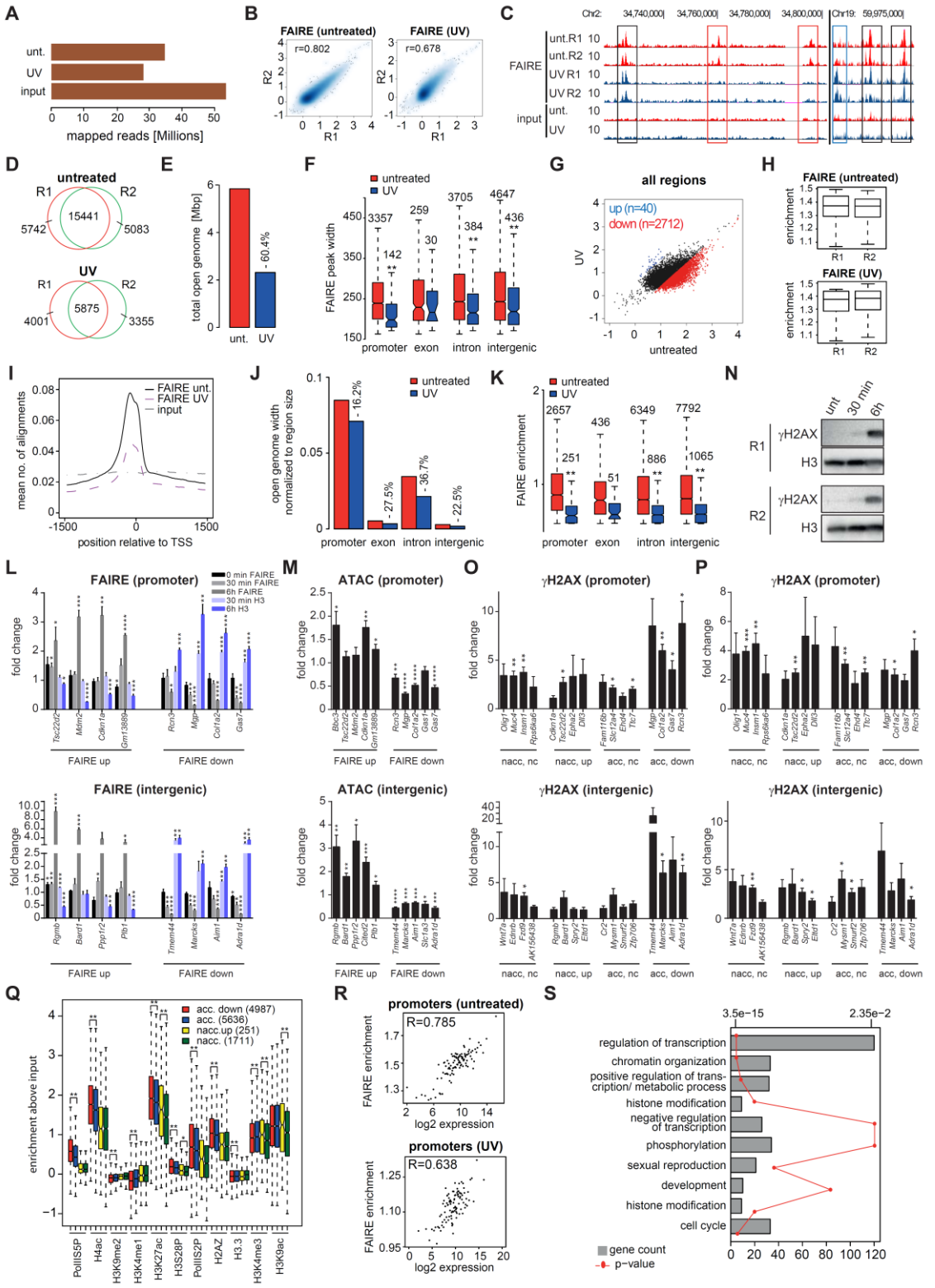
86. Yu, W., *Genome-wide analysis of DNA damage and repair.* PhD thesis, 2014.
87. Falk, M., E. Lukasova, and S. Kozubek, *Chromatin structure influences the sensitivity of DNA to gamma-radiation.* Biochim Biophys Acta, 2008. **1783**(12): p. 2398-414.
88. Zavala, A.G., et al., *High-resolution characterization of CPD hotspot formation in human fibroblasts.* Nucleic Acids Res, 2014. **42**(2): p. 893-905.
89. Lenhard, B., A. Sandelin, and P. Carninci, *Metazoan promoters: emerging characteristics and insights into transcriptional regulation.* Nat Rev Genet, 2012. **13**(4): p. 233-45.
90. Suganuma, T. and J.L. Workman, *Signals and combinatorial functions of histone modifications.* Annu Rev Biochem, 2011. **80**: p. 473-99.
91. Zhou, V.W., A. Goren, and B.E. Bernstein, *Charting histone modifications and the functional organization of mammalian genomes.* Nat Rev Genet, 2011. **12**(1): p. 7-18.
92. Li, T. and L. Lu, *Functional role of CCCTC binding factor (CTCF) in stress-induced apoptosis.* Exp Cell Res, 2007. **313**(14): p. 3057-65.
93. Thakurela, S., et al., *Dynamics and function of distal regulatory elements during neurogenesis and neuroplasticity.* Genome Res, 2015.
94. Lu, L., L. Wang, and B. Shell, *UV-induced signaling pathways associated with corneal epithelial cell apoptosis.* Invest Ophthalmol Vis Sci, 2003. **44**(12): p. 5102-9.
95. Stokes, M.P., et al., *Profiling of UV-induced ATM/ATR signaling pathways.* Proc Natl Acad Sci U S A, 2007. **104**(50): p. 19855-60.
96. Sahu, S.K., et al., *JNK-dependent gene regulatory circuitry governs mesenchymal fate.* EMBO J, 2015. **34**(16): p. 2162-81.
97. Trapnell, C., et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nat Biotechnol, 2010. **28**(5): p. 511-5.
98. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.
99. Langmead, B., *Aligning short sequencing reads with Bowtie.* Curr Protoc Bioinformatics, 2010. **Chapter 11**: p. Unit 11 7.
100. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
101. Gaidatzis, D., et al., *QuasR: quantification and annotation of short reads in R.* Bioinformatics, 2014.
102. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic Acids Res, 2009. **37**(1): p. 1-13.
103. Chen, J., et al., *ToppGene Suite for gene list enrichment analysis and candidate gene prioritization.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W305-11.
104. Zambelli, F., G. Pesole, and G. Pavesi, *PscanChIP: Finding over-represented transcription factor-binding site motifs and their correlations in sequences from ChIP-Seq experiments.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W535-43.
105. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Molecular Cell, 2010. **38**(4): p. 576-89.

# SUPPLEMENTARY FIGURES




***Supplementary Figure 1***

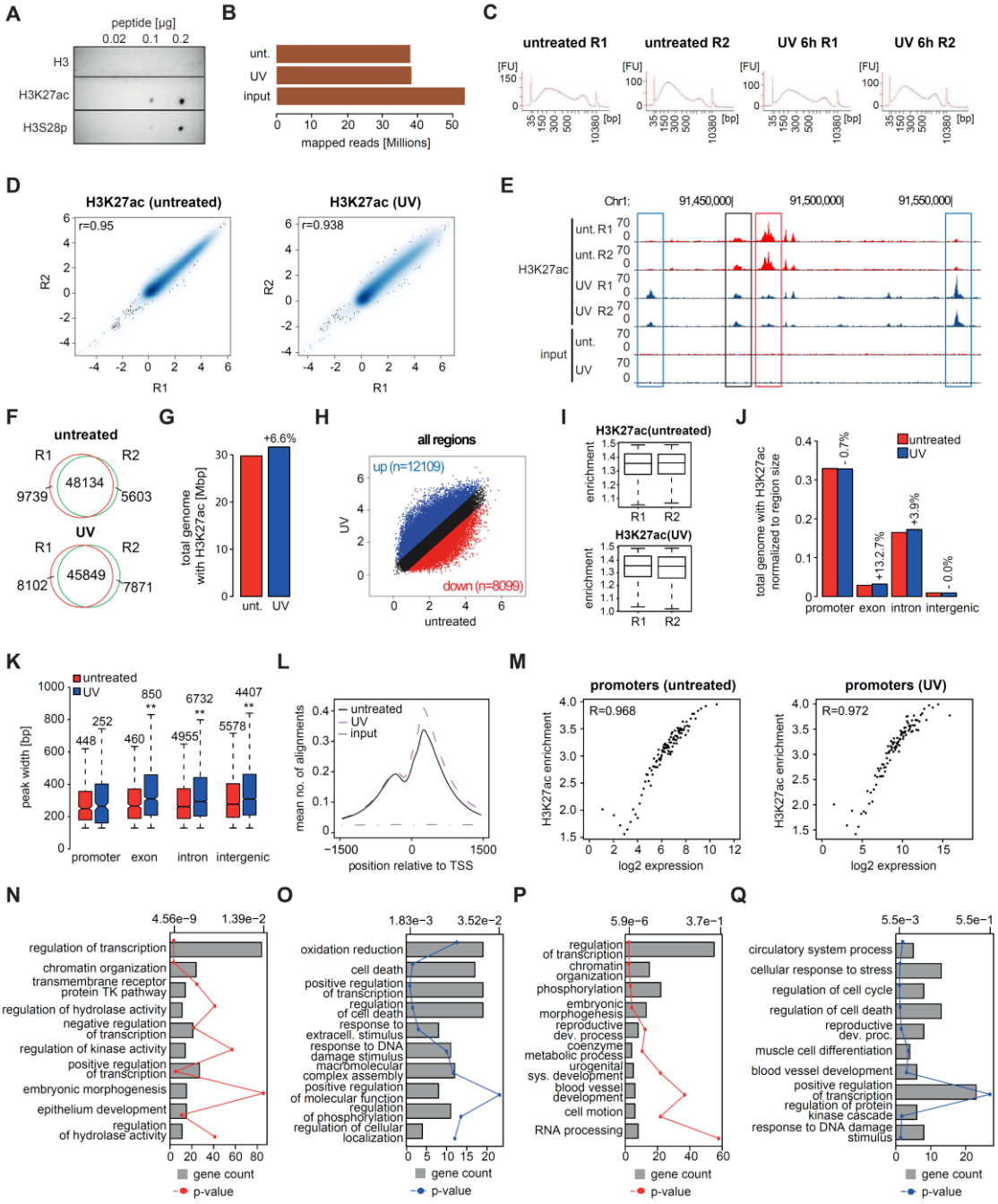
(A). Bar plot showing total number of mapped reads for each RNA-seq sample. (B). Correlation matrix of gene expression values between RNA-seq replicates using normalized tag counts. Values displayed are Spearman coefficients. (C). Density plots showing distribution of normalized tag counts in untreated (red lines) and UV condition (blues lines). Left panel: density plot for protein-coding genes (n=19,069). Middle panel: density plot for differentially expressed genes only (n=2068). Right panel: density plot for genes not differentially expressed (n=17,001). (D). Bar plot showing enrichment of GO terms associated to cellular component for up- and down-regulated genes upon UV treatment. Bar length shows the count of genes (primary x-axis), while p-value is shown as line graph with respect to alternate x-axis. (E). Gene ontology analysis performed using "GSEA preranked (v1)" module from GenePattern for differentially expressed genes. GO-terms associated to biological processes with the best enrichment values are displayed. GSEA was used with the default parameters and top 10 GO terms are displayed sorted by their FDR q-value. Bar length represents the gene counts (primary x-axis), while FDR p-value is shown as line graph with respect to upper x-axis. (F). Bar plot representing the protein concentration of a defined number of untreated and UV treated cells lysed in the same volume (n=4, error bars represent s.e.m.; upper panel). An unpaired t-test did not reveal a significant change upon UV treatment. The same amount of lysate from each sample was separated on an acrylamide gel and Coomassie Blue stained (M = marker, R = replicate, - : untreated, +: UV treated; lower panel). (G). MA-plot showing the changes in expression of lincRNA genes six hours after UV treatment in NIH3T3 cells. The log2 ratio of change in gene expression for each lincRNA (y-axis) is plotted against log2 mean expression value in untreated cells (x-axis). Differentially expressed lincRNAs are highlighted (blue: up-regulated, n=86; red: down-regulated, n=12). (H). Scatter plot showing the relationship between log2 ratio of change in lincRNA expression (x-axis) and the log2 ratio of expression change of the nearest gene (y-axis). LincRNAs found significantly changing following UV in figure S1G are displayed. The nearest genes that are significantly up-regulated are displayed in blue and the down-regulated genes are displayed in red. Dotted lines mark the cut-off of 1.5 fold expression change. (I). Browser track showing an up-regulated lincRNA, whose nearest gene also showed an increased in expression upon UV irradiation (*Btg2*). Tracks for expression, FAIRE and H3K27ac are shown for untreated (red) and UV (blue) conditions. Input for untreated and UV conditions are shown for comparison.




***Supplementary Figure 2***

(A). Replicates of each condition of the FAIRE-seq experiments have been merged and number of total mapped reads per condition are shown. (B). Scatterplots showing correlation between peak enrichments from replicate 1 and 2 called independently (left panel: untreated condition, right panel: UV condition). Values displayed are Spearman correlation coefficients. (C). Browser track showing reproducible patterns of peaks between replicates of the same condition. Red tracks show replicates for untreated and blue tracks for UV condition. The black box shows an example of peaks visible in all replicates, the red boxes show peaks visible for both untreated replicates only and the blue box shows an example of peaks only found in UV condition. Input tracks are displayed for untreated and UV conditions for comparison. (D). Venn diagrams showing the number of overlapping enriched peaks between independently called peaks from replicate 1 and replicate 2. Top diagram shows the overlap in untreated situation and bottom diagram the overlap of replicates in UV condition. (E).

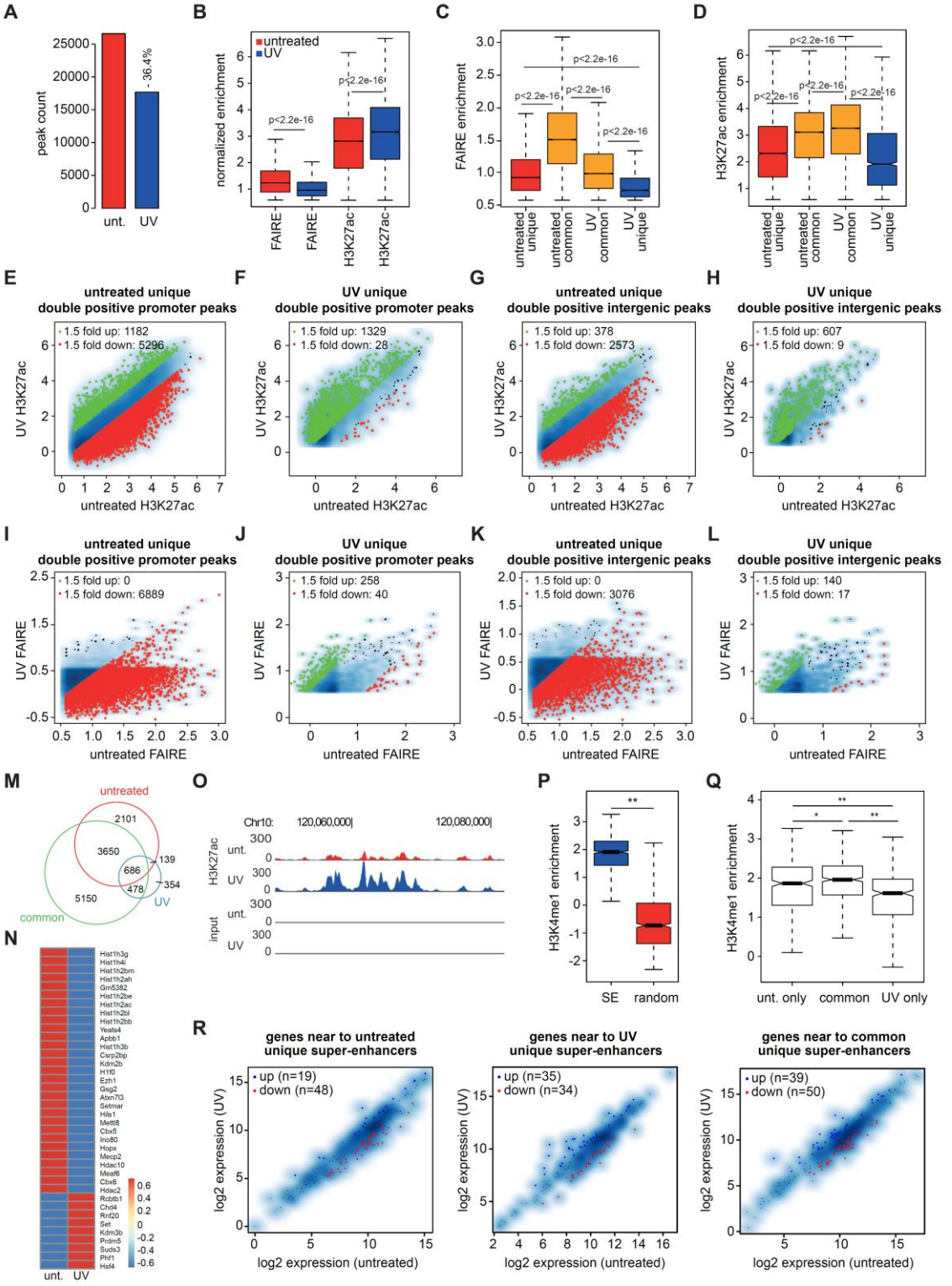
Bar plot showing total open genome size in Mbp in untreated and UV treated NIH3T3 cells calculated from the peaks detected in both replicates. Total open genome size was calculated as the sum of the widths of all FAIRE peaks found to be enriched above background in each condition. (F). Box plots showing distribution of peak width for unique accessible regions in untreated and UV-treated cells calculated with the peaks found in both replicates. Significance between peak width in untreated and UV condition is indicated by asterisks (**: p-value <0.01, Wilcoxon rank-sum test for non-normal distributions; Promoters: p-value = $1.2x10^{-11}$; exons: p-value = 0.3887; introns: p-value = $7.6x10^{-10}$; intergenic regions: p-value = $1.8x10^{-8}$). The number above each box plot displays the number of unique peaks for each condition. (G). Scatter plot showing dynamics of FAIRE enrichment changes between untreated and UV conditions for peaks found in both replicates using independent peak calling. Regions showing more than 1.5-fold gain of enrichment 6h following UV exposure are depicted in blue, while the regions showing a decrease after UV treatment of more than 1.5-fold are depicted in red. (H). Box plots showing the enrichment values in individual replicates of FAIRE sites found to be enriched 1.5-fold above input only in the merged approach. (I). Density plot representing the reads falling around TSS for untreated condition, UV condition and input in merged samples. (J). Bar plot showing the total amount of open genome in different genomic regions (promoter, intron, exon and intergenic regions) normalized by the size of the regions they occupy in the genome. (K). Distribution of FAIRE peak enrichments in different genomic regions for peaks only enriched in untreated (red) or UV-treated (blue) cells. Wilcoxon rank-sum tests for non-normal distribution were performed to test the equality of distribution between the untreated and UV conditions (**: p-value<0.01; promoters: p-value = $1.581x10^{-9}$; exons: p-value = 0.774; introns: p-value = $1.762x10^{-15}$; intergenic regions: p-value = $8.188x10^{-7}$). (L). FAIRE-qPCR for promoter (upper panel) and intergenic regions (lower panel) detected by FAIRE-seq to change accessibility 6h after UV irradiation (indicated below). FAIRE-qPCR was performed in untreated cells and at different time points after UV irradiation (immediately, 30 minutes, 6h). The results were normalized to input and *Ctcf* and plotted as fold change to the respective untreated sample (n=3; error bars represent s.e.m.). Furthermore, H3 occupancy at these loci has been measured by ChIP-qPCR, normalized to input and plotted as fold change (n=3, error bars represent s.e.m.). Statistical significance for FAIRE- and ChIP-qPCR results were calculated with an unpaired t-test (*: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001; ****: p-value < 0.0001). (M). ATAC-qPCR for promoter (upper panel) and intergenic (lower panel) regions detected by FAIRE-seq to change accessibility by 6h after UV (indicated below). The results are plotted as fold change between UV and untreated condition (n=4; error bars represent s.e.m.). Statistical significance for ATAC-qPCR results were calculated with an unpaired t-test (*: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001; ****: p-value < 0.0001). (N). Western Blot analysis of γH2AX in untreated NIH3T3 cells as well as 30 minutes or 6h after UV irradiation. Two independent replicates (R1 and R2) are shown with H3 as a loading control. (O). γH2AX ChIP-qPCR results for promoters (upper panel) and intergenic (lower panel) regions, which are either accessible (acc) or not (nacc) in untreated condition and change (up or down) or do not change (nc) their accessibility 6h after UV treatment. The data are normalized to input and fold changed between UV and untreated condition is plotted (n=3; error bars represent s.e.m.). Statistical significance for ChIP-qPCR results were calculated with an unpaired t-test (*: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001; ****: p-value < 0.0001). (P). Same as in figure S2O, but the data was additionally normalized to H3 enrichment and fold changes are shown (n=3; error bars represent s.e.m.). Statistical significance for ChIP-qPCR results were calculated with an unpaired t-test (*: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001; ****: p-value < 0.0001). (Q). Distribution of enrichment values of various chromatin features for sites of different accessibility change patterns. In red, sites found accessible in untreated cells and going down upon UV. In blue, sites accessible in untreated cells but not showing significant change upon UV. In yellow, sites gaining accessibility upon UV. In green, inaccessible sites whose status does not change upon UV. Eventual significant differences between regions belonging to red and blue categories on one hand, yellow and green categories on the other hand are displayed. Statistical significance were calculated with Wilcoxon rank-sum test (*: p-value < 0.05; **: p-value < 0.01). (R). Scatter plots showing correlation between gene expression and accessibility for untreated (upper panel) and UV condition (lower panel). Expression levels have been divided into 100 bins of equal size and expression mean of each bin has been plotted against mean of FAIRE-seq enrichments at promoters of the corresponding genes. Spearman correlation coefficients are displayed inside the graph. (S). Bar plot showing enrichment of biological processes for down-regulated genes showing a loss of accessibility at their promoters (red quadrant of figure 2F). Bar length shows the number of genes (primary x-axis), while p-values are shown as a line graph with respect to alternate x-axis.




***Supplementary Figure 3***

(A). Dot plot analysis of H3, H3K27ac and H3S28p antibody binding to a human H3.3 peptide (aa 1-42) containing acetylated K27 and phosphorylated S28. The H3 antibody recognizes only the first 20 amino acids of an unmodified H3. (B). Replicates of each condition of the H3K27ac ChIP-seq experiments have been merged and number of total mapped reads per condition are shown. (C). Bioanalyzer profiles for both replicates of the untreated and UV irradiated sonicated samples used for ChIP-seq experiments. (D). Scatterplots showing correlation between peak enrichments from replicate 1 and 2 called independently (left panel: untreated condition, right panel: UV condition). Values displayed are Spearman's coefficients of correlation. (E). Browser track showing reproducible patterns of peaks between replicates of the same condition for H3K27ac ChIP-seq and input samples. The red tracks show replicates for the untreated condition and blue tracks for UV condition. The black box shows an example of peaks visible in all conditions, the red box marks peaks higher enriched in untreated replicates compared to UV replicates, and the blue boxes show two examples of peaks only present in UV condition. (F). Venn diagrams showing the number of overlapping peaks for replicate 1 and 2 of the ChIP-seq experiment using independent peak calling. Top diagram shows the overlap in untreated situation and bottom diagram for UV condition. (G). Bar plot showing total size of the genome marked with H3K27ac, calculated by summing the widths of all H3K27ac peaks found to be enriched above input in both replicates. (H). Scatter plot showing dynamics of H3K27ac enrichment changes between untreated and UV conditions for peaks found in both replicates using independent peak calling. Regions showing more than 1.5-fold gain in enrichment in UV condition compared to untreated condition are depicted in blue, while the regions showing a decrease after UV treatment of more than 1.5-fold are depicted in red. (I). Box plots showing the enrichment values in individual replicates of H3K27ac sites found to be enriched 1.5-fold above input only in the merged approach. (J). Bar plot showing total size of H3K27ac enriched regions found in each of four genomic regions (promoter, intron, exon and intergenic regions) normalized by the size of the regions they occupy in the genome. (K). Peak-width distribution across

different genomic regions for H3K27ac-enriched regions uniquely found in untreated or UV-treated cells. Wilcoxon rank-sum tests for non-normal distribution were performed to test the equality of distribution between untreated and UV conditions (**: p-value<0.01; promoters: p-value = 0.7776; exons: p-value: $1.281x10^{-6}$; introns: p-value < $2.2x10^{-16}$; intergenic: p-value < $2.2x10^{-16}$). (L). Density plot representing the reads falling around TSS for untreated condition, UV condition and input in merged samples. (M). Scatter plots showing correlation between gene expression and H3K27ac enrichment for untreated (left panel) and UV (right panel) condition. Expression levels have been divided into 100 bins of equal size and expression mean of each bin has been plotted against mean of H3K27ac enrichments at promoters of the corresponding genes. Spearman correlation coefficients are displayed inside the graph. (N). Gene ontologies associated with down-regulated genes showing a loss of H3K27ac enrichments at their promoters following UV exposure (red quadrant in figure 3F). Bar lengths show the count of genes (primary x-axis), while p-values are shown as a line graph with respect to alternate x-axis. (O). Gene ontologies associated with up-regulated genes showing increase in H3K27ac enrichments at their promoters following UV exposure (blue quadrant in figure 3F). Bar lengths show the count of genes (primary x-axis), while p-values are shown as a line graph with respect to alternate x-axis. (P). Gene ontologies associated to genes falling in the red quadrant in figure 3K. Bar lengths show the count of genes (primary x-axis), while p-values are shown as a line graph with respect to alternate x-axis. (Q). Gene ontologies associated to genes falling in the blue quadrant in figure 3K. Bar lengths show the count of genes (primary x-axis), while p-values are shown as a line graph with respect to alternate x-axis.




***Supplementary Figure 4***

(A). Bar plots showing number of H3K27ac and FAIRE double-positive sites in untreated (red) and UV (blue) conditions. Overlapping peaks were decided based on minimal 20% overlap of FAIRE and H3K27ac peaks. (B). Box plots showing changes in FAIRE and H3K27ac enrichments at double-positive sites in untreated (red) and UV (blue) conditions. The y-axis shows normalized enrichment in log2. P-values are calculated using a paired Wilcoxon test. (C). Box plots showing FAIRE enrichments at double-positive regions either specific to untreated condition (red), specific to UV condition (blue), or present in both (orange). In the latter case, the values of peak enrichments are shown for untreated condition ("untreated common") and UV condition ("UV common"). The y-axis shows normalized FAIRE enrichment in log2 scale. P-values are calculated using a paired Wilcoxon test. (D). Same as in C but showing enrichment for H3K27ac. P-values are calculated using a paired Wilcoxon test. (E-H). Scatter plot showing changes in H3K27ac enrichments at unique double positive sites found in the untreated condition (E, G) or UV condition (F, H) at promoter (E, F) or intergenic regions (G, H). The x-axis shows log2 of H3K27ac enrichment in the untreated condition, while the y-axis represents log2 of H3K27ac enrichment in the UV condition. (I-L) Same as in E-H, but for FAIRE enrichment. (M). Venn diagram showing overlap of genes that are near to double-positive sites unique to

untreated (untreated) or UV (UV) condition, or enriched for FAIRE and H3K27ac in untreated and UV condition (common). (N). Heat map for genes associated with the GO term category 'chromatin organization' from figure 4E. The heatmap illustrates the direction of gene expression changes 6h after UV irradiation. (O). Browser track showing a super-enhancer induced upon UV irradiation. Tracks showing H3K27ac signal in untreated (red track) and UV (blue track) condition are displayed, along with the corresponding inputs. (P). Boxplots showing the distribution of H3K4me1 enrichment above input for super-enhancers and random genomic intervals selected based on the average super-enhancer size (published H3K4me1 ChIP-seq dataset performed in untreated 3T3 fibroblast cells was used: SRR349993). (Q). Boxplots showing H3K4me1 enrichment in untreated 3T3 cells at super-enhancers regions unique to untreated condition, common to untreated and UV conditions, and unique to UV condition (dataset used: SRR349993). Wilcoxon rank-sum tests for non-normal distribution were performed to test the equality of distribution between two of three groups (*: p-value<0.05; untreated only versus common: p-value = $1.34x10^{-2}$; untreated only versus UV only: p-value = $2.64x10^{-6}$; common versus UV only: p-value = $8.19x10^{-15}$). (R). Scatter plots showing gene expression changes of genes near to untreated unique super-enhancers (left panel), UV unique super-enhancers (middle panel) and common super-enhancers (right panel). The x-axis shows log2 expression in the untreated condition, while the y-axis represents log2 expression in the UV condition. Differentially expressed genes are highlighted (down-regulated genes in red, up-regulated genes in blue).

# SUPPLEMENTARY TABLES

For supplemental tables please refer to the Excel-files published together with this manuscript.

**Supplementary Table 1.**

Tables showing gene ontology analysis for different set of genes indicated in the sheet name.

**Supplementary Table 2.**

Tables show the results of the Pscan motif analysis for untreated unique, UV unique and common between both conditions accessible and H3K27ac-marked sites.

**Supplementary Table 3.**

Primers used for RT-qPCR and ChIP-, FAIRE- or ATAC-qPCRs and for ATAC library preparation.

**Supplementary Table 4.**

Details of the reagents and antibodies used in the study.

# 2.2

# Identifying Novel Transcriptional Regulators with Circadian Expression

Sandra Schick[1,¶], Kolja Becker[2,¶], Sudhir Thakurela[1], David Fournier[1], Mareike Hildegard Hampel[3], Stefan Legewie[2,*], Vijay K. Tiwari[1,*]

1. Epigenetic Regulation of Development and Disease Group, Institute of Molecular Biology, Mainz, Germany
2. Modelling of Biological Networks Group, Institute of Molecular Biology, Mainz, Germany
3. Dr. Senckenbergisches Chronomedizinisches Institut, Frankfurt am Main, Germany

*Corresponding authors: S.Legewie@imb-mainz.de (SL) and V.Tiwari@imb-mainz.de (VKT)

¶ SS and KB contributed equally to this work.

Running title: Novel Genes with Circadian Expression

Keywords: Circadian rhythm, gene expression, genomics, bioinformatics, systems biology

**Published in the journal Molecular and Cellular Biology (MCB):**

Mol Cell Biol. 2015 Dec 7;36(4):545-58. doi: 10.1128/MCB.00701-15.

## ABSTRACT

Organisms adapt their physiology and behavior to the 24-h day-night cycle to which they are exposed. On a cellular level, this is regulated by intrinsic transcriptional-translational feedback loops that are important for maintaining the circadian rhythm. These loops are organized by members of the core clock network, which further regulate transcription of downstream genes, resulting in their circadian expression. Despite progress in understanding circadian gene expression, only a few players involved in circadian transcriptional regulation, including transcription factors, epigenetic regulators, and long noncoding RNAs, are known. Aiming to discover such genes, we performed a high-coverage transcriptome analysis of a circadian time course in murine fibroblast cells. In combination with a newly developed algorithm, we identified many transcription factors, epigenetic regulators, and long intergenic noncoding RNAs that are cyclically expressed. In addition, a number of these genes also showed circadian expression in mouse tissues. Furthermore, the knockdown of one such factor, Zfp28, influenced the core clock network. Mathematical modeling was able to predict putative regulator-effector interactions between the identified circadian genes and may help for investigations into the gene regulatory networks underlying circadian rhythms.

## INTRODUCTION

Organisms adapt to the 24-h day-night cycle, which leads to oscillations in physiology and behavior. This is coordinated by an intrinsic molecular clock originating from the interplay of two transcriptional-translational feedback loops [1]. In mammals, the core loop consists of the transcriptional activators Clock and Bmal1 and the repressors Period (Per) and cryptochrome (Cry). In a second loop, retinoid-related orphan receptors (ROR) activate transcription while Rev-Erb factors (Nr1d1 and Nr1d2) repress transcription [2]. These core clock components also regulate the expression of additional genes, possibly resulting in an oscillating transcription of these so-called clock-controlled genes and finally in the circadian phenotype.

Recent genome-wide studies of circadian time courses in various mouse tissues indicated that each tissue expresses its own particular set of cyclical genes, which only partly overlap each other [3-7]. Nearly half of all genes in the mouse genome show circadian oscillation in at least one tissue [7]. The basic mechanisms causing transcriptional rhythms of the core clock components are similar among all tissues, but how tissue-specific circadian output is achieved remains unknown, although several mechanisms have been proposed [2]. Among these are the use of tissue-specific transcription factors (TFs) or co-regulators [6, 8], and different temporal control of RNA polymerase II recruitment [9, 10], as well as defined rhythms in histone modifications accompanied by differential gene regulation [9-19].

To gain further insight into the transcriptional control of the circadian rhythm, we aimed to identify novel factors, particularly transcription factors and epigenetic regulators, which are under circadian control and might either feed back into the core clock network or are involved in establishing the circadian phenotype via downstream target regulation. Another set of genes with an emerging role in the regulation of transcription are long intergenic noncoding RNAs (lincRNAs). It is not yet fully understood how lincRNAs control gene expression, but case-specific studies suggest that they might act as a scaffold to target chromatin-modifying complexes and transcriptional regulatory proteins to the genome [20]. Recently, circadian expression of long noncoding RNAs was reported [7, 21].

To identify and investigate cyclically expressed factors, we synchronized NIH3T3 cells, a commonly used murine model system, in their circadian rhythm. We performed RNA-

sequencing (RNA-seq) to derive an extensive time course data set and applied three established as well as one novel computational approach to identify a number of transcription factors, epigenetic regulators, and lincRNAs that show cyclical expression. Among the identified TFs, we further investigated a cyclically expressed transcription factor, Zfp28, which influences the core clock network. Proposed interactions between transcriptional regulators and selected downstream targets could be tested using mathematical modeling based on ordinary differential equations, which we subsequently evaluated experimentally. Overall, the study revealed potential novel candidates involved in circadian rhythm control as well as possible interactions between them. Furthermore, the obtained data set provides an extensive resource for further circadian research.

# RESULTS

### Identification of genes with circadian expression in NIH 3T3 cells

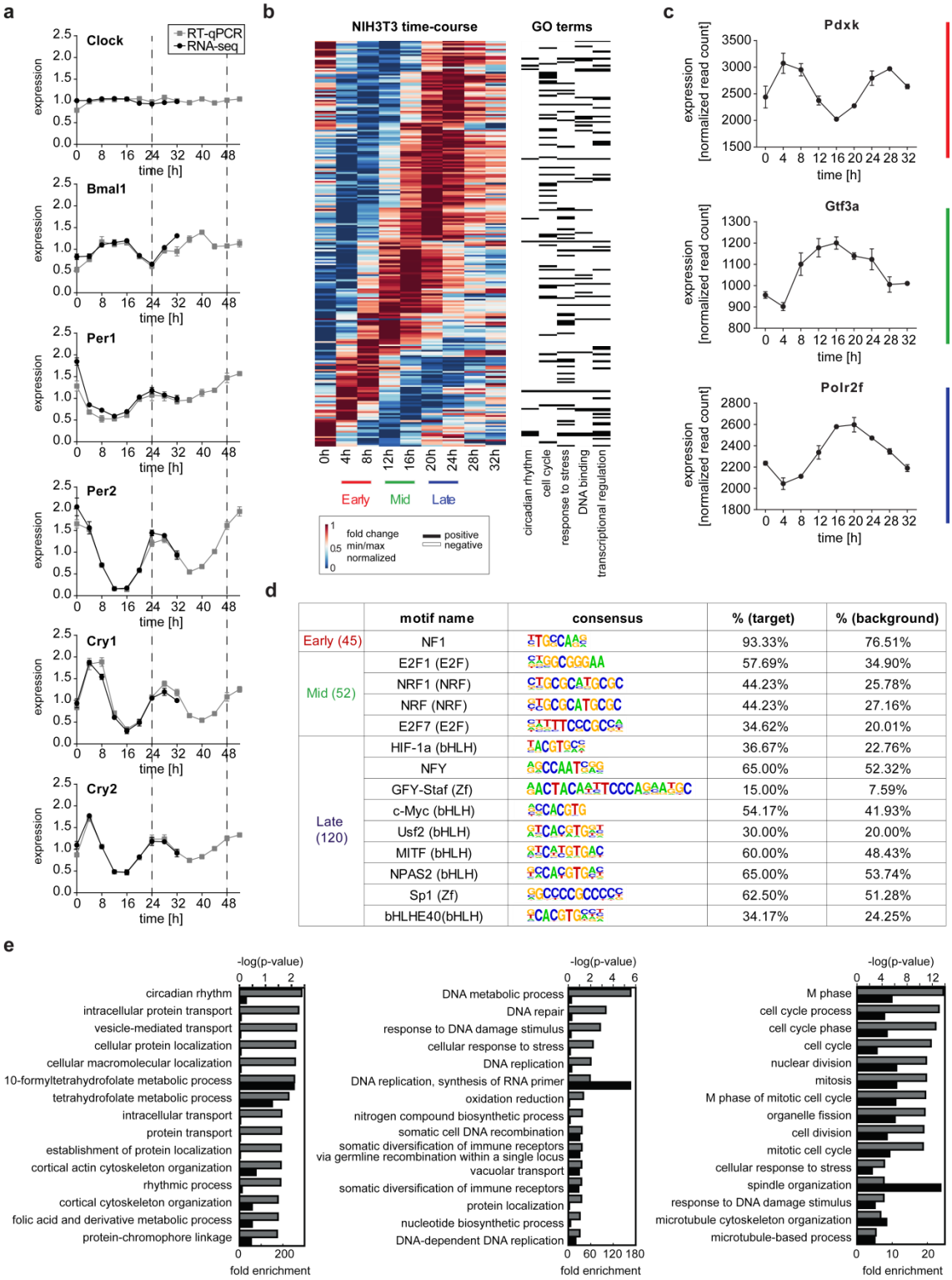
To study transcriptional regulation during circadian rhythm in a mammalian system, we employed NIH 3T3 cells synchronized with dexamethasone. This cell line has been routinely used as a model system for circadian rhythmicity [22-29]. NIH 3T3 Per2:luc cells showed a cyclical luminescence signal over time, thereby validating the synchronization protocol (see Fig. S1a in the supplemental material). Furthermore, we confirmed the cyclical expression pattern of known core clock genes in NIH 3T3 cells by using RT-quantitative PCR over two 24-h cycles. Consistent with previous studies [30], Clock showed nearly stable expression, whereas Cry and Per mRNAs showed cyclical expression peaking around 4 h, 28 h, and again at 52 h (Fig. 1a; see also Fig. S1b). Bmal1 (Arntl) expression peaked in between the Per and Cry maxima, at approximately 12 h and 36 h (Fig. 1a). We further performed genome-wide RNA expression analysis over a 32-h time course with 4-h resolution via high-throughput RNA sequencing (for the total and aligned number of reads, see Table S1 in the supplemental material). Previously, similar data sets were generated via microarray analysis [5, 24]; however, RNA-sequencing technology enables a deeper coverage and a more extensive analysis of the transcriptome [31]. Furthermore, it allows analysis of not-yet-annotated as well as noncoding RNAs. The data therefore adds substantially to the published microarray data sets.

Based on our RNA-seq data, half of the genes (50.5%, 11,082 genes) were expressed at least at one time point and were included for further analysis. The expression patterns of core clock genes (Clock, Cry, Per, and Bmal1 genes) coincided well with the RT-qPCR measurements (Fig. 1a). To identify further genes that show cyclical expression with a period length between 20 h and 28 h, we applied four different methods. JTK_CYCLE identifies oscillating genes in time-resolved gene expression data sets based on a non-parametric test (JT test) and a measure of rank correlation (Kendall's tau) combined in a Jonckheere-Terpstra-Kendall (JTK) algorithm [32]. By assigning a $q$ value of $<0.05$, the algorithm returned 328 cyclical transcripts (3% of all expressed genes) (see Fig. S1c in the supplemental material). Among the found cycling genes were established core clock components, such as the Period and cryptochrome genes, as well as the Bmal1, Nr1d1, and Nr1d2 genes. The algorithm RAIN, which is an extension of the JTK

algorithm but less stringent regarding the shape of the waveform, identified 4,063 cyclical genes with a period length between 20 h and 28 h ($q$ value <0.05) [33]. We further applied Arser, a combination of autoregressive spectrum analysis and harmonic regression, to identify cyclically expressed genes [34]. Arser detected 810 cyclically expressed genes (see Fig. S1c). As noted also by others, we found little overlap between the oscillating genes detected by Arser, JTK_CYCLE and RAIN [9, 35] (see Fig. S1d). We further developed our own approach (classification by nonlinear regression analysis [CBNLR]) to identify cyclical genes; this approach uses nonlinear regression to identify rhythmically expressed genes with a period length between 20 h and 28 h. In contrast to Arser, this method explicitly takes into account the experimental variability between biological replicates by using a bootstrapping approach (see Materials and Methods for details). In addition, by applying nonlinear regression analysis, the period length for identifying cyclical genes is not dependent on the sampling rate of the data set, which is the case for JTK_CYCLE or RAIN. All estimators of our approach are fitted freely and do not need to be fixed. This approach identified 292 cyclical expressed genes, a number similar to that determined with JTK_CYCLE (see Fig. S1c). Again, this approach identified most of the core clock components. The overall overlap of identified cyclical genes between all four different methods was limited (see Fig. S1d). A comparison of CBNLR to JTK_CYCLE and Arser using a benchmark data set showed competitive performance (see Text S1 in the supplemental material) and that, considering its different approach, it complements the existing methods in identifying cyclically expressed genes.

The common set of genes identified by all four methods consists of 49 genes and includes genes for known cyclical core clock components, such as Cry1, Cry2, Nr1d2, Per2, and Per3 (see Fig. S1d and e in the supplemental material). For further analysis, we selected high-confidence cyclical genes from our data set by considering only genes which were classified as cyclical by at least three of the four analysis methods. This resulted in a set of 230 genes (see Table S2 in the supplemental material). The expression patterns of these genes fall into various oscillation phases, as shown in the exemplary time courses (Fig. 1b and 1c).

When we compared the identified cyclical genes to the results of previous microarray studies performed in NIH 3T3 cells [5, 24], we found only minimal overlap between all studies, possibly due to differences in the synchronization protocols or classification




| | motif name | consensus | % (target) | % (background) |
|---|---|---|---|---|
| Early (45) | NF1 | | 93.33% | 76.51% |
| Mid (52) | E2F1 (E2F) | | 57.69% | 34.90% |
| | NRF1 (NRF) | | 44.23% | 25.78% |
| | NRF (NRF) | | 44.23% | 27.16% |
| | E2F7 (E2F) | | 34.62% | 20.01% |
| Late (120) | HIF-1a (bHLH) | | 36.67% | 22.76% |
| | NFY | | 65.00% | 52.32% |
| | GFY-Staf (Zf) | | 15.00% | 7.59% |
| | c-Myc (bHLH) | | 54.17% | 41.93% |
| | Usf2 (bHLH) | | 30.00% | 20.00% |
| | MITF (bHLH) | | 60.00% | 48.43% |
| | NPAS2 (bHLH) | | 65.00% | 53.74% |
| | Sp1 (Zf) | | 62.50% | 51.28% |
| | bHLHE40(bHLH) | | 34.17% | 24.25% |

***Figure 1: Characterization of circadian genes in NIH 3T3 cells.***

(a) Expression of core clock components as measured by RNA-seq data (*n*=2, normalized to average expression) or measured by RT-qPCR (*n*=4, normalized to Hsp90ab1 and to average expression; error bars represent standard errors of the means [SEM]). Dashed lines mark the 24-h and 48-h time points. (b) Heat map showing the expression of 230 cyclically expressed genes (detected by at least three methods). The expression values per gene were minimum-maximum normalized. Genes are sorted by the phase determined with JTK_CYCLE. GO terms associated with each gene are indicated in black on the right. (c) RNA-seq data for exemplary genes with circadian expression showing different peak times (*n*=2, normalized read counts, error bars represent SEM). Colored lines indicate the time point of their highest expression during the time course as marked in panel b (early, mid, and late). (d) Enriched sequence motifs in the promoters (±1 kb of TSS) of early, middle, and late peaking genes, performed with HOMER. The percentage of target as well as background sequences with motif are shown. (e) GO term enrichment for early, middle, and late peaking genes (top 15 are shown, based on *P* values). Grey bars indicate the negative logarithm of the *P* value (top axis); black bars show fold enrichment (bottom axis).

method of cyclical genes (see Fig. S1f in the supplemental material). The total overlap between the two microarray studies was restricted to only four core clock genes (Per2, Per3, Cry1, and Tef genes). Our data showed greater overlap with each of the microarray experiments, and our results also revealed the four above-mentioned genes as well as a number of core clock genes not identified by the microarray studies.

We next investigated, whether there was evidence that our identified cyclically expressed genes are under circadian control in other cell or tissue types. CircaDB, a database that compiles published transcriptome data of circadian time courses in different tissues and cellular systems, is a valuable resource for this purpose [36]. We found that 176 out of 230 cyclically expressed genes in our data set also oscillate in at least one other cell or tissue type (76 %) (see Fig. S1g. in the supplemental material). Moreover, we examined if the promoters of these genes are bound by core clock factors in different chromatin immunoprecipitation sequencing (ChIP-seq) studies performed in liver cells or macrophages [8, 9, 37-42]. Strikingly, 225 of these genes have been determined to be bound by at least one core clock factor (Clock, Per1, Per2, Cry1, Cry2, Npas2, Bmal1, Nr1d1, Nr1d2, or Rora; 97.8%, 1.36-fold increase above background; $P < 0.001$) (see Fig. S1h and i in the supplemental material), indicating that most of the 230 cyclically expressed genes might be under control of the circadian network. These genes include circadian regulators, stress response genes, and cell cycle genes (based on GO term enrichment analysis for biological processes), as well as many genes that bind to nucleotides and/or are implied to be transcriptional regulators (based on GO term enrichment analysis for molecular function) (Fig. 1b; see also Table S3 in the supplemental material).

Groups of genes with common functions may peak at certain phases of the circadian cycle. To analyze this, we performed gene set enrichment analysis (GSEA) for all cyclically expressed genes, sorted by their circadian phase [43, 44]. Interestingly, some GO terms showed enrichment among genes within a certain range of phases, while others displayed nearly uniform distribution (see Fig. S1j in the supplemental material). For example, genes related to metabolic processes were nearly evenly distributed throughout the time course, while genes involved in cell cycle tended to peak late in the circadian cycle (see Fig. S1j). To investigate this further, we determined the maximal expression of the cycling genes between 4 h and 24 h of the time course and divided them in early (maximal peak at 4 h or 8 h), middle (maximal peak at 12 h or 16 h), and

late (maximal peak at 20 h or 24 h) cycling genes (Fig. 1b and c). For these three sets of genes, we performed GO term analysis [45, 46] as well as motif prediction [47] (Fig. 1d and e; see also Tables S3 and S4 in the supplemental material). This revealed that genes related to circadian rhythm were mainly enriched in the set of genes showing high expression early in the time course. The genes of this class also exhibited enrichment of the NF1 motif in their promoter. Genes of the second category, peaking at 12 h or 16 h, were enriched for the E2F1, NRF, and E2F7 motifs and GO terms related to processes such as DNA metabolic process, DNA repair, cellular response to stress, and DNA replication. As observed by the GSEA analysis, the cycling genes peaking late were highly enriched for cell cycle related genes and, furthermore, for stress response genes. Motif analysis detected enriched HIF-1a, NFY, and the E-box-containing motifs c-MYC, USF2, MITF, NPAS2, and bHLHE40 in the promoters of these genes.

**Cyclical expression of transcriptional regulators and lincRNAs**

The GO term analysis indicated that in the set of cyclically expressed genes many genes are associated to transcriptional regulation (GO:0006355). Since the fate of a cell is largely defined by its transcriptome, we investigated the genes grouped under the GO terms "DNA binding" and "transcriptional regulation" in more detail. In addition, we screened the list of 230 cyclically expressed genes for putative transcription factors and epigenetic regulators, based either on prior knowledge from the literature or on the functional domains within the proteins. The combination of these genes resulted in a list of 70 putative transcriptional regulators with cyclical expression (Fig 2a). The list also contained genes previously studied in the context of circadian rhythm (e.g., Tef, Dbp, Nono, Mta1, Id1) [48-53].

It has been shown that the overlap of genes with circadian expression between different tissues is low [3-7]. However, if any of these factors plays a general role in the regulation of circadian rhythm, it should be cyclically expressed in most tissues. We therefore analyzed which of the 70 putative transcriptional regulators were also cyclically expressed in other tissues, again using the data sets provided by CircaDB [36]. Indeed, we found all core clock components and well-established players in circadian rhythm to be cyclically expressed in nearly all investigated tissues (Fig 2a, right column; see also Fig. S2a in the supplemental material). Furthermore, among the 70 potential transcriptional regulators, we found additional genes showing cyclical expression in

several tissues. This suggests that some of the identified factors might play a general role in the regulation of circadian rhythm or in circadian-regulated processes.

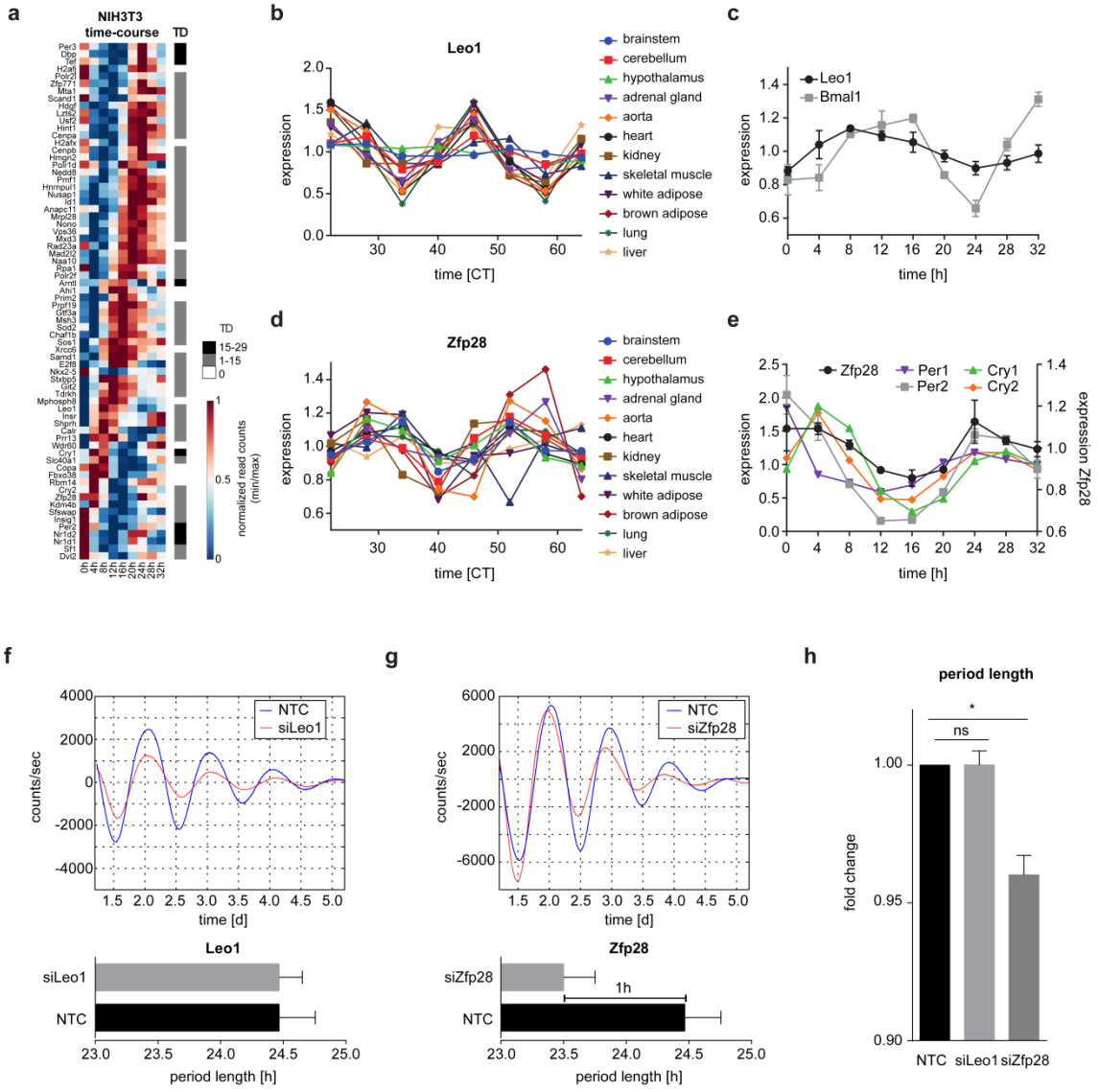



***Figure 2: Cyclically expressed transcription factors and epigenetic regulators can affect the core clock network.***

(a) Heat map of transcription factors, epigenetic regulators, and additional DNA-binding genes with circadian expression in NIH 3T3 cells (minimum-maximum normalized). Genes were sorted by the estimated phase given by JTK_CYCLE. The right column (TD) indicates in how many times out of 29 published time course data sets (obtained from CircaDB) a gene was classified by JTK_CYCLE to be cyclically expressed. White, in no data set; gray, in 1 to 15 data sets; black, in >15 data sets. (b) Expression of Leo1 in different mouse tissue time course experiments (data obtained from GEO, GSE54651). Data were normalized to average expression of Leo1 per tissue. (c) Expression of Leo1 compared to Bmal1 in the NIH 3T3 time course RNA-seq data. Normalized read counts of two replicates were normalized to average expression and are plotted with error bars (standard errors of the means [SEM]). (d) Expression of Zfp28 in different mouse tissue time course experiments (data obtained from GEO, GSE54651). Data were normalized to average expression of Zfp28 per tissue. (e) Expression of the Zfp28 compared to Cry and Per genes in NIH 3T3 time course RNA-seq data. Normalized read counts of two replicates were normalized to average expression and are plotted with error bars (SEM). (f) Luminescence measurements of NIH 3T3 Bmal1:luc cells under knockdown of Leo1 (siLeo1) and control (NTC) conditions (top). The bar graph shows the calculated period lengths of the luminescence signal under siLeo1 and NTC conditions of three independent replicates (bottom). (g) Luminescence measurements of NIH 3T3 Bmal1:luc cells under knockdown of Zfp28 (siZfp28) and control (NTC) conditions (top). The bar graph shows the calculated period lengths of the luminescence signal under siZfp28 and NTC conditions of three independent replicates (bottom). (h) Fold change in period length of knockdown cells (siLeo1, siZfp28) compared to control cells (NTC), measured by luminescence in NIH 3T3 Bmal1:luc cells ($n$=3). *, $P$ = 0.0318, Mann-Whitney U test.

For a more extensive characterization, we selected 2 candidates out of these 70 factors that had not been studied in detail with regard to circadian rhythm. Leo1 is a component of the PAF complex, which interacts with RNA polymerase II. Inspecting recently published transcriptome data of time courses in different mice tissues covering all germ layers, Leo1 showed circadian expression in 10 out of 12 tested mouse tissues (Fig. 2b) [7]. Interestingly, Leo1 expression, which we validated by RT-qPCR (see Fig. S2b in the supplemental material), is cyclically expressed in phase with Bmal1 in all of the 10 tissues, as well as in NIH 3T3 cells (Fig. 2c). Anafi et al. [54] provide a ranked list of 1,000 potential core clock components based on multiple metrics (cycling, phenotype, network interaction, ubiquity, and phylogenic conservation) derived from various data sets. Leo1 was ranked number 200 on this list of potential core clock factors.

An additional factor, the zinc finger protein Zfp28, showed cyclical expression in phase with downstream targets of Clock and Bmal1 (Fig. 2 d and e; see also Fig. S2b in the supplemental material). Although cyclical expression of Zfp28 in several tissues was clearly visible (Fig. 2d), CircaDB failed to identify it as such, given the standard parameters ($q$ value of <0.05) (see Fig. S2a). Perhaps this was a result of the low fold change in Zfp28 expression.

We performed knockdown experiments of Leo1 and Zfp28 in order to explore their impacts on the core clock network. While knockdown of Clock and/or Bmal1 abolished the cyclical expression of luciferase in NIH 3T3 Bmal1:luc cells (see Fig. S2c in the supplemental material), knockdown of Leo1 and Zfp28 did not suppress circadian cycling (Fig. 2f and g; see also Fig. S2d). Hence, Leo1 and Zfp28 are not essential for the circadian rhythm. However, Zfp28 knockdown decreased the period length of the circadian rhythm of Bmal1 by approximately 1 h (Fig. 2g and h; see also Fig. S2d), indicating a possible role of Zfp28 in modulating the core clock network. In contrast, Leo1 did not affect the period length of circadian Bmal1 cycling (Fig. 2f and h; see also Fig. S2d).

Recently, lincRNAs have emerged as critical regulators of gene expression [20, 55]. Having performed high-coverage RNA sequencing, we were able to investigate the expression patterns of lincRNAs (see Table S5 in the supplemental material). Again, we performed the analysis using JTK_CYCLE, RAIN, Arser, and our own approach. We identified around 130 to 400 cyclically expressed lincRNAs with each method (JTK_CYCLE identified 134; RAIN, 392; Arser, 161; CBNLR, 211) (Fig. 3 a and b; see also

Fig. S3a in the supplemental material); among them, 31 lincRNAs were detected by all four methods (see Fig. S3a and b).

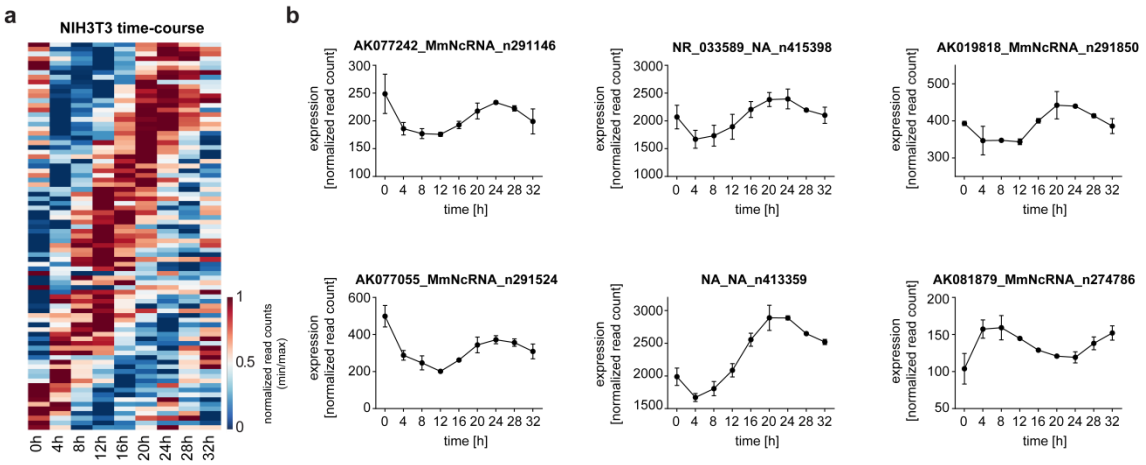



***Figure 3: Cyclically expressed lincRNAs.***

(a) Heat map showing expression of 83 lincRNAs detected to be cyclical by at least three out of four computational methods (minimum-maximum normalized). lincRNAs were sorted by the phase, determined by JTK_CYCLE. (b) RNA-seq expression data for six cycling lincRNAs ($n$=2; error bars represent standard errors of the means [SEM]).

## Mathematical modeling can predict putative interactions between circadian genes

Having identified a number of novel circadian genes, we next investigated by using a mathematical modeling approach if they could regulate each other. Previous theoretical work on the circadian clock established qualitative networks of clock-controlled genes [56, 57], or formulated mechanistic models of the core clock, mainly based on prior knowledge [24, 58-61]. In contrast, we focused on a network reconstruction approach, in which we asked in an unbiased way whether the cyclical genes identified from the RNA-seq data could, in principle, be regulated by one of the core clock factors or any other cyclically expressed transcription factor.

We formulated ordinary differential equation (ODE) models that used the mRNA expression time course of a putative regulator as input. The ODE models describe the translation of the putative regulator into protein, its nonlinear impact on the transcription of a putative target, and also the degradation of these molecular species (Fig. 4a; see also equation set 1 in Materials and Methods). For simplicity, we only considered one-to-one regulatory interactions between putative regulators and targets. The models were calibrated by fitting to the available RNA-seq data of target gene expression by minimizing the weighted least-squares measure ($X^2$ value) (see Materials and Methods), which is a measure of the difference between the data and simulated mRNA expression values. By evaluating the empirical distribution of the $X^2$ values, via a

parametric bootstrap approach we were able to reject models of regulation based on their calculated *P* value.

Since the transcriptional regulators Zfp28 and Leo1, which we studied in more detail, showed distinct phases of oscillation, we first considered regulation of these genes by core clock genes. Leo1 shows cyclical expression in phase with Bmal1 in many tissues. Similar to the expression of Bmal1, Leo1 could be regulated by ROR factors, since its promoter contains a RORE motif (AACTAGGTCA; 66bp upstream). Moreover, in liver of Rev-ErbA$^{-/-}$ mice, Leo1 is derepressed [39], suggesting that, additionally, Rev-ErbA (Nr1d1) represses the transcription of Leo1. Nr1d1, Nr1d2, and Rora have also been found to bind to the Leo1 promoter in liver cells or macrophages (see Table S2 in the supplemental material) [39-42]. Another possibility is that Leo1 is regulated by the cryptochrome genes, which have been shown to bind to the intron of Leo1 in liver cells [9]. Accordingly, we have designed ODE models of these interactions. Analysis of these models showed that Leo1 is possibly regulated by Cry1 but not by Cry2 alone (Fig. 4b; see also Table S6 in the supplemental material). Based on the *P* value of <0.05, regulation of Leo1 by Rev-ErbA (Nr1d1) without contributions from other factors is possible but not likely. Regulation of Leo1 by Rev-ErbB (Nr1d2) or Bmal1 alone was rejected. Indeed, in Bmal1 knockdown experiments, Leo1 expression was not affected (Fig. 4c).

Zfp28 is cyclically expressed with typical Bmal1 downstream targets, indicating that it might be regulated by Bmal1. Clock and Bmal1 were found to be enriched at the promoter of Zfp28 in liver [37, 38]; also, Nr1d1, Nr1d2, and Per2 showed binding to the Zfp28 promoter [9, 42]. The models considering regulation of Zfp28 by Bmal1, Cry1, Cry2, Nr1d1, or Nr1d2 could indeed not be rejected, implying that one of these core clock factors might establish transcriptional activation of Zfp28. Of all tested regulators, Bmal1 was considered the most likely (Fig. 4d; see also Table S6 in the supplemental material), and Bmal1 knockdown experiments revealed a significant downregulation of Zfp28 (Fig. 4e), showing regulation of Zfp28 by the core clock.

Encouraged by these promising predictions, we wanted to test further if the cyclical genes identified from the RNA-seq data could in principle be regulated by core clock factors. The transcription factor Bmal1, which is an essential factor that participates in both core clock feedback loops, was chosen as the possible input, and cyclical genes with a corresponding binding site in their promoter (E-Box; CCGGTCACGTGA) were

considered potential targets. Twenty-four of these 28 genes showed Bmal1 binding in their promoter in the liver [8, 9, 37, 38]. Also, Nr1d1 (Rev-ErbA), a core clock factor that exhibits a phase shift relative to Bmal1 and is part of the ROR/Bmal1/Rev-Erb feedback loop, was included as a putative regulator of genes containing a Rev-ErbA motif (GTAGGTCACTGGGTCA) in their promoter. Among these 20 genes, 15 have been found to be bound by Nr1d1 in liver tissue or macrophages (see Fig. S4a in the supplemental material) [40, 42]. Fourteen of 20 potential one-to-one interactions with Nr1d1 as the only regulator needed to be rejected (see Fig. S4a and Table S6). All of the remaining potential Nr1d1 targets showed Nr1d1 (Rev-ErbA) binding in at least one of the considered ChIP-seq studies.

Out of 28 potential Bmal1 targets, 14 regulator-effector interactions were rejected by our analyses with respect to the data and underlying model structure (Fig. 4f; see also Table S6 in the supplemental material). We evaluated the performance of the approach by performing knockdown of Bmal1, and found that 50% of potential Bmal1 targets were differentially downregulated, as predicted by the modeling approach (Fig. 4g; see also Fig. S4c in the supplemental material). Forty-five percent of rejected targets were also significantly deregulated upon Bmal1 knockdown. However, as this set contained mostly core clock genes, this was expected. Bmal1 is widely known as a transcriptional activator [2] and, consequently, all but one deregulated gene was downregulated. In accordance with an observation in Bmal1$^{-/-}$ mice [8], Cry1 was upregulated upon Bmal1 knockdown. However, the regulation of Cry1 by Bmal1 was rejected by the modeling analysis, reflecting the fact that additional factors cooperate with Bmal1 in regulation of Cry1 [27].

The Usf2 gene was identified as a cyclically expressed gene by three out of four methods. Interestingly, the Usf2 binding motif was also enriched in the set of identified circadian genes. In total, we found 87 occurrences of the Usf2 motif distributed among the promoters of 58 cyclically expressed genes. No preference for a specific phase, fold change, or period length was found among these 58 genes compared to genes that did not show the Usf2 motif (data not shown). To answer the question whether Usf2 alone could explain the expression data of each of the 58 genes containing a Usf2 motif in their promoter, we designed regulatory models for these genes with Usf2 as the main input. Following this analysis, 36 out of the 58 genes with a Usf2 binding motif in their

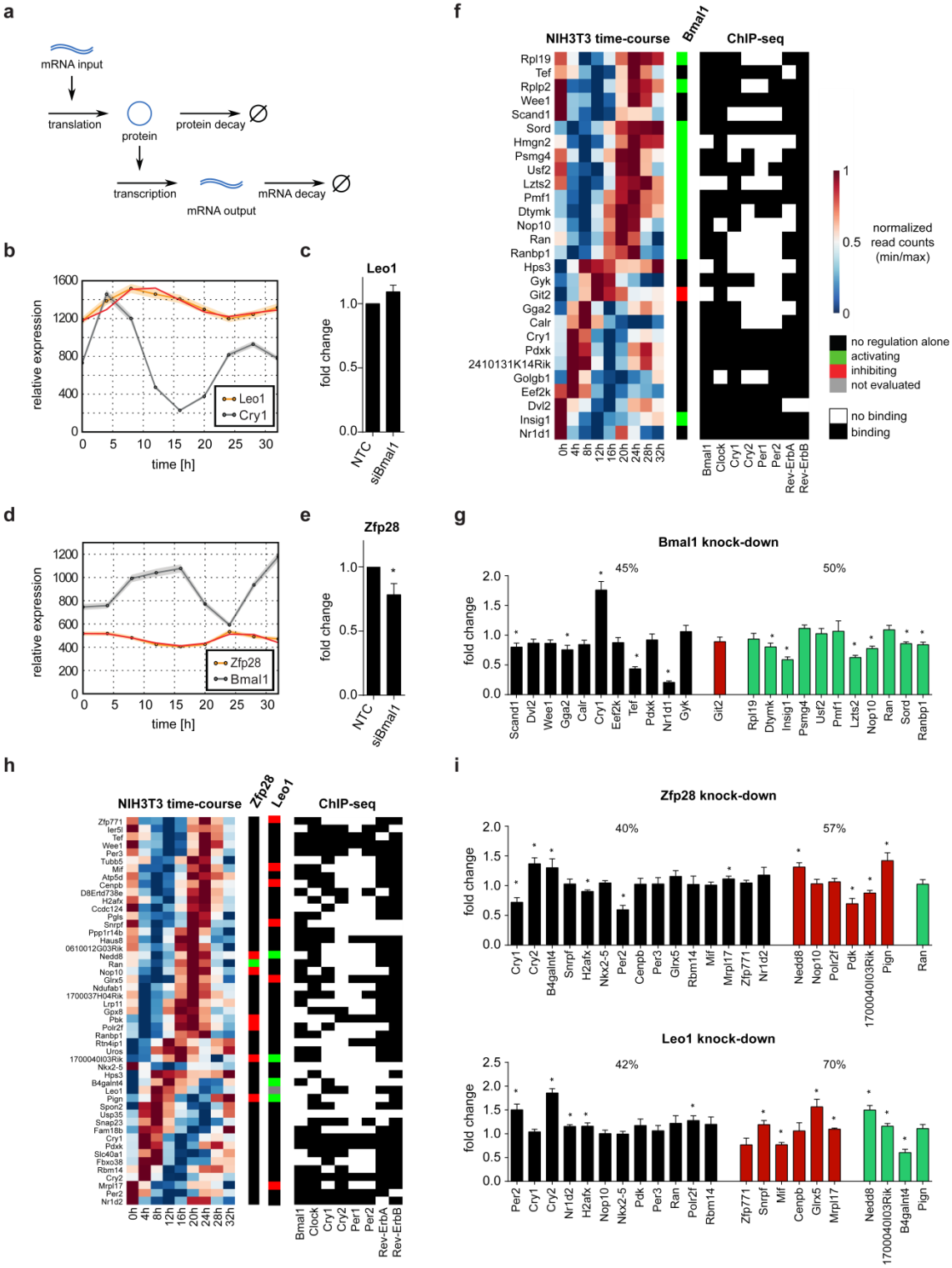



***Figure 4: Potential interactions tested by model evaluation.***

(a) Schematic representation of the model used for evaluating if a given target gene can be regulated by only one specific input gene. The model includes production of protein by input mRNA, protein decay, protein-dependent regulation of mRNA transcription, and mRNA decay. (b) RNA-seq data and simulation of the model considering Leo1 regulation by Cry1. Gray, expression of the input gene; orange, expression of the target gene and the corresponding standard deviation (shaded area). Red, the model simulation resulting from the best fit. (c) Fold change of Leo1 expression in NIH 3T3 Per2:luc cells treated with siBmal1 for 3 days, compared to control cells (NTC), measured by RT-qPCR ($n$=4, normalized to Hsp90ab1 and NTC control [$\Delta\Delta C_T$]). Error bars represent standard errors of the means (SEM). No significant change was observed (Mann-Whitney U-test). (d) RNA-seq data and simulation of the model considering Zfp28 regulation by Bmal1 (Arntl). Gray, expression of the input gene; orange, expression of the target gene and the corresponding standard deviation (shaded area). Red, the model simulation resulting from the best fit. (e) Fold change of Zfp28 expression in NIH 3T3 Per2:luc cells treated with siBmal1 for 3 days compared to control cells (NTC), measured by RT-qPCR ($n$=4, normalized to Hsp90ab1 and NTC control [$\Delta\Delta C_T$]). Error bars represent the SEM. A significant change was observed (Mann-Whitney U-test, $P$=0.011). (f) Heat map of potential Bmal1 targets (genes identified by three out of four methods as cyclical and containing an E-Box motif in their promoter, minimum-maximum normalized, sorted by the estimated phase given by JTK_CYCLE). The middle panel indicates whether the regulation of the gene by Bmal1

alone was rejected (black) or not (red/green). Red represents inhibition of the target gene by Bmal1, whereas green indicates activation. Columns on the right show binding of core clock factors to target genes (black) in at least one of the analyzed ChIP-seq data sets of the respective factor. (g) Fold change in expression of putative Bmal1 target genes in NIH 3T3 Per2:luc cells treated with siBmal1 for 3 days compared to control cells (NTC), measured by RT-qPCR ($n$=4, normalized to Hsp90ab1 and NTC control [$\Delta\Delta C_T$]). Error bars represent the SEM. Black bars indicate rejected targets, while green and red bars indicate inhibited or activated potential Bmal1 target genes, respectively. The fraction of deregulated genes in rejected (black) and potential target group (red and green) are given above the respective bars. Significance was assessed by a Mann-Whitney U-test; *, $P < 0.05$. (h) Heat map of potential Zfp28 or Leo1 targets (genes identified by all four methods as cyclical, minimum-maximum normalized, sorted by estimated phase given by JTK_CYCLE). The middle panels indicate whether the regulation of the target gene by Zfp28 or Leo1 alone was rejected (black) or not (red/green). Red represents inhibition of the target gene by the input factor, whereas green indicates activation. Grey indicates the model was not evaluated. Columns on the right show binding of core clock factors to target genes (black) in at least one of the analyzed ChIP-seq data sets of the respective factor. (i) Fold change in expression of putative Zfp28 (top) or Leo1 (bottom) target genes in NIH 3T3 Per2:luc cells treated with the respective siRNA for 3 days compared to control cells (NTC) measured by RT-qPCR ($n$=4, normalized to Hsp90ab1 and NTC control [$\Delta\Delta C_T$]). Error bars represent the SEM. Black bars indicate rejected targets, while green and red bars indicate inhibited or activated potential Zfp28 or Leo1 target genes, respectively. The fraction of deregulated genes in rejected (black) and potential target group (red and green) are given above the respective bars. Significance was assessed by a Mann-Whitney U-test; *, $P < 0.05$.

promoter could not be regulated by Usf2 alone, and additional regulators needed to be considered (see Fig. S4b and Table S6 in the supplemental material). In principle, the remaining 22 genes are potentially regulated by Usf2 only. These 22 genes are mainly distributed among the genes that peak late during the day-night cycle.

With regard to the transcriptional regulators Zfp28 and Leo1, which we have characterized in more detail, we do not possess information about putative binding sites. Therefore, we chose the 49 genes identified by all four methods as potential target genes of Leo1 or Zfp28 and tested which of these could possibly be explained by one of these factors alone. In the case of Zfp28, almost all putative targets (42 out of 49) could not be explained by Zfp28 alone, and additional regulators needed to be taken into account (Fig. 4h; see also Table S6 in the supplemental material). For Leo1, this number was only slightly higher; 10 out of the 49 putative targets can be explained by regulation of Leo1 alone (Fig. 4h; see also Table S6). Many potential Zfp28 or Leo1 putative targets showed binding of at least one core clock factors to their promoter, meaning that in principle they could also be directly regulated by the core clock network. However, in knockdown experiments of Zfp28 or Leo1, we found that indeed many predicted targets were deregulated (Fig. 4i; see also Fig. S4c in the supplemental material). Moreover, in the predicted target group, we found more deregulated genes than in the group of rejected genes, indicating the predictive power of our modeling approach.

## DISCUSSION

Establishment of circadian rhythm in cells is to a large extent dependent on transcriptional regulation. We therefore characterized circadian expression of coding and noncoding genes in NIH 3T3 fibroblast cells by time-resolved genome-wide RNA sequencing. Cyclically expressed genes were identified by four different computational approaches (three published and one new). Each method identified a different set of genes, which partly overlapped with each other. This was in agreement with earlier reports indicating that the detected cyclical genes in a data set depend partly on the applied approach [9, 35].

We found that cyclically expressed genes are involved in various biological processes, including circadian rhythm, stress response, and cell cycle regulation. Some groups of genes with a common functional annotation tend to peak in the same phase of the cycle. For example, cell cycle genes were preferentially enriched for the middle to late phases during the time course, consistent with the finding of enriched cell cycle regulator motifs in the promoters of these genes, e.g., E2F1, c-MYC, HIF-1a, and MITF [62-66]. The middle and late categories were also enriched for stress response genes, which is consistent with the motif enrichments of factors involved in stress responses, such as E2F7, HIF-1a, and SP1 [67-70]. Furthermore, gene promoters of each category were enriched for motifs related to circadian rhythm. NF1, detected in the early class, has been shown to play a role in circadian rhythm regulation in *Drosophila* [71]. Previous findings show that NRF1, enriched in the middle class, regulate numerous circadian regulatory genes in NIH 3T3 cells [72]. The transcription factor NF-Y, enriched in the late class, has been found to regulate the transcription of the core clock gene *Bmal1* [73]. Overall, GO terms and motifs assigned to each group of genes (early, middle, and late) are consistent with respect to their assigned function. Furthermore, the detected motifs are in agreement with identified motifs enriched in promoters of clock-controlled genes in other cell and tissue types [56, 74].

Among the cyclic genes, we found many potential transcription factors and epigenetic regulators. This emphasizes that transcriptional control is an important mechanism to regulate circadian rhythm. Two factors that were previously not implicated in circadian rhythmicity were characterized further: Leo1, a component of the PAF complex, displays cyclical expression in phase with Bmal1 in many tissues. Based on a RORA motif in its

promoter, derepression of Leo1 in liver of Rev-ErbA$^{-/-}$ mice [39], and published cryptochrome ChIP data [9], Leo1 could potentially be regulated by ROR, by Rev-Erb, and/or by cryptochrome factors. By the evaluation of potential one-to-one regulatory models, Leo1 can only be regulated by Rev-ErbA (Nr1d1) if additional factors are considered. However, a simple model exhibiting regulation of Leo1 by Cry1 alone could not be rejected by our analysis. The cyclical expression of Leo1 in many tissues implies a potential important role of Leo1 for the circadian phenotype. However, it is most likely not involved in the core clock network, as its knockdown did not affect cyclical expression of Bmal1. We also found that Bmal1 knockdown did not affect Leo1 expression. Leo1 has been shown to be important for the recruitment of the Paf1 complex to nucleosomes as well as for the quantity of the active histone mark H3K4me3 in yeast [75, 76], suggesting that Leo1 plays an essential role in transcriptional regulation. Its coordinated expression with Bmal1 further indicates that Leo1 might enhance transcriptional activation of clock-controlled genes by Clock and Bmal1.

The zinc finger protein Zfp28 has so far not been studied extensively. In humans, it shows strong expression in various adult tissues, but expression in embryonic tissue is development specific [77]. Apart from being cyclically expressed in NIH 3T3 cells, the cyclical behavior is also present in various tissues among different lineages. However, its amplitude is very low, which could explain why it is rarely detected as being cyclically expressed by computational approaches. Zfp28 cycles in phase with the classical downstream targets of Clock and Bmal1, such as Period and cryptochrome genes. As mentioned before, there is indication that in liver Zfp28 is bound at the promoter by Clock or Bmal1 [37, 38]. By model evaluation as well as experimentally, we showed that indeed Zfp28 is regulated by Bmal1. Knockdown of Zfp28 shortened the period of Bmal1 oscillations by around 1 h, indicating a potential role of this factor in the core clock regulatory network. To elucidate the mechanisms behind this, it would be interesting to investigate the genomic occupancy of Zfp28 by ChIP analysis and Zfp28 interaction partners by mass spectrometry.

We further identified cyclically expressed lincRNAs in our time course data. lincRNAs constitute a different layer of transcriptional regulation, but their specific role in circadian rhythm remains to be elucidated. For example, it seems that lincRNAs often

accomplish their gene regulatory function in *trans* at distal binding sites [20]. Therefore, it would be interesting to further investigate the binding sites of the cyclically expressed lincRNAs and their interaction partners, as well as their impact on the transcription of coding genes.

Mathematical models have been widely used in the field of circadian rhythms and are indeed very useful to our understanding of the core clock network and its impact on core clock genes. Several of these studies formulate mechanistic models of (parts of) the core clock and conceptually analyzed how different transcriptional regulatory modes among core clock TFs affect the dynamic behavior of clock-controlled genes [58, 59, 78, 79]. Others created a network description of the core clock genes based on large-scale promoter analyses [56] or based on a careful analysis of different TF knockout or mutant strain data sets as well as binding site predictions [57, 80]. Hence, most of these studies draw from prior information provided in the literature, are limited to a conceptual level, or are restricted to only a small number of clock genes. In the mathematical modeling approach developed in this study, however, we attempted to infer novel regulatory relationships among any of the cyclically expressed genes.

By formulating models of potential interactions, we were able to investigate regulator-effector relationships among circadian genes. This analysis revealed that in most cases, even though a gene contains a specific motif, such as Bmal1, Nr1d1, or Usf2, most likely more than one factor is involved in establishing the transcriptional output, as a large number of tested interactions had to be rejected. In addition, we have identified cyclical genes which are potentially regulated by Bmal1, Nr1d1, Usf2, Leo1, or Zfp28 alone. We evaluated this *in silico* approach by performing knockdown experiments and found that, on average, more potential targets were deregulated compared to targets which needed to be rejected by our computational analysis.

These interactions are modeled with mRNA expression as input and consider only constitutive translation and protein decay; however, circadian regulation has also been described to occur on the posttranscriptional and posttranslational level [25, 81-93]. Therefore, further studies should aim to quantify degradation and production rates of mRNA and/or protein or assess protein levels in circadian rhythm, as done in several studies [90, 92, 93], and consequently expand and improve the modeling approach [94-96].

Overall, we demonstrated that with this data set we could identify transcriptional and epigenetic regulators of circadian rhythm with potential general roles during circadian rhythm across lineages. Further, we proposed a modeling approach to identify potential regulatory pairs of circadian genes, which we validated experimentally.

# MATERIALS AND METHODS

### Cell Culture, synchronization, small interfering RNA (siRNA) knockdown, and luminescence measurement

NIH 3T3 Per2:luc cells (kindly provided by Hiroki R. Ueda [29]) were cultured at 37°C, 5 or 7% $CO_2$, and in high humidity. Culture medium was Dulbecco's modified Eagle's medium (DMEM; catalog number 21969035; Life Technologies) supplemented with 10% fetal bovine serum (catalog number 10270106; Life Technologies), 2 mM L-glutamine (catalog number 25030024; Life Technologies), and 1x nonessential amino acids (catalog number 11140-035; Life Technologies). Two days after seeding the cells at a density of 3,000 cells/$cm^2$, the culture medium was removed and changed to medium containing 100 nM dexamethasone (catalog number D4902; Sigma-Aldrich) and, 2 h later, medium was changed back to normal medium (time point 0 h) [27, 88, 97-103].

For luminescence measurements, NIH 3T3 Bmal1:luc cells were generated by transfecting NIH 3T3 cells (CRL-1658; ATCC) with a plasmid expressing Bmal1:luc (Bmal1:luc-pT2A) and a Tol2 transposase-expressing plasmid (pCAGGS-TP) (plasmids were kindly provided by K. Yagita [104]) followed by selection with 250 µg/ml hygromycine. Two days before the start of the measurement, cells were seeded at a density of 3,000 cells/$cm^2$ and at the same time transfected with 30 pmol ON-TARGETplus SMARTpool siRNA (i.e., a mixture of four siRNAs targeting the same gene) against the gene of interest or nontargeting control (NTC; Dharmacon) per 3.5-cm plate by using Lipofectamine RNAiMAX (catalog number 13778150; Invitrogen). After 2 days, cells were synchronized with dexamethasone as described above, and after 2 h medium was changed to DMEM without phenol red (catalog number 31053-028; Life Technologies) containing 10% fetal bovine serum, 2 mM L-glutamine, 1x nonessential amino acids, 1x sodium pyruvate (catalog number 11360-039; Life Technologies), 25 mM HEPES (catalog number H0887; Sigma), and 500 µM beetle luciferin (catalog number E1601; Promega), and cells were transfected again with siRNA. Luminescence measurements were performed with a LumiCycle 32 (ActiMetrics) and analyzed with the LumiCycle analysis program, version 2.3. A running average of 24 h was used for the baseline fit, and the period length was calculated by fitting a damped sine wave to the data.

For the mathematical model evaluation, cells were seeded at a density of 3,000 cells/cm$^2$ and at the same time transfected with 30 pmol ON-TARGETplus SMARTpool siRNA (i.e., a mixture of four siRNA targeting the same gene) (Dharmacon) per well of a 6-well plate by using Lipofectamine RNAiMAX (catalog number 13778150; Invitrogen). After 3 days, cells were harvested for further analysis.

The statistical significance of deregulation observed in the experiments was tested using a Mann–Whitney *U* test (with significance set at a *P* value of <0.05).

**Quantitative reverse transcription-PCR**

Total RNA was prepared using TRIzol (catalog number 15596; Invitrogen) or a Purelink RNA Mini Kit (catalog number 12183025; Life Technologies) and reverse transcribed with a First-Strand cDNA synthesis kit (catalog number K1612; Fermentas). Transcripts were quantified by PCR using SYBR green PCR master mix (catalog number 4334973; Life Technologies) on a ViiA7 PCR machine (Life Technologies).

**RNA-seq data analysis**

RNA-seq data for multiple time points of the circadian cycle were generated in biological duplicates by using Illumina sequencing [50-bp reads; poly(A) RNA sequencing, nonstrand specific; Illumina HiSeq 2000]. Reads were aligned to the mouse genome (mm9) by using TopHat (version 2.0.9) with default options (for the total and aligned number of reads, see Table S1 in the supplemental material) [105]. Normalized read counts for each samples were calculated using the DESeq package after applying library size normalization [106]. Library size-normalized .wig files were created for all the samples by using the QuasR package [107]. The RNA-seq data were deposited with NCBI's Gene Expression Omnibus database [108]. Only genes with a normalized read count above 64 at any of the time points were considered for further analyses. The lincRNA annotations were collected from NonCode [109], ENSEMBL [110], UCSC [111] and other published resources [112]. Normalized read counts for each sample were generated using the DESeq package after applying library size normalization [106]. Only lincRNAs expressed above 32 normalized read count at any of the time points and which were found to be cyclically expressed by three out of four methods were considered for further analysis. This list of lincRNAs as well as a list of cyclically

expressed lincRNAs with an expression value above 10 at any of the time points are provided in Table S5 in the supplemental material.

**Identification of cyclically expressed genes in synchronized NIH 3T3 cells**

We identified cyclically expressed genes by using four different methods: JTK_CYCLE [32], RAIN [33], Arser [34], and also a novel method based on nonlinear optimization (CBNLR). The published methods were applied as described in the respective manuals. For all four methods, genes with a period between 20 h and 28 h were selected to be cyclical. Additional criteria for selection were a false-discovery rate below 0.05 for JTK_CYCLE, RAIN, and Arser.

In the CBNLR approach (classification by nonlinear regression analysis), a sine function dependent on three parameters (amplitude, period length, and phase shift) is fitted to the detrended time course data of each gene by using multistart local optimization by Levenberg-Marquardt (5 different initial conditions). From this analysis, we selected only genes with an estimated period length between 20 h and 28 h. As nonlinear optimization alone did not yield any information on significance levels, we classified genes based on the robustness of the estimated period length *(T)*. Based on the variability between measurements, we created 1,000 nonparametric bootstrap samples for each gene with a period length between 20 h and 28 h. Bootstraps were generated by resampling each data point based on a normal distribution with a standard deviation corresponding to the experimental error, and models were again fitted to each bootstrapped data set. The result is a distribution of parameters over the bootstrap samples. For the classification of cyclical genes, we only selected genes with a mean period *(T)* between 20 h and 28 h and a small relative error. A thorough comparison of the CBNLR method and two other methods for classification of cyclical genes using benchmark data can be found in Text S1 in the supplemental material.

**GO term analysis**

Functional annotation tables were generated using the Gene Ontology (GO) browser supplied by the Mouse Genome Database (MGD; Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, ME) [113]. GO terms included were GO:0007623~circadian rhythm, "GO:0006950~response to stress, GO:0007049~cell

cycle, GO:0003677~DNA binding, and GO:000635~regulation of transcription, DNA-templated, all with corresponding child terms.

GO term analysis for all cyclical genes and subgroups were performed using the DAVID functional annotation tool [45, 46].

Gene set enrichment analysis (GSEA) was performed by sorting the cyclical genes according to their phase, which was determined with JTK_CYCLE [43, 44] (http://www.broadinstitute.org/gsea/index.jsp).

**Motif analysis**

Motif analysis was carried out using the HOMER software (http://homer.salk.edu/homer/motif/) [47]. A promoter region was defined as bp -1000 to +1000 around the transcription start site. The cutoff values for the different promoters were set to the default values given by the HOMER program.

**Tissue data and CircaDB**

Tissue data were obtained from reference [7]. Cyclical genes of these and other published tissue data sets, as well as cell line data sets, were obtained from CircaDB database (http://bioinf.itmat.upenn.edu/circa/) [36].

**Analysis of ChIP-seq data**

ChIP-seq data were downloaded from the Gene Expression Omnibus (GEO) database (NCBI). Fastq files were processed using Bowtie (version 0.12.7) with default parameters for alignment to mouse genome mm9 (available from UCSC) [114]. Positions of promoters (-800 to +200 with respect to the transcription start site [TSS]) were determined using the GenomicFeatures R bioconductor package. Promoter enrichments over inputs were determined by using the same method as described by Schick et al. [115]. Only peaks showing an enrichment 1.5-fold above input were considered for further analysis.

The data sets used were the following: for Bmal1 ZT08, Clock ZT08, and Cry1 ZT20 in liver, GSE53828 (corresponding inputs at ZT08 and ZT20 were kindly obtained from the authors for that GEO entry); for the Bmal1 time course from ZT02 to ZT22 in liver and associated input, GSE26602; for Bmal1 ZT08, Clock ZT08 and input in liver, GSE36916; for Rora in liver at ZT22, GSE59486 (associated input at ZT22, GSE26345); for Rev-ErbB

in liver at ZT10 and ZT22, GSE36375 (associated inputs at ZT10 and ZT22, GSE26345); for Rev-ErbA in liver, Rev-ErbB in liver, and associated input, GSE34020; for Rev-ErbA_macrophage, Rev-ErbB_macrophage, and associated input, GSE45914.

**Model evaluation**

We used differential equations to model pairwise regulation of one target gene by a single input gene (u). The model equations included factors for protein synthesis of the input gene ($y_{\mathrm{prot}}$) as well as for transcription of the target gene ($y_{\mathrm{mRNA}}$):

$$\begin{aligned}\frac{dy_{\mathrm{prot}}}{dt} &= R^{TL}u(t) - \lambda^{\mathrm{prot}}y_{\mathrm{prot}} \\ \frac{dy_{\mathrm{mRNA}}}{dt} &= R^{TC}\frac{y_{\mathrm{prot}}^{h}}{y_{\mathrm{prot}}^{h}+K^{h}} - \lambda^{\mathrm{mRNA}}y_{\mathrm{mRNA}}\end{aligned} \qquad (1)$$

with the initial conditions

$$\begin{aligned}y_{\mathrm{prot}}^{0} &= y_{\mathrm{prot}}(t=0) = \frac{R^{TL}y_{\mathrm{mRNA}}}{\lambda^{\mathrm{prot}}} \\ y_{\mathrm{mRNA}}^{0} &= y_{\mathrm{mRNA}}(t=0)\end{aligned} \qquad (2)$$

The concentration u(t) of the input gene between measurements is given by the linear interpolation between data points. $R^{TL}$ and $R^{TC}$ denote the translation rate of the input protein and transcription rate of the target mRNA, respectively. λ denotes decay rates of either target mRNA or input protein. *K* is the input protein concentration at which the target gene promoter is half-saturated. h denotes the Hill coefficient. Ranges considered are summarized in Table 1. Note that a repressive interaction can be modeled by a negative Hill coefficient.

Assuming steady state at t=0, $R^{TC}$ can be expressed as a function of the other parameters. Models were fit to the data by minimizing the $X^2$-value, defined as follows:

$$\sum \frac{\left(y^{\mathrm{data}}-y^{\mathrm{model}}\right)^2}{\sigma^2} \qquad (3)$$

where σ corresponds to the error in measurement. As an optimization algorithm, we used the multistart Levenberg-Marquardt method from 100 initial conditions sampled via Latin hypercube sampling.

In order to evaluate models, the empirical distribution of the $X^2$ values needs to be calculated. Therefore, bootstrap samples for each target gene's expression were generated from a linear error model, where we adopted linear regression to find the best fit between expression values and corresponding standard deviations. Each data point was resampled by assuming independent and normally distributed noise with

mean zero and variance $\sigma^2$. A total of 1000 bootstrap samples were generated in this manner and refitted to obtain the distribution of $X^2$ values. By calculating the empirical cumulative density function, we were able to extract a *P* value of the initial fit (original $X^2$ value). A right-sided test with a cutoff *P* value of <0.05 served as rejection criteria for the proposed models [116].

**Table 1. Parameters used for model evaluation**

| Parameter | Lower limit | Higher limit |
|---|---|---|
| *K* | 1.00E-03 | 1.00E+06 |
| h | -5 | 5 |
| $\lambda^{prot}$ (h) | 0.15 | 48.0 |
| $\lambda^{mRNA}$ (h) | 0.15 | 48.0 |
| $R^{TL}$ | 1.00E-03 | 1.00E+06 |

**Nucleotide sequence accession number**

The RNA-seq data have been deposited in NCBI's GEO database [108] and are accessible through series accession number GSE66243.

## ACKNOWLEDGEMENTS

We thank members of the Tiwari and Legewie labs for their cooperation and critical feedback during the project. We gratefully acknowledge support from the Core Facilities of the Institute of Molecular Biology (IMB), Mainz, especially from the Genomics Core Facility. We are thankful to John Hogenesch (University of Pennsylvania, USA) for his help in extracting data from CircaDB. We also thank Hiroki R. Udea (RIKEN Center for Developmental Biology, Japan) for providing the NIH 3T3 Per2:luc cells and Kazuhiro Yagita (Kyoto Prefectural University of Medecine, Japan) for providing Bmal1:luc-pT2A and pCAGGS-TP plasmid.

## AUTHOR CONTRIBUTIONS

S.S. performed experiments (all experimental work, supported during the luminescence experiments by M.H.H.) and wrote the manuscript. K.B. performed computational analysis and wrote the manuscript. S.T. and D.F. performed computational analysis. M.H.H. supported the luminescence experiments and data acquisition. S.L. and V.K.T. designed the study, analyzed data and wrote the manuscript. All authors read and approved the final manuscript.

## FUNDING INFORMATION

Wilhelm Sander Stiftung provided funding to Sandra Schick, Sudhir Thakurela, David Fournier, and Vijay K. Tiwari under the grant number 2012.009.2. EpiGeneSys RISE1 program provided funding to Sandra Schick, Sudhir Thakurela, David Fournier, and Vijay K. Tiwari. Marie Curie Fellows Association provided funding to Sandra Schick, Sudhir Thakurela, David Fournier, and Vijay K. Tiwari under the grant number CIG 322210. German Ministry of Research (BMBF) provided funding to Kolja Becker and Stefan Legewie. Deutsche Forschungsgemeinschaft (DFG) provided funding to Sandra Schick, Sudhir Thakurela, David Fournier, and Vijay K. Tiwari under the grant number TI 799/1-1. Research of David Fournier is supported by the Focus Program Translational Neurosciences Mainz (FTN).

# REFERENCES


1. Lowrey, P.L. and J.S. Takahashi, *Genetics of circadian rhythms in Mammalian model organisms.* Adv Genet, 2011. **74**: p. 175-230.
2. Partch, C.L., C.B. Green, and J.S. Takahashi, *Molecular architecture of the mammalian circadian clock.* Trends Cell Biol, 2014. **24**(2): p. 90-9.
3. Panda, S., et al., *Coordinated transcription of key pathways in the mouse by the circadian clock.* Cell, 2002. **109**(3): p. 307-20.
4. Storch, K.F., et al., *Extensive and divergent circadian gene expression in liver and heart.* Nature, 2002. **417**(6884): p. 78-83.
5. Menger, G.J., et al., *Circadian profiling of the transcriptome in NIH/3T3 fibroblasts: comparison with rhythmic gene expression in SCN2.2 cells and the rat SCN.* Physiol Genomics, 2007. **29**(3): p. 280-9.
6. Miller, B.H., et al., *Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation.* Proc Natl Acad Sci U S A, 2007. **104**(9): p. 3342-7.
7. Zhang, R., et al., *A circadian gene expression atlas in mammals: implications for biology and medicine.* Proc Natl Acad Sci U S A, 2014. **111**(45): p. 16219-24.
8. Menet, J.S., et al., *Nascent-Seq reveals novel features of mouse circadian transcriptional regulation.* Elife, 2012. **1**: p. e00011.
9. Koike, N., et al., *Transcriptional architecture and chromatin landscape of the core circadian clock in mammals.* Science, 2012. **338**(6105): p. 349-54.
10. Le Martelot, G., et al., *Genome-wide RNA polymerase II profiles and RNA accumulation reveal kinetics of transcription and associated epigenetic changes during diurnal cycles.* PLoS Biol, 2012. **10**(11): p. e1001442.
11. Etchegaray, J.P., et al., *Rhythmic histone acetylation underlies transcription in the mammalian circadian clock.* Nature, 2003. **421**(6919): p. 177-82.
12. Doi, M., J. Hirayama, and P. Sassone-Corsi, *Circadian regulator CLOCK is a histone acetyltransferase.* Cell, 2006. **125**(3): p. 497-508.
13. Etchegaray, J.P., et al., *The polycomb group protein EZH2 is required for mammalian circadian clock function.* J Biol Chem, 2006. **281**(30): p. 21209-15.
14. Nakahata, Y., et al., *The NAD+-dependent deacetylase SIRT1 modulates CLOCK-mediated chromatin remodeling and circadian control.* Cell, 2008. **134**(2): p. 329-40.
15. Jones, M.A., et al., *Jumonji domain protein JMJD5 functions in both the plant and human circadian systems.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21623-8.
16. DiTacchio, L., et al., *Histone lysine demethylase JARID1a activates CLOCK-BMAL1 and influences the circadian clock.* Science, 2011. **333**(6051): p. 1881-5.
17. Masri, S., et al., *The circadian clock transcriptional complex: metabolic feedback intersects with epigenetic control.* Ann N Y Acad Sci, 2012. **1264**: p. 103-9.
18. Vollmers, C., et al., *Circadian oscillations of protein-coding and regulatory RNAs in a highly dynamic mammalian liver epigenome.* Cell Metab, 2012. **16**(6): p. 833-45.
19. Valekunja, U.K., et al., *Histone methyltransferase MLL3 contributes to genome-scale circadian transcription.* Proc Natl Acad Sci U S A, 2013. **110**(4): p. 1554-9.
20. Vance, K.W. and C.P. Ponting, *Transcriptional regulatory functions of nuclear long noncoding RNAs.* Trends Genet, 2014. **30**(8): p. 348-55.
21. Coon, S.L., et al., *Circadian changes in long noncoding RNAs in the pineal gland.* Proc Natl Acad Sci U S A, 2012. **109**(33): p. 13319-24.
22. Nagoshi, E., et al., *Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells.* Cell, 2004. **119**(5): p. 693-705.
23. Nagoshi, E., et al., *Circadian gene expression in cultured cells.* Methods Enzymol, 2005. **393**: p. 543-57.
24. Hughes, M.E., et al., *Harmonics of circadian gene transcription in mammals.* PLoS Genet, 2009. **5**(4): p. e1000442.
25. Morf, J., et al., *Cold-inducible RNA-binding protein modulates circadian gene expression posttranscriptionally.* Science, 2012. **338**(6105): p. 379-83.
26. Engelen, E., et al., *Mammalian TIMELESS is involved in period determination and DNA damage-dependent phase advancing of the circadian clock.* PLoS One, 2013. **8**(2): p. e56623.
27. Ma, Y.T., et al., *O-GlcNAcylation of BMAL1 regulates circadian rhythms in NIH3T3 fibroblasts.* Biochem Biophys Res Commun, 2013. **431**(3): p. 382-7.
28. Bieler, J., et al., *Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells.* Mol Syst Biol, 2014. **10**: p. 739.

29. Isojima, Y., et al., *CKIepsilon/delta-dependent phosphorylation is a temperature-insensitive, period-determining process in the mammalian circadian clock.* Proc Natl Acad Sci U S A, 2009. **106**(37): p. 15744-9.
30. Hamilton, E.E. and S.A. Kay, *SnapShot: circadian clock proteins.* Cell, 2008. **135**(2): p. 368-368 e1.
31. Zhao, S., et al., *Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells.* PLoS One, 2014. **9**(1): p. e78644.
32. Hughes, M.E., J.B. Hogenesch, and K. Kornacker, *JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets.* J Biol Rhythms, 2010. **25**(5): p. 372-80.
33. Sakata, R., et al., *Long-term effects of the rain exposure shortly after the atomic bombings in Hiroshima and Nagasaki.* Radiat Res, 2014. **182**(6): p. 599-606.
34. Yang, R. and Z. Su, *Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation.* Bioinformatics, 2010. **26**(12): p. i168-74.
35. Doherty, C.J. and S.A. Kay, *Circadian control of global gene expression patterns.* Annu Rev Genet, 2010. **44**: p. 419-44.
36. Pizarro, A., et al., *CircaDB: a database of mammalian circadian gene expression profiles.* Nucleic Acids Res, 2013. **41**(Database issue): p. D1009-13.
37. Rey, G., et al., *Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver.* PLoS Biol, 2011. **9**(2): p. e1000595.
38. Annayev, Y., et al., *Gene model 129 (Gm129) encodes a novel transcriptional repressor that modulates circadian gene expression.* J Biol Chem, 2014. **289**(8): p. 5013-24.
39. Fang, B., et al., *Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo.* Cell, 2014. **159**(5): p. 1140-52.
40. Cho, H., et al., *Regulation of circadian behaviour and metabolism by REV-ERB-alpha and REV-ERB-beta.* Nature, 2012. **485**(7396): p. 123-7.
41. Bugge, A., et al., *Rev-erbalpha and Rev-erbbeta coordinately protect the circadian clock and normal metabolic function.* Genes Dev, 2012. **26**(7): p. 657-67.
42. Lam, M.T., et al., *Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription.* Nature, 2013. **498**(7455): p. 511-5.
43. Mootha, V.K., et al., *PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.* Nat Genet, 2003. **34**(3): p. 267-73.
44. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
45. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic Acids Res, 2009. **37**(1): p. 1-13.
46. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.
47. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Molecular Cell, 2010. **38**(4): p. 576-89.
48. Wuarin, J. and U. Schibler, *Expression of the liver-enriched transcriptional activator protein DBP follows a stringent circadian rhythm.* Cell, 1990. **63**(6): p. 1257-66.
49. Fonjallaz, P., et al., *The two PAR leucine zipper proteins, TEF and DBP, display similar circadian and tissue-specific expression, but have different target promoter preferences.* EMBO J, 1996. **15**(2): p. 351-62.
50. Humphries, A., et al., *cDNA array analysis of pineal gene expression reveals circadian rhythmicity of the dominant negative helix-loop-helix protein-encoding gene, Id-1.* J Neuroendocrinol, 2002. **14**(2): p. 101-8.
51. Brown, S.A., et al., *PERIOD1-associated proteins modulate the negative limb of the mammalian circadian oscillator.* Science, 2005. **308**(5722): p. 693-6.
52. Li, D.Q., et al., *Metastasis-associated protein 1 is an integral component of the circadian molecular machinery.* Nat Commun, 2013. **4**: p. 2545.
53. Kowalska, E., et al., *NONO couples the circadian clock to the cell cycle.* Proc Natl Acad Sci U S A, 2013. **110**(5): p. 1592-9.
54. Anafi, R.C., et al., *Machine learning helps identify CHRONO as a circadian clock component.* PLoS Biol, 2014. **12**(4): p. e1001840.
55. Kornienko, A.E., et al., *Gene regulation by the act of long non-coding RNA transcription.* BMC Biol, 2013. **11**: p. 59.
56. Bozek, K., et al., *Regulation of clock-controlled genes in mammals.* PLoS One, 2009. **4**(3): p. e4882.
57. Yan, J., et al., *Analysis of gene regulatory networks in the mammalian circadian rhythm.* PLoS Comput Biol, 2008. **4**(10): p. e1000193.
58. Ueda, H.R., et al., *System-level identification of transcriptional circuits underlying mammalian circadian clocks.* Nat Genet, 2005. **37**(2): p. 187-92.

59. Korencic, A., et al., *The interplay of cis-regulatory elements rules circadian rhythms in mouse liver.* PLoS One, 2012. **7**(11): p. e46835.
60. Tovin, A., et al., *Systematic identification of rhythmic genes reveals camk1gb as a new element in the circadian clockwork.* PLoS Genet, 2012. **8**(12): p. e1003116.
61. Guilding, C., et al., *A riot of rhythms: neuronal and glial circadian oscillators in the mediobasal hypothalamus.* Mol Brain, 2009. **2**: p. 28.
62. Amati, B., K. Alevizopoulos, and J. Vlach, *Myc and the cell cycle.* Front Biosci, 1998. **3**: p. d250-68.
63. Carreira, S., et al., *Mitf cooperates with Rb1 and activates p21Cip1 expression to regulate cell cycle progression.* Nature, 2005. **433**(7027): p. 764-9.
64. Goda, N., et al., *Hypoxia-inducible factor 1alpha is essential for cell cycle arrest during hypoxia.* Mol Cell Biol, 2003. **23**(1): p. 359-69.
65. Koshiji, M., et al., *HIF-1alpha induces cell cycle arrest by functionally counteracting Myc.* EMBO J, 2004. **23**(9): p. 1949-56.
66. Johnson, D.G., et al., *Expression of transcription factor E2F1 induces quiescent cells to enter S phase.* Nature, 1993. **365**(6444): p. 349-52.
67. Zalmas, L.P., et al., *DNA-damage response control of E2F7 and E2F8.* EMBO Rep, 2008. **9**(3): p. 252-9.
68. Majmundar, A.J., W.J. Wong, and M.C. Simon, *Hypoxia-inducible factors and the response to hypoxic stress.* Molecular Cell, 2010. **40**(2): p. 294-309.
69. Carvajal, L.A., et al., *E2F7, a novel target, is up-regulated by p53 and mediates DNA damage-dependent transcriptional repression.* Genes Dev, 2012. **26**(14): p. 1533-45.
70. Li, H., et al., *Integrated high-throughput analysis identifies Sp1 as a crucial determinant of p53-mediated apoptosis.* Cell Death Differ, 2014. **21**(9): p. 1493-502.
71. Williams, J.A., et al., *A circadian output in Drosophila mediated by neurofibromatosis-1 and Ras/MAPK.* Science, 2001. **293**(5538): p. 2251-6.
72. Zhu, W., *Mechanisms and Functional Roles of Nuclear Respiratory Factor 1 (NRF1) Binding Sites in the Human Genome*, in *PhD thesis*2011. http://d-scholarship.pitt.edu/8605/, PhD thesis, University of Pittsburgh.
73. Xiao, J., et al., *Transcription factor NF-Y is a functional regulator of the transcription of core clock gene Bmal1.* J Biol Chem, 2013. **288**(44): p. 31930-6.
74. Bozek, K., et al., *Circadian transcription in liver.* Biosystems, 2010. **102**(1): p. 61-9.
75. Dermody, J.L. and S. Buratowski, *Leo1 subunit of the yeast paf1 complex binds RNA and contributes to complex recruitment.* J Biol Chem, 2010. **285**(44): p. 33671-9.
76. Chu, X., et al., *Structural insights into Paf1 complex assembly and histone binding.* Nucleic Acids Res, 2013. **41**(22): p. 10619-29.
77. Zhou, L., et al., *Identification and characterization of two novel zinc finger genes, ZNF359 and ZFP28, in human development.* Biochem Biophys Res Commun, 2002. **295**(4): p. 862-8.
78. Westermark, P.O. and H. Herzel, *Mechanism for 12 hr rhythm generation by the circadian clock.* Cell Rep, 2013. **3**(4): p. 1228-38.
79. Korencic, A., et al., *Timing of circadian genes in mammalian tissues.* Sci Rep, 2014. **4**: p. 5782.
80. Relogio, A., et al., *Tuning the mammalian circadian clock: robust synergy of two loops.* PLoS Comput Biol, 2011. **7**(12): p. e1002309.
81. Kim, S.H., et al., *Rhythmic control of mRNA stability modulates circadian amplitude of mouse Period3 mRNA.* J Neurochem, 2015. **132**(6): p. 642-56.
82. Chen, W., et al., *Regulation of Drosophila circadian rhythms by miRNA let-7 is mediated by a regulatory cycle.* Nat Commun, 2014. **5**: p. 5549.
83. Kojima, S. and C.B. Green, *Circadian genomics reveal a role for post-transcriptional regulation in mammals.* Biochemistry, 2015. **54**(2): p. 124-33.
84. Beckwith, E.J. and M.J. Yanovsky, *Circadian regulation of gene expression: at the crossroads of transcriptional and post-transcriptional regulatory networks.* Curr Opin Genet Dev, 2014. **27**: p. 35-42.
85. Reddy, A.B. and G. Rey, *Metabolic and nontranscriptional circadian clocks: eukaryotes.* Annu Rev Biochem, 2014. **83**: p. 165-89.
86. Ki, Y., et al., *Warming Up Your Tick-Tock: Temperature-Dependent Regulation of Circadian Clocks.* Neuroscientist, 2015.
87. Woo, K.C., et al., *Mouse period 2 mRNA circadian oscillation is modulated by PTB-mediated rhythmic mRNA degradation.* Nucleic Acids Res, 2009. **37**(1): p. 26-37.
88. Kwak, E., T.D. Kim, and K.T. Kim, *Essential role of 3'-untranslated region-mediated mRNA decay in circadian oscillations of mouse Period3 mRNA.* J Biol Chem, 2006. **281**(28): p. 19100-6.

89. Woo, K.C., et al., *Circadian amplitude of cryptochrome 1 is modulated by mRNA stability regulation via cytoplasmic hnRNP D oscillation.* Mol Cell Biol, 2010. **30**(1): p. 197-205.

90. Luck, S., et al., *Rhythmic degradation explains and unifies circadian transcriptome and proteome data.* Cell Rep, 2014. **9**(2): p. 741-51.

91. Mauvoisin, D., et al., *Proteomics and circadian rhythms: it's all about signaling!* Proteomics, 2015. **15**(2-3): p. 310-7.

92. Mauvoisin, D., et al., *Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver.* Proc Natl Acad Sci U S A, 2014. **111**(1): p. 167-72.

93. Robles, M.S., J. Cox, and M. Mann, *In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism.* PLoS Genet, 2014. **10**(1): p. e1004047.

94. Becker, K., et al., *Reverse-engineering post-transcriptional regulation of gap genes in Drosophila melanogaster.* PLoS Comput Biol, 2013. **9**(10): p. e1003281.

95. Schwanhausser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-42.

96. de Sousa Abreu, R., et al., *Global signatures of protein and mRNA expression levels.* Mol Biosyst, 2009. **5**(12): p. 1512-26.

97. Lee, K.H., et al., *Rhythmic interaction between Period1 mRNA and hnRNP Q leads to circadian time-dependent translation.* Mol Cell Biol, 2012. **32**(3): p. 717-28.

98. Takarada, T., et al., *Clock genes influence gene expression in growth plate and endochondral ossification in mice.* J Biol Chem, 2012. **287**(43): p. 36081-95.

99. Lee, K.H., et al., *AUF1 contributes to Cryptochrome1 mRNA degradation and rhythmic translation.* Nucleic Acids Res, 2014. **42**(6): p. 3590-606.

100. Onishi, Y. and Y. Kawano, *Rhythmic binding of Topoisomerase I impacts on the transcription of Bmal1 and circadian period.* Nucleic Acids Res, 2012. **40**(19): p. 9482-92.

101. Izumo, M., et al., *Quantitative analyses of circadian gene expression in mammalian cell cultures.* PLoS Comput Biol, 2006. **2**(10): p. e136.

102. Balsalobre, A., et al., *Resetting of circadian time in peripheral tissues by glucocorticoid signaling.* Science, 2000. **289**(5488): p. 2344-7.

103. Goriki, A., et al., *A novel protein, CHRONO, functions as a core component of the mammalian circadian clock.* PLoS Biol, 2014. **12**(4): p. e1001839.

104. Yagita, K., et al., *Development of the circadian oscillator during differentiation of mouse embryonic stem cells in vitro.* Proc Natl Acad Sci U S A, 2010. **107**(8): p. 3846-51.

105. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics, 2009. **25**(9): p. 1105-11.

106. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.

107. Gaidatzis, D., et al., *QuasR: quantification and annotation of short reads in R.* Bioinformatics, 2014.

108. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.* Nucleic Acids Res, 2002. **30**(1): p. 207-10.

109. Liu, C., et al., *NONCODE: an integrated knowledge database of non-coding RNAs.* Nucleic Acids Res, 2005. **33**(Database issue): p. D112-5.

110. Flicek, P., et al., *Ensembl 2014.* Nucleic Acids Res, 2014. **42**(Database issue): p. D749-55.

111. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.

112. Ramos, A.D., et al., *Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny in vivo.* Cell Stem Cell, 2013. **12**(5): p. 616-28.

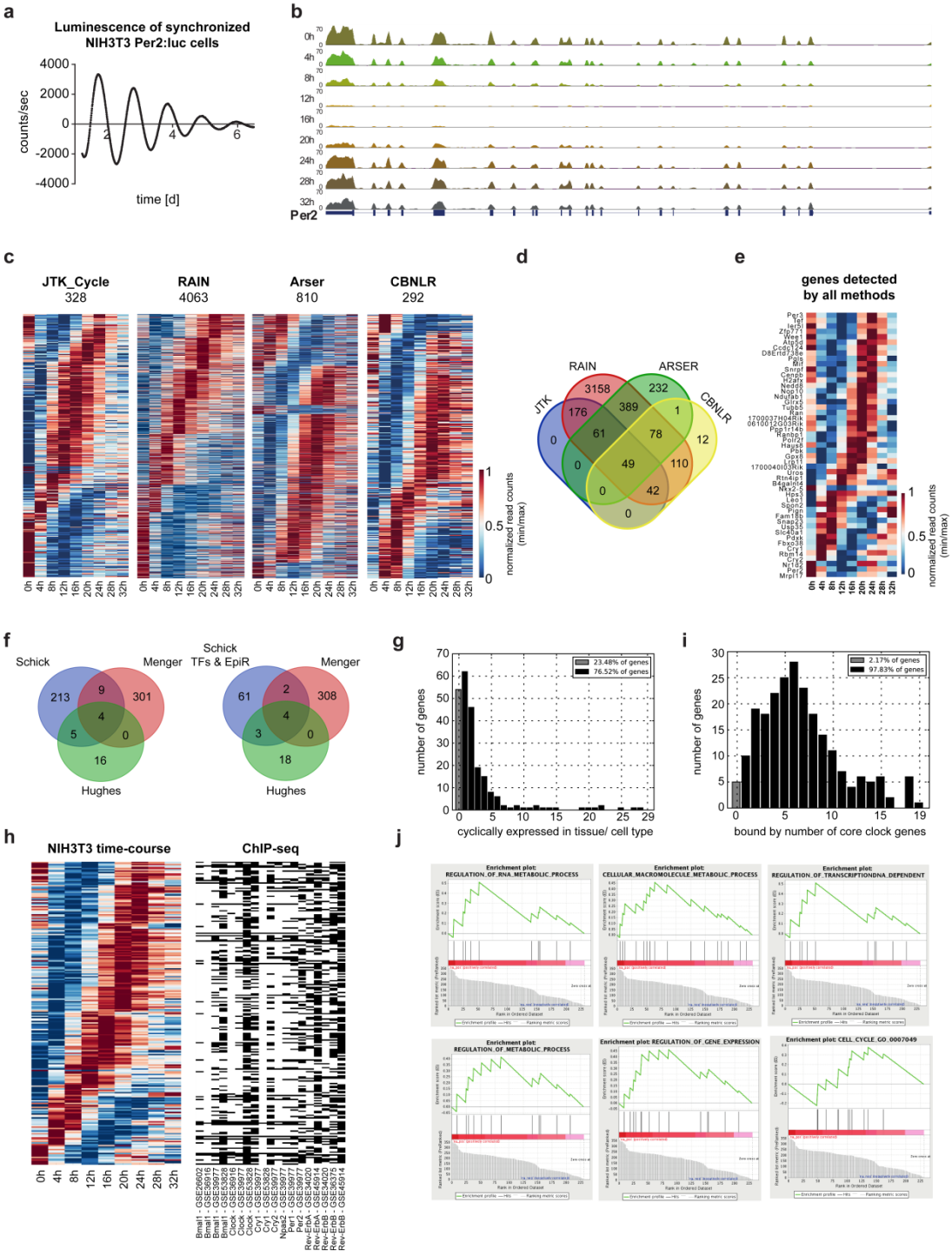
113. Blake, J.A., et al., *The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse.* Nucleic Acids Res, 2014. **42**(Database issue): p. D810-7.

114. Langmead, B., *Aligning short sequencing reads with Bowtie.* Curr Protoc Bioinformatics, 2010. **Chapter 11**: p. Unit 11 7.

115. Schick, S., et al., *Dynamics of chromatin accessibility and epigenetic state in response to UV damage.* J Cell Sci, 2015.

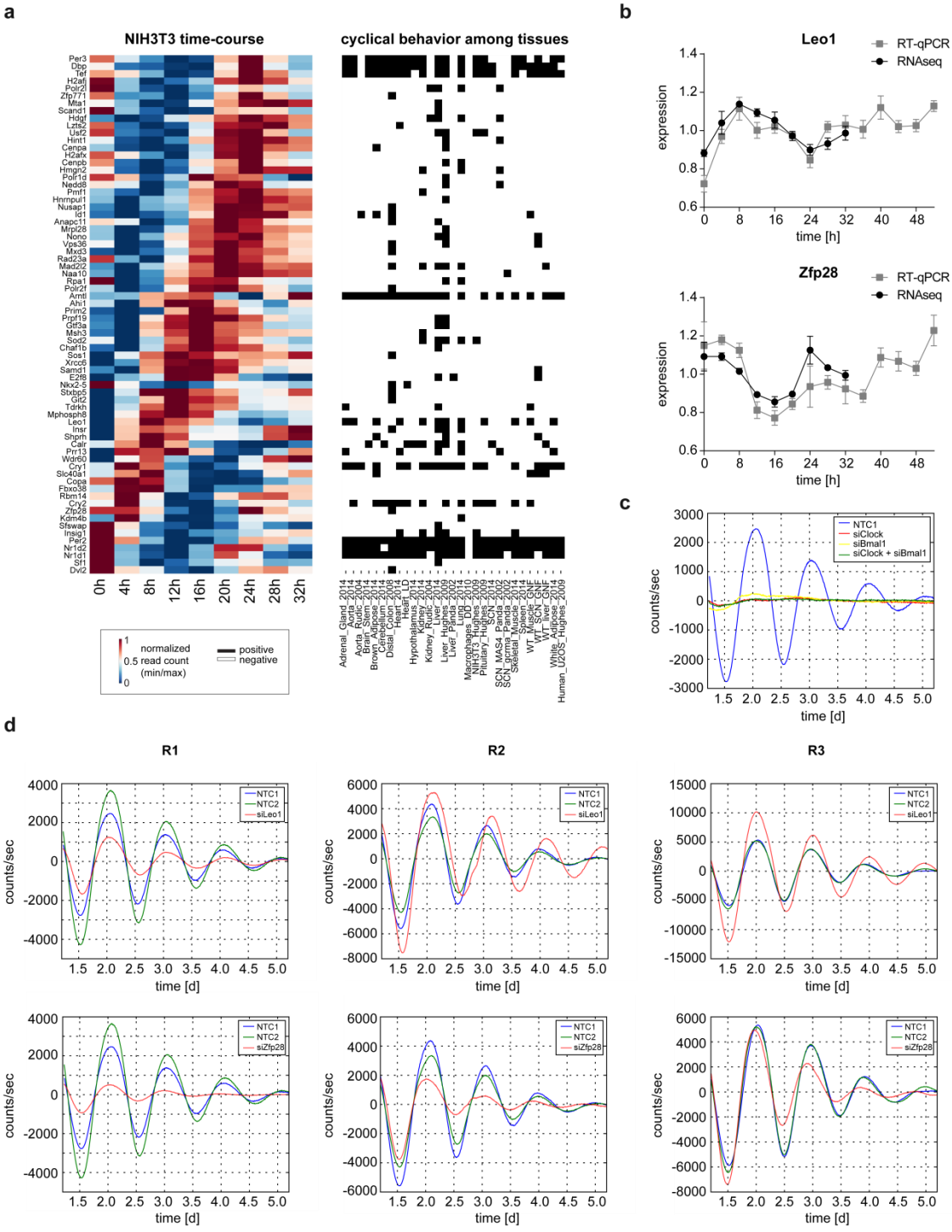
116. Johansson, R., P. Stralfors, and G. Cedersund, *Combining test statistics and models in bootstrapped model rejection: it is a balancing act.* BMC Syst Biol, 2014. **8**: p. 46.

# SUPPLEMENTARY FIGURES



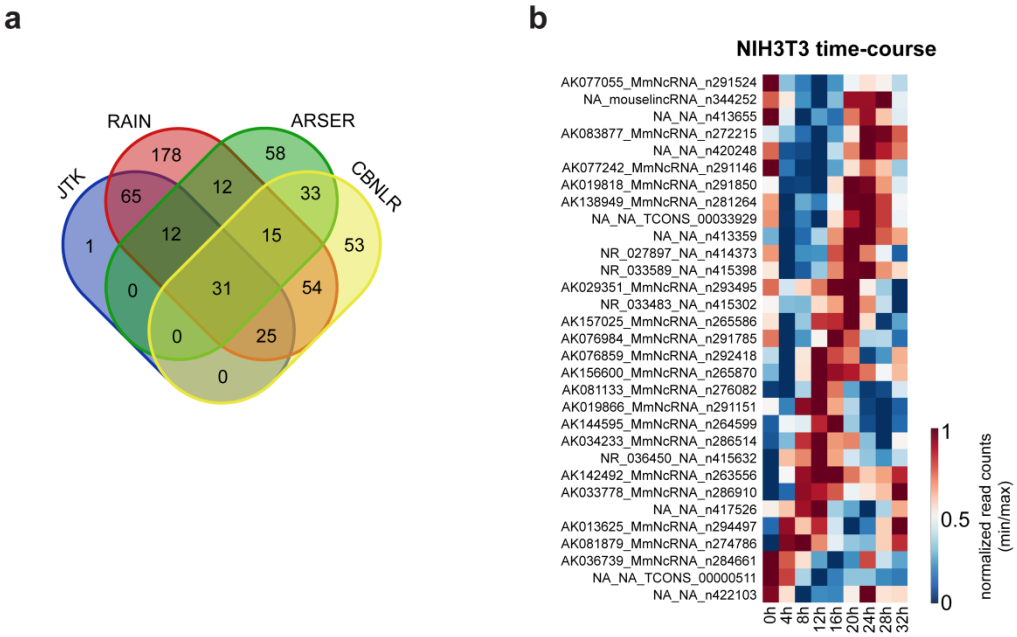

***Supplementary Figure 1: Identification of circadianly expressed genes in NIH3T3 cells.***

a) Background-subtracted luminescence signal of dexamethasone synchronized NIH3T3 Per2:luc cells. b) UCSC genome browser view of one replicate of the RNAseq data at the Per2 locus. c) Heat map of cyclically expressed genes after minimum-maximum normalization (blue – lowest expression, red – highest expression) detected by either JTK_CYCLE, RAIN, Arser or the CBNLR approach. Genes are sorted according to the phase given by the corresponding method. d) Venn diagram of cyclically expressed genes detected by JTK-Cycle, RAIN, Arser and the CBNLR approach. e) Heat map of cyclically expressed genes detected with all four methods after minimum-maximum normalization (blue – lowest expression, red – highest expression). Genes are sorted by phase given by JTK_CYCLE. f) Venn diagrams comparing cyclically expressed genes found in NIH3T3 cells in Menger et al. 2007, Hughes et al. 2009 and this study (left: genes detected by three out of four applied methods, right: transcription factors and epigenetic regulators (TF and EpiR)). g) Bar plot showing the number of genes identified by three out of four methods found to be cyclically expressed in different cell or tissue types. Black bars indicate genes cyclically expressed in at least one cell or tissue type. h) Heat map of cyclically expressed genes detected by three out of four methods after minimum-maximum normalization. Columns on the right indicate binding of a core clock factor in the promoter (-800, +200 around TSS) of the respective gene. i) Bar plot showing the number of genes identified by three out of four methods found to be bound by at least one core clock factor (black) compared to genes not bound (grey). j) Examples of Gene Set Enrichment Analysis results. Genes are sorted by the phase determined by JTK_CYCLE, with highest phase value on the left.



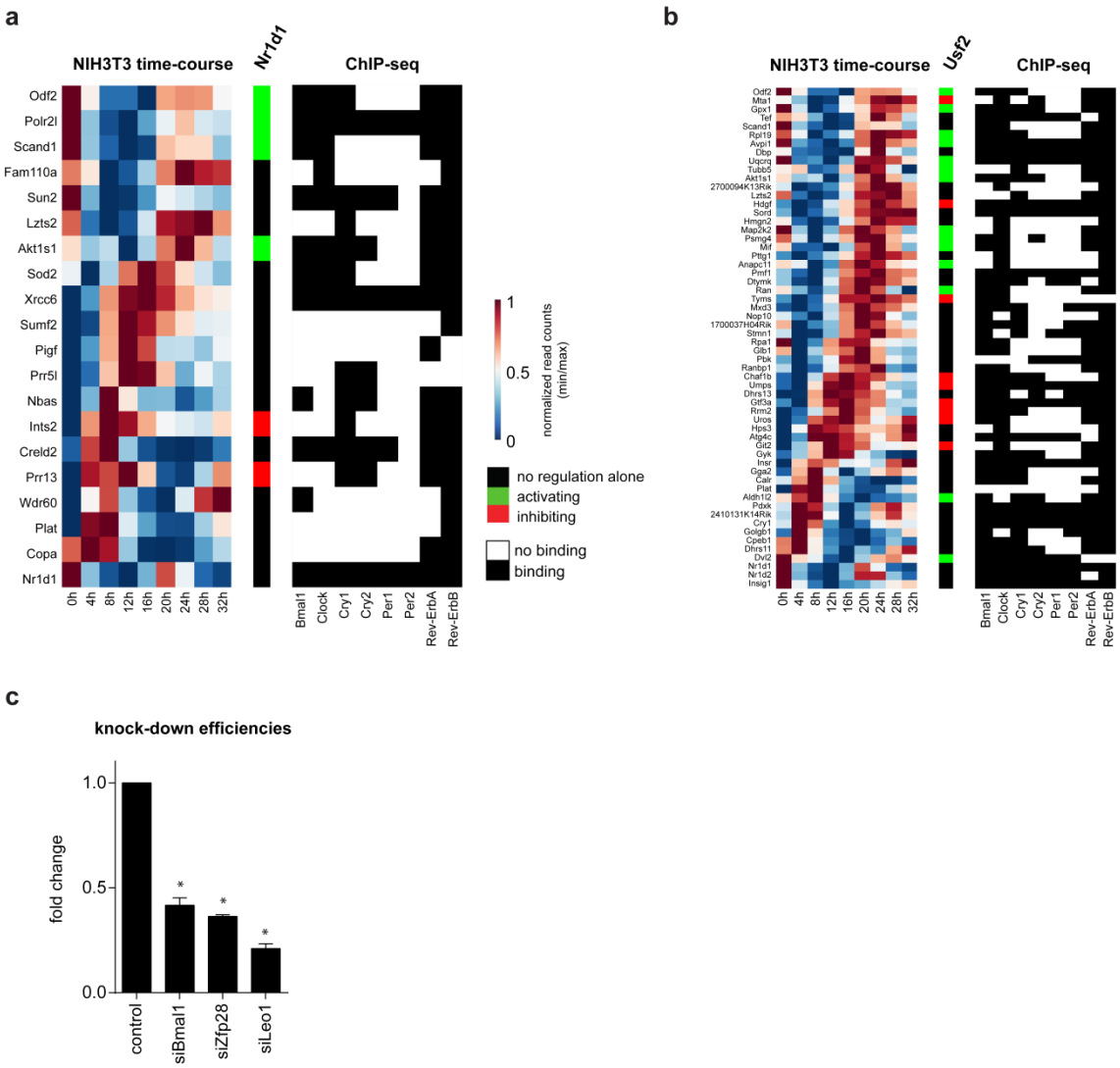

***Supplementary Figure 2: Many transcriptional regulators are also identified as cyclically expressed in other tissue and cell types.***

a) Heat map of cyclically expressed transcription factors, epigenetic regulators and other DNA-binding genes in NIH3T3 cells (minimum-maximum normalized, blue – lowest expression, red – highest expression). Genes are sorted by the estimated phase given by JTK_CYCLE. On the right, occurrence of this gene to be cyclically expressed in 29 published time course data (black- detected as cycling in this data set). Tissue and cell line data was collected from CircaDB. b) Expression of Leo1 (top) and Zfp28 (bottom) determined by RNA-seq (n=2, normalized read counts normalized to average expression, error bars represent s.e.m.) and RT-qPCR (n=4, relative expression normalized to Hsp90ab1 and to the average expression: ΔΔCT, error bars represent s.e.m.). c) Luminescence measurement of NIH3T3 Bmal1:luc cells with knock-down of Clock (siClock), Bmal1 (siBmal1), or Clock + Bmal1 (siClock + siBmal1), respectively. Data is background subtracted. d) Luminescence measurement of NIH3T3 Bmal1:luc cells with knock-down of Leo1 or Zfp28 of three replicates (R1-R3), respectively. Data is background subtracted.




***Supplementary Figure 3: Identification of circadianly expressed lincRNAs in NIH3T3 cells.***

a) Venn diagram showing cyclically expressed long intergenic non-coding RNAs (lincRNAs) detected by JTK_CYCLE, RAIN, Arser, and the CBNLR approach. b) Heat map of cyclically expressed lincRNAs detected by all four methods (minimum-maximum normalization, blue – lowest expression, red – highest expression).




***Supplementary Figure 4: Model evaluation of potential Nr1d1 or Usf2 targets.***

a) Heat map of potential Nr1d1 (Rev-ErbA) targets (genes identified by three out of four methods as cyclical and containing a Rev-ErbA motif in their promoter, minimum-maximum normalized, sorted by estimated phase given by JTK_CYCLE). The middle panel indicates whether the regulation of this gene by Nr1d1 alone was rejected (black) or not (red/green). Red represents inhibition of the target gene by Nr1d1, whereas green indicates activation. Columns on the right show binding of core clock factors to target genes (black) in at least one of the analyzed ChIP-seq data sets of the respective factor. b) Heat map of potential Usf2 targets (genes identified by three out of four methods as cyclical and containing an Usf2 motif in their promoter, minimum-maximum normalized, sorted by estimated phase given by JTK_CYCLE). The middle panel indicates whether the regulation of this gene by Usf2 alone was rejected (black) or not (red/green). Red represents inhibition of the target gene by Usf2, whereas green indicates activation. Columns on the right show binding of core clock factors to target genes (black) in at least one of the analyzed ChIP-seq data sets of the respective factor. c) Bar graphs showing the expression of the respective target gene after knock-down with siRNA in NIH3T3 Per2:luc cells measured by RT-qPCR. The expression is normalized to Hsp90ab1 as well as to the expression in the control cells (NTC) (ΔΔCT, n=4, error bars represent s.e.m.). Significance was assessed by a Mann-Whitney U-test (*: $p<0.05$).

**S1 Text Evaluation of methods for classification of cyclical genes**

In order to evaluate CLNBR, the proposed method of classifying circadian genes, and compare it to the existing ones used in this paper we have constructed a benchmark dataset according to Yang & Su 2011: 5000 time-courses for both stationary signals (non-decaying) and non-stationary signals (decaying in signal amplitude and signal strength) were constructed with a signal-to noise ratio ranging from 1 to 5. The parameters chosen for data generation were taken as in Yang & Su 2011. Test datasets A and B were constructed by combining stationary or non-stationary periodic signals with a white-noise background (non-periodic signals) respectively. In test datasets C and D the respective periodic signals were combined with a background composed of autoregressive processes of order AR(1). In the AR(1)-background model the expression value at a certain time is linearly dependent on previous expression values. It is assumed that this more realistically captures the autocorrelation properties of time-course expression data. Timing and amount of measurements were simulated according to the time points of the RNAseq experiment presented in this study. Two realizations of the noise were generated and the mean and standard deviation calculated in order to simulate experimental duplicates.

The Matthews correlation coefficient (MCC), a measure of the performance of binary classifications, was calculated for each dataset and all three methods evaluated (Fig ST1). Although CBNLR performs less well compared to the other two methods on datasets with a white-noise background (Fig ST2 and ST3), it shows better performance than Arser on datasets C and D, which are the respective datasets with an AR(1)-background (Fig ST4 and ST5). JTK shows lower MCC values in datasets B and D as compared to the MCC value of JTK for datasets A and C, indicating possible problems with non-stationary signals. Although CBNLR cannot meet the performance of JTK, the method shows a relatively high performance even in the context of a low signal-to-noise ratio. MCC values resolved for each dataset and signal-to-noise ratio are given in figures ST2-5.

If we combine signals identified as cyclical by all three methods (intersection of the three sets) the precision value is > 99% for all four datasets with a recall of 80-90%. Choosing signals identified by at least two out of three methods yields a good compromise between precision (94-97%) and recall (95-99%).

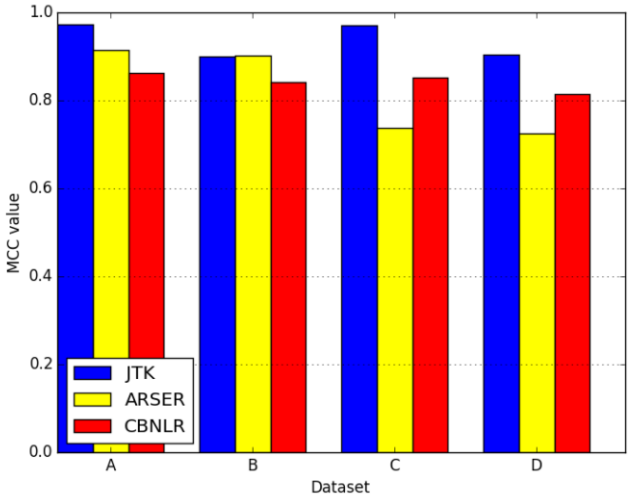



***Figure ST 1: Overview of performance on all dataset***

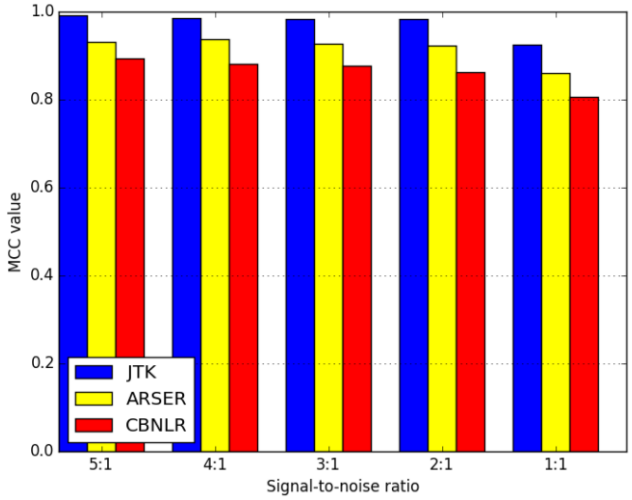



***Figure ST 2: Dataset A: stationary signal - white noise background***

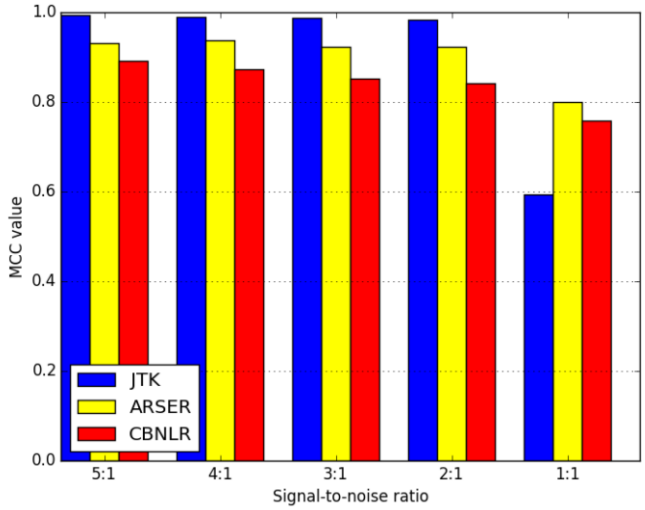



***Figure ST 3: Dataset B: non-stationary signal - white noise background***

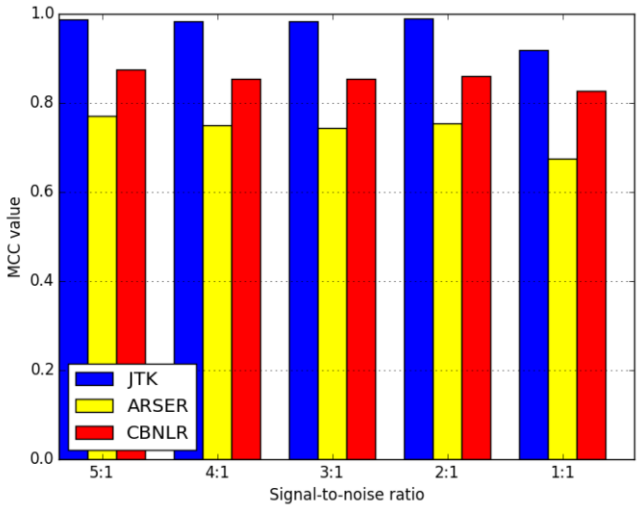



***Figure ST 4: Dataset C: stationary signal - AR(1)-background***

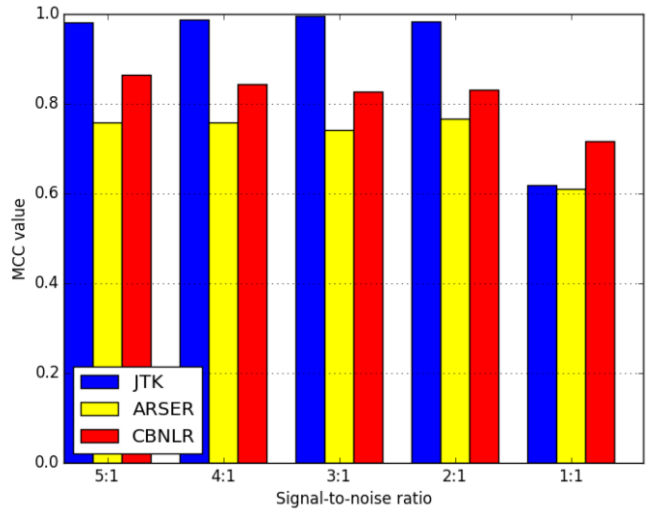



***Figure ST 5: Dataset D: non-stationary signal -AR(1)-background***

## SUPPLEMENTARY TABLES

For supplemental tables please refer to the Excel-files published together with this manuscript.

**S1 Table. Read Statistics of RNA sequencing**

This table contains the numbers of total reads, aligned reads as well as the percentage of aligned reads of the RNA sequencing samples.

**S2 Table. Circadian Expressed Genes in NIH3T3.**

The table contains normalized read counts and fold change of the genes detected to be cyclically expressed in NIH3T3 by at least three of four methods (JTK_CYCLE, RAIN, Arser, CBNLR). Particular parameters of each method, GO terms (see Fig. 1b) and cyclical expression in other published circadian time course data (obtained from CircaDB) are indicated for each gene.

**S3 Table. GO Term Results for Cyclically Expressed Genes in NIH3T3.**

GO term results obtained in this study using DAVID functional annotation tool as well as output from GSEA analysis.

**S4 Table. Motif Analysis Results**

The table shows the results of the motif analysis using HOMER software for early, mid, and late peaking genes.

**S5 Table. Circadian Expressed lincRNAs in NIH3T3**

The table contains normalized read counts and fold change of lincRNAs detected to be cyclically expressed in NIH3T3 by at least three of four methods (JTK_CYCLE, RAIN, Arser, CBNLR) for two different expression cutoffs (at any time point expressed above 32 or 10 normalized read counts). Particular parameters of each method as well as their genomic location are indicated for each lincRNA.

**S6 Table. Results of the Tested one-to-one Regulatory Models.**

Results of the model evaluation where it was tested if a particular input gene alone can in principle explain the expression of a potential target gene (one-to-one interaction). The following cases were evaluated: Bmal1, Nr1d1 and Usf2 regulating circadianly

expressed genes harboring the respective motif in their promoter (+/- 1 kb), Leo1 or Zfp28 regulating genes detected to be cyclical by JTK_CYCLE, RAIN, Arser, and CBNLR, as well as regulation of Leo1 or Zfp28 by specific core clock components. Included in the tables are $X^2$-values of the original fit (chisqr), p-values and estimated Hill coefficients (h).

# 2.3

# Mapping gene regulatory circuitry of Pax6 during neurogenesis

Sudhir Thakurela[1,5], Neha Tiwari[2,5], Sandra Schick[1], Angela Garding[1], Robert Ivanek[3], Benedikt Berninger[2,4], Vijay K. Tiwari[1]

1. Institute of Molecular Biology (IMB), Ackermannweg 4, Mainz, Germany
2. Institute of Physiological Chemistry, University Medical Center of the Johannes Gutenberg University Mainz, Hanns-Dieter-Hüsch-Weg 19, Mainz, Germany
3. Department of Biomedicine, University of Basel, Basel, Switzerland
4. Focus Program Translational Neuroscience, Johannes Gutenberg University Mainz, Langenbeckstr. 1, Mainz, Germany

[5] These authors contributed equally to this work.

Correspondence: Vijay K Tiwari; E-mail: v.tiwari@imb-mainz.de

Running title: Uncovering Pax6-dependent gene expression program

Key words: Chromatin; Gene regulation; Neural progenitors; Neurogenesis; Transcription Factors

**Published in the journal Cell Discovery:**

Cell Discovery, 2, 2016, article number: 15045, doi:10.1038/celldisc.2015.45

## ABSTRACT

Pax6 is a highly conserved transcription factor among vertebrates and is important in various aspects of the central nervous system development. However, the gene regulatory circuitry of Pax6 underlying these functions remains elusive. We find that Pax6 targets a large number of promoters in neural progenitor cells. Intriguingly, many of these sites are also bound by another progenitor factor, Sox2, which cooperates with Pax6 in gene regulation. A combinatorial analysis of Pax6-binding data set with transcriptome changes in Pax6-deficient neural progenitors reveals a dual role for Pax6, in which it activates the neuronal (ectodermal) genes while concurrently represses the mesodermal and endodermal genes, thereby ensuring the unidirectionality of lineage commitment towards neuronal differentiation. Furthermore, Pax6 is critical for inducing activity of transcription factors that elicit neurogenesis and repress others that promote non-neuronal lineages. In addition to many established downstream effectors, Pax6 directly binds and activates a number of genes that are specifically expressed in neural progenitors but have not been previously implicated in neurogenesis. The *in utero* knockdown of one such gene, *Ift74*, during brain development impairs polarity and migration of newborn neurons. These findings demonstrate new aspects of the gene regulatory circuitry of Pax6, revealing how it functions to control neuronal development at multiple levels to ensure unidirectionality and proper execution of the neurogenic program.

# INTRODUCTION

The paired box protein, Pax6, is a highly conserved transcription factor of 422 amino acids comprising two DNA-binding domains, an amino-terminal paired domain and a homeodomain along with a carboxy-terminal Proline/Serine/Threonine-rich transactivation domain [1, 2]. Pax6 was first discovered to be required for proper segmentation in Drosophila [3, 4] and later shown to be essential for eye development in Drosophila [5], a role that was further found to be conserved in human and mouse eye development [6, 7]. During mammalian brain development, Pax6 is expressed in a specific spatiotemporal manner and is restricted to mainly neuronal tissues [2, 8]. Pax6 is now established to be essential for maintaining the pool of neural stem cells (NSCs) and thereby regulating embryonic as well as adult neurogenesis, as shown by its expression in neuroepithelial and radial glial cells, which can divide symmetrically to produce NSCs or asymmetrically to become a NSC and a neuron [9, 10].

The discovery of a plethora of known Pax6 functions has been facilitated by various Pax6 mutants. One such very useful mutant, the small eye (Sey) mouse mutant, contains a single-base substitution [11], resulting in the production of a functionally inactive and truncated Pax6 lacking the DNA-binding homeodomain and the C-terminal activation domain. Importantly, the Sey mutant mouse phenotype is similar to that of an artificially targeted Pax6-deficient mouse ($Pax6^{-ax}$), showing small eyes and numerous neural defects, including reduced neurons in the cerebral cortex [11-13]. These phenotypic similarities in the Sey mutant and $Pax6^{-ax}$ substantiate the use of Sey homozygous mutant mice as Pax6-null mutants. It was further shown that Sey mutant embryonic stem (ES) cells generate misspecified neurons that undergo death because of high expression of the neurotrophin receptor p75NTR [14].

It is well established that Pax6 is crucial for the development of the central nervous system, eyes, nose, pancreas and pituitary gland [13, 15, 16]. Recent studies have shown that Pax6 functions upstream of gene networks involved in brain patterning, neuronal migration and neural circuit formation [17]. Despite the established role of Pax6 in neurogenesis, its genomic targets, their chromatin status and its cooperativity with other transcription factors during neurogenesis remain unclear. Furthermore, while a number of players functioning downstream of Pax6 have been identified, these are not enough to explain the plethora of functions Pax6 is known for. Here we reveal that in

neural progenitors, Pax6 binds a large number of gene promoters that exhibit epigenetic state that is hallmark of open chromatin. Many Pax6-bound promoters are also targeted by Sox2 and functionally cooperate in gene regulation underlying neuronal specification. Pax6 directly binds and silences genes important for mesoderm and endoderm development as they get de-repressed in progenitors lacking Pax6. In addition, Pax6 targets that are downregulated in mutant progenitors are known to be critically involved in neuronal development. Pax6-driven gene-expression program further induces activity of neurogenic transcription factors and repress others that promote non-neuronal lineages. Importantly, our analysis also revealed a number of Pax6-induced genes that are highly expressed during brain development but their function has not yet been explored during neurogenesis. Here we show that one such gene, *Ift74*, which is directly bound and activated by Pax6 in NP cells, is required for the proper migration of newborn neurons. Furthermore, our analysis revealed that Pax6 directly targets the promoter of Notch signaling components and induces their expression, which then further contribute to Ift74 expression. These observations reveal the manner by which Pax6 controls multiple components of the network underlying neuronal development and uncovers Ift74 as a novel regulator of neurogenesis.

# RESULTS

### Pax6 binds to a number of gene promoters in neural progenitor cells

We first determined the expression patterns of Pax6 in various embryonic tissues and cortical layers. As shown previously, Pax6 is specifically highly expressed in ventricular zone (VZ) and is gradually lost as cells progress through the subventricular zone (SVZ) to the cortical plate (CP; Supplementary Figure S1A). An analysis of other ectoderm (epidermis), mesoderm (heart and mouse embryonic fibroblasts) and endoderm tissues (lung and pancreas) showed relative absence of Pax6 expression, with the exception of the pancreas that exhibited low levels of Pax6, confirming previous reports [18] (Supplementary Figure S1A). We next use a highly refined and established differentiation model of neurogenesis, in which mouse ES cells first differentiate into Pax6-positive NP (radial glial-like) cells (also referred as cellular aggregates, in short CA_D8) and subsequently into terminally differentiated glutamatergic pyramidal neurons (TN) with high purity (>95%) and synchrony and is known to closely recapitulate the stages of embryonic neurogenesis [19-21]. The expression analysis of Pax6 in this system revealed its highest expression in cellular aggregate cells, thereby presenting a system for investigating Pax6 function *in vitro* (Supplementary Figure S1B).

To shed light on Pax6 function, we performed Pax6 chromatin immunoprecipitation (ChIP) in NPs and investigated its genome-wide binding pattern using a previously described ChIP-chip platform in biological replicates [22]. These arrays cover 10% of the mouse genome, including all well-annotated promoters, several large multigene loci and the complete chromosome 19 [23]. The visual inspection of the genomic regions suggested that Pax6 is targeted to distinct genomic sites and also occupies a number of promoters ($n$=5086, promoter enrichment >0.25; Figure 1a and b, Supplementary Figure S1C and D, Supplementary Table S1). A comprehensive and unbiased analysis of Pax6 binding along the fully tiled chromosome 19 revealed its relatively high enrichment at promoters (Figure 1c). These observations were validated at selected gene promoters in independent ChIP assays (Figure 1d). Such targeting of Pax6 to gene promoters prompted us to investigate its relationship with the chromatin state of target sites and the transcriptional states of associated genes at the progenitor stage. We analyzed the ChIP-seq datasets for RNA Pol II, H3K4me2, which is an established active histone modification, and the Polycomb group repressive mark, H3K27me3, at the NP

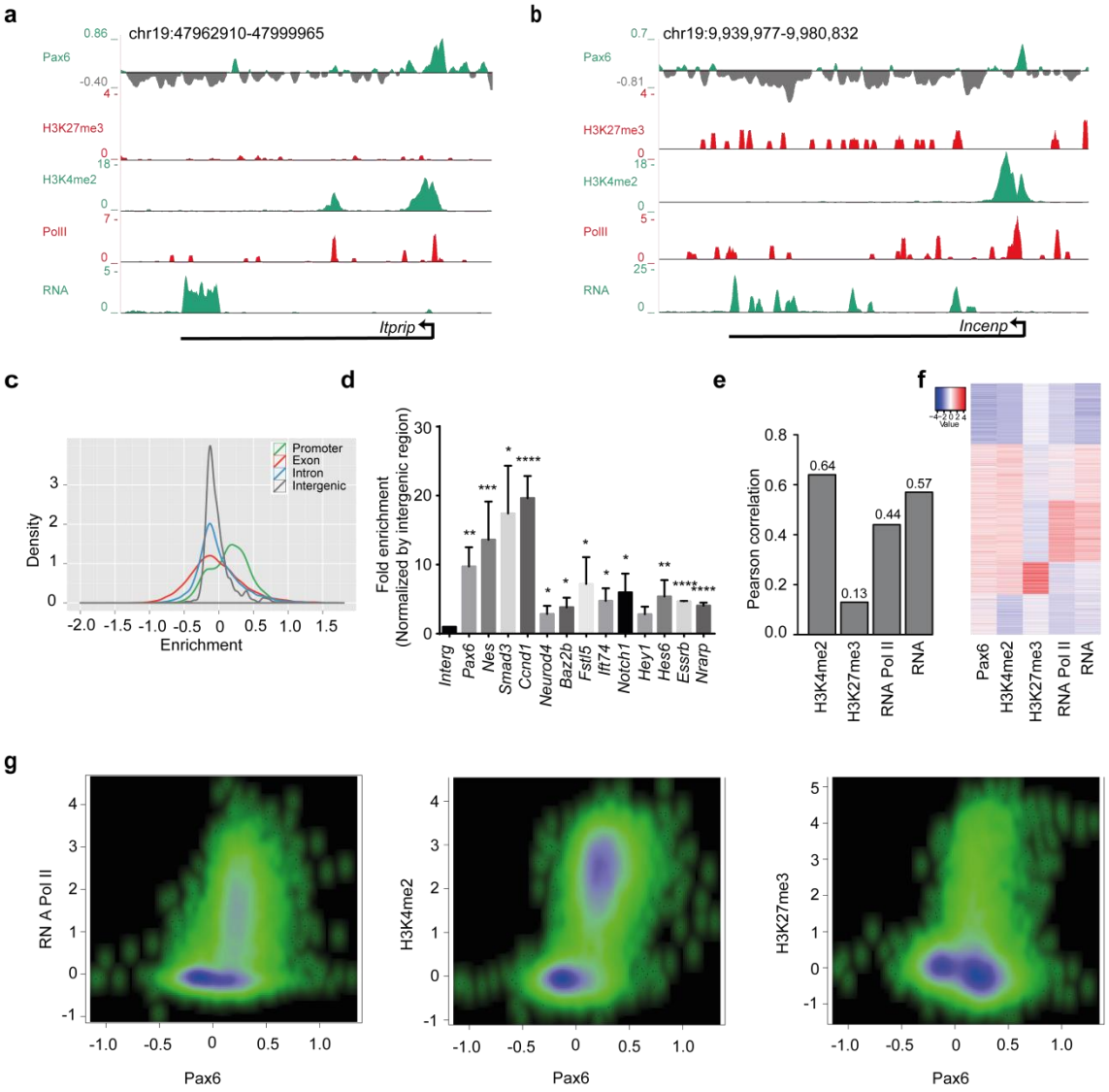



***Figure 1: Pax6 targets many gene promoters that largely exhibit an active chromatin state.***

(**a, b**) Representative examples of genome bowser tracks showing specific occupancies of Pax6 at promoter and its relative de-enrichment at other regions of a selected gene (*Itprip*-A, *Incenp*-B). UCSC browser tracks from ChIP-seq datasets for H3K4me2, H3K27me3 and RNA Pol II as well as RNA-seq from neural progenitor cells are shown. (**c**) Density plot showing genome-wide enrichment of Pax6 at promoters compared with other genomic regions on chromosome 19. To calculate the relative enrichment for each region, Pax6 enrichment was normalized with respect to the total size of that region. (**d**) ChIP-quantitative PCR validations of selected target genes showing the enrichment of Pax6 at their promoters. Pax6 and Nes are used as a positive control while intergenic region is a non-target (negative) control. Error bars reflect s.d. Statistical significance were calculated with an unpaired t-test (*$P$<0.05; **$P$<0.01; ***$P$<0.001; ****$P$<0.0001) (**e**) Bar plot showing Pearson correlation coefficients between Pax6 occupancy and H3K4me2, H3K27me3, and RNA Poll II RNA levels. (**f**) Heat map showing a comparison of Pax6 occupancy with H3K4me2, H3K27me3 and RNA Poll II RNA levels. Red indicates high values and blue low values. (**g**) Scatter plot comparing enrichment of Pax6 occupancy at promoters with enrichments of RNA Pol II, H3K4me2 and H3K27me3 in the same *in vitro* differentiation system. Each dot represents a promoter and *x* axis shows the enrichment of Pax6, while *y* axis represents enrichment of RNA Pol II, H3K4me2 or H3K27me3. Higher density of data points is represented as dark blue, while relatively less density is shown as light green.

stage in the same differentiation system, and correlated these data with Pax6 occupancy at the target gene promoters (Figure 1e and f). Further analysis showed that Pax6 occupancy was most strongly correlated with the active mark H3K4me2 ($R^2$= 0.64; Figure 1e). In addition, a large number of Pax6-bound promoters were RNA Pol II bound ($R^2$= 0.44) and actively transcribed ($R^2$= 0.57; Figure 1e). Furthermore, Pax6

target promoters were mostly devoid of the repressive mark H3K27me3 ($R^2$= 0.13; Figure 1e). Heat map visualization at promoters supported these observations, revealing that the majority of Pax6-bound genes displayed the H3K4me2 mark, a significant fraction of which were Pol II bound and actively transcribed (Figure 1f). Furthermore, comparison of Pax6 occupancy at promoters with enrichment of RNA Pol II, H3K4me2 and H3K27me3 in the same differentiation system showed similar patterns (Figure 1g). A comparison with promoter targets of Pax6 recently identified by ChIP-seq assay in E12.5 forebrain tissue [24] showed that out of 240 promoter targets discovered in this study, 141 promoters were also detected as Pax6 targets in our study (data not shown), supporting the comprehensiveness of our data.

**Pax6 targets are misregulated in Pax6 mutant NPs**

To further investigate the genes under the direct transcriptional control of Pax6, we differentiated Pax6 mutant ES cells (isolated from the blastocysts of homozygote *Sey* mutants, referred to thereafter as 'mutant cells') [14] into NPs and performed genome-wide transcriptome profiling. Sey mutant ES cells generate misspecified neurons that undergo death owing to high expression of the neurotrophin receptor p75NTR [14]. Comparing the transcriptome of Sey mutant cells with that of wild-type (WT) NP cells revealed 675 differentially downregulated and 623 differentially upregulated genes exhibiting enrichment for the nervous system development and metabolism related Gene Ontology (GO), respectively (Figure 2a, Supplementary Figure S2A and B and Supplementary Table S2). Promoters of most of the genes downregulated in mutant progenitors were bound by Pax6 and very highly expressed in the WT progenitors, suggesting their robust transcription in the presence of Pax6 (Figure 2b and c). By directly comparing Pax6 binding at promoters to the transcriptional changes in mutant progenitors, we found that nearly all differentially expressed gene promoters (90%) were Pax6 bound (promoter enrichment >0) in the WT cells (Figure 2d, left bar plots; Supplementary Figure S2C) (Hypergeometric *P*-value, upregulated genes: 1.71e-265; downregulated genes: 0). To retain only those target promoters that were highly bound by Pax6, we increased the cut-off to a higher level (promoter enrichment >0.25) for Pax6 enrichment, which revealed 406 downregulated and 249 upregulated genes (Figure 2d and e; hypergeometric *P*-value, upregulated genes: 1.70e-22; downregulated genes: 1.05e-99). The observation that a large fraction of genes downregulated in the

mutant progenitors were highly enriched for Pax6 in the WT cells (*n*=406) is consistent with our previous observation that the majority of Pax6 targets were highly expressed in the WT NPs (Figure 2b and c). Interestingly, during the differentiation of ES cells into neurons, a large number of Pax6-bound/mutant-downregulated and Pax6-bound/mutant-upregulated genes were either majorly expressed and repressed in the NPs or an early neurogenesis stage, respectively (Figure 2f and g). In summary, Pax6 directly binds at the regulatory elements of many genes to govern their proper transcriptional dynamics during neuronal development.

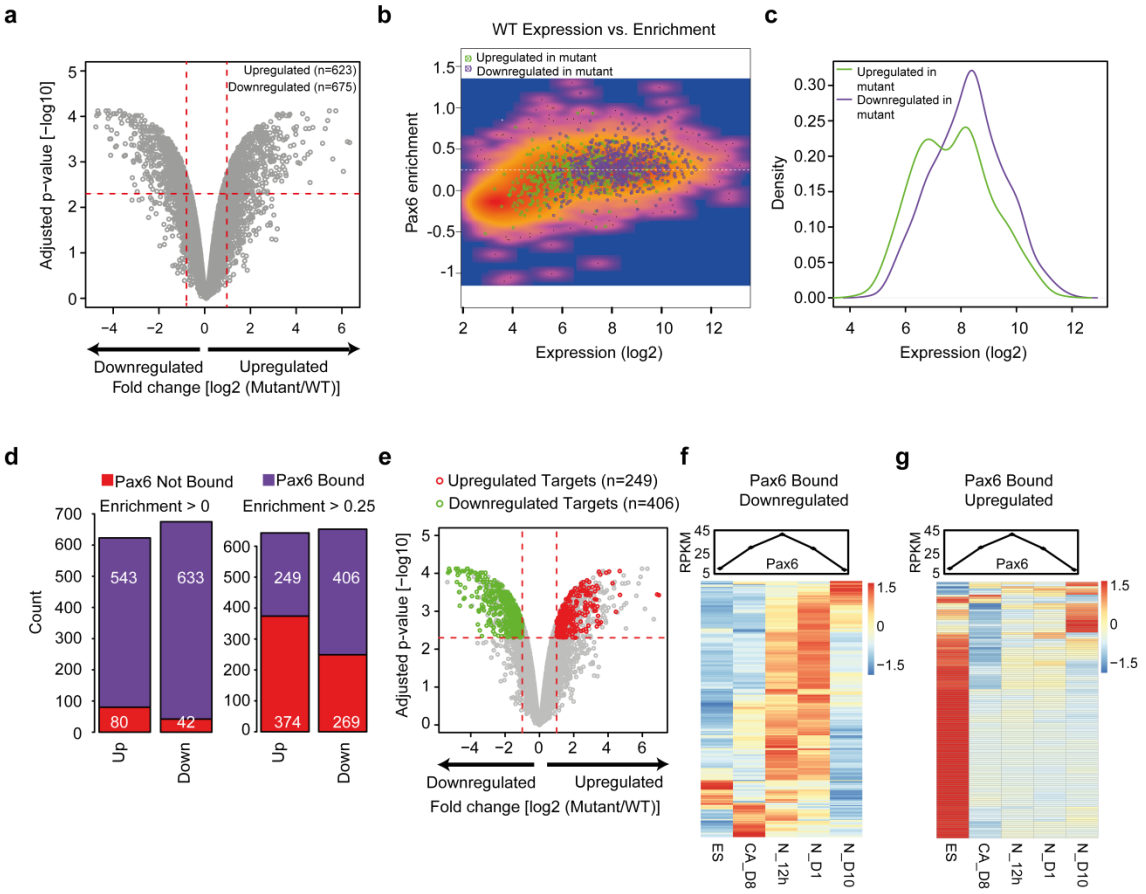



***Figure 2: Pax6 targets are misregulated in the absence of Pax6.***

(**a**) Volcano plot showing changes in expression of genes in WT and *Pax6* mutant (mutant). *x* axis represents fold change in log2 scale between WT and mutant, and the *y* axis shows level of significance as in -log10 of adjusted *P*-value. Vertical dotted red lines reflect a two-fold cutoff for expression while horizontal dotted red lines represent a *P*-value cutoff of 0.005. (**b**) Scatter plot showing log2 of expression levels in WT (*x* axis) and Pax6 enrichment on *y* axis. Upregulated (green) and downregulated (purple) genes in *Pax6* mutant are highlighted as dots. (**c**) Density plot to show the WT expression of upregulated and downregulated genes shown in **b**. *y* axis represents the density while *x* axis represents the expression levels. (**d**) Stacked bar plots showing that the substantial subset of differentially regulated genes in Pax6 mutants are Pax6 targets (two different Pax6 promoter enrichment cutoff: >0 (low stringency, left panel), >0.25 (high stringency, right panel). (**e**) Volcano plot as described in **a** but highlighting only upregulated (red dots) and downregulated (green dots) genes that are Pax6 targets in WT cells. (**f**, **g**) Heat maps showing the expression in ES, CA, N_12h (neurons at 12 hours), N_D1 (day 1 neurons) and N_D10 (neurons at day 10) of Pax6-bound genes that are significantly downregulated (**f**) and upregulated (**g**) in Pax6 mutant progenitors. Line plots above the heat maps show expression of Pax6 in the same stages.

**Pax6 activates neuronal development genes and represses genes from other lineages**

We next performed a GO enrichment analysis for genes that are Pax6-bound (promoter enrichment >0.25) and differentially expressed between the WT and mutant NPs. Genes bound by Pax6 and downregulated in mutant cells were exclusively enriched for neuronal development (Figure 3a). Interestingly, genes bound by Pax6 and upregulated in mutant progenitors showed enrichment for terms related to mesoderm (cardiovascular system development) and endoderm (for example, respiratory system development) development (Figure 3b). Considering these GO term enrichments, we further analyzed expression of these Pax6-bound differentially expressed genes in representative embryonic tissues from the three lineages. A majority of downregulated genes were expressed in the three layers of the embryonic cortex (VZ, SVZ and CP) [25], while upregulated genes were much higher expressed in tissues from mesoderm (heart and mouse embryonic fibroblast) [26, 27] and endoderm (lung and pancreas) lineages [28, 29] (Figure 3c and d). Furthermore, Pax6-bound upregulated and downregulated transcription factor genes also showed similar patterns in different lineages, as well as during *in vitro* neurogenesis where downregulated factors are mainly expressed in NPs or an early neurogenesis stage while upregulated factors show high expression in other lineages and ES cells (Figure 3e and f and Supplementary Figure S3A). To further substantiate these observations, we performed enrichment analysis on bound and differentially expressed genes based on known phenotypes associated with these genes. The downregulated genes were significantly associated with phenotypes related to brain development (Figure 3g), while upregulated genes were linked to phenotypes related to abnormal development of various mesodermal or endodermal tissues (Figure 3h). This further substantiates our previous observations and also provides additional insights into how Pax6 contributes to the gene expression program underlying neurogenesis.

To further uncover other aspects of the Pax6-dependent regulatory network, we performed a signaling pathway enrichment analysis. Genes regulated by Notch signaling, a pathway that is established to be critical for self-renewal of NSCs, were most highly enriched among the Pax6-target mutant-downregulated genes [30-32] (Figure 3i and j). This was followed by the Hedgehog signaling pathway, which is also

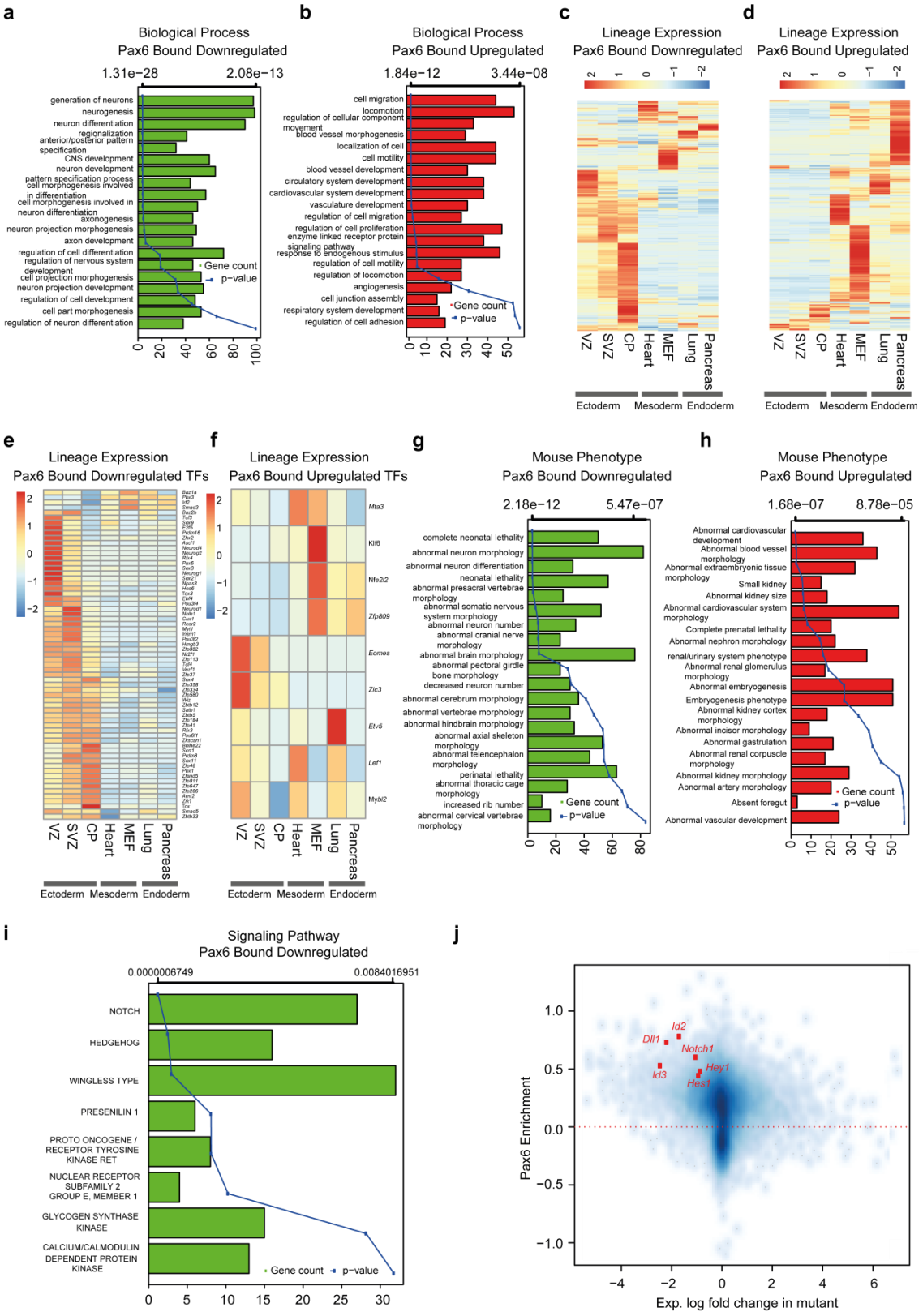



***Figure 3: Pax6 activates neuronal genes while repressing mesodermal and endodermal genes.***

(**a**, **b**) Bar and line plots showing GO term enrichment analysis of Pax6-bound genes that were downregulated (**a**) or upregulated (**b**) in *Pax6* mutant progenitors. Bar plots show number of genes for each enriched GO term (main *x* axis), and lines represent *P*-values for corresponding GO terms (alternate *x* axis). (**c**, **d**) Expression of upregulated (**c**) and downregulated (**d**) genes in tissues from different germ layers. (**e**, **f**) Same as in **c** and **d** but only for differentially expressed transcription factors. (**g**, **h**) Same as in **a** and **b** but an enrichment analysis was performed for mouse phenotypes enriched in Pax6-bound downregulated (**g**) and upregulated (**h**) genes. (**i**) Similar bar plot as in **a**, but the enrichment analysis was performed for signaling pathways using Genomatix. (**j**) Scatter plot showing changes in expression of core Notch signaling pathway components in WT and *Pax6* mutant cells and their enrichments for Pax6. The *x* axis represents the fold change (log2) between WT and *Pax6* mutant cells, and the *y* axis shows Pax6 enrichment.

shown to be important for specification of NPs [33] (Figure 3i and Supplementary Figure S3B). By contrast, Pax6-target mutant-upregulated genes showed enrichment for FGF signaling, that has been shown to be involved in mesodermal and endodermal specification [34-36] (Supplementary Figure S3C and D). Although previously Pax6 has been indirectly implicated in the control of Notch pathway [10], our analysis revealed that Pax6 directly binds at the promoters of a large number of genes associated with Notch signaling. Furthermore, these genes were downregulated in mutant NPs, indicating that Pax6 has a direct role in the activation of Notch signaling in NP cells (Figure 3j). This targeting by Pax6 at Notch signaling components provides potential mechanism regarding how this master transcription factor acts at multiple levels to define progenitor identity and differentiation towards neurons. Overall, these analyses identify a dual role for Pax6, in which it mediates the activation of neuronal (ectodermal) genes while concurrently represses the mesodermal and endodermal genes.

**Pax6 influences transcription factor network to confer unidirectionality towards neuronal differentiation**

On the basis of our observations that Pax6 mutant cells showed upregulation of non-neuronal and downregulation of neuronal genes, we next probed whether activities of any particular transcriptional factors are altered in the absence of Pax6 that in turns could explain part of gene-expression program alterations. Towards this we applied integrated system for motif activity response analysis (ISMARA), which predicts the transcription factors that can potentially regulate the differentially expressed genes on the basis of binding motifs at the promoters of these genes [37, 38]. ISMARA analysis predicted a number of transcription factors whose activity significantly changed in Pax6 mutant NPs. This included Pax6 and Sox2 that showed downregulation in their activity (Figure 4a-f). In line with these findings, ISMARA predicted targets of Pax6 and Sox2 were also found to be significantly downregulated in Sey cells (Figure 4b and e).

Furthermore, ISMARA also predicted interaction networks of Pax6 and Sox2 with other transcription factors many of which are known to be important for neurogenesis (Figure 4c and f). Two transcription factors, TFAP2B and TCF4, were commonly identified in both Pax6 and Sox2 interaction networks (Figure 4c, f and g). Surprisingly, predicted targets of TFAP2B or TCF4 were highly enriched for genes related to neurogenesis

(Supplementary Table S3). The role of TFAP2B in neuronal development as well as its interaction with Pax6 and Sox2 is unknown, however, our prediction provides potential insights of how cooperativity between different transcription factors contributes to neurogenesis. Targets of TFAP2B were also very significantly downregulated in Sey cells (Figure 4h) and network analysis further predicted its interaction with Pax6 as well as Sox2 in addition to many other interesting factors known to be required for neurogenesis (for example, Zeb1; Figure 4i). Overall, these observations suggest a cooperative function of transcription factors Pax6, Sox2 and TFAP2B in WT progenitors in gene activation as their activity and consequently their targets are downregulated in mutant cells.

Furthermore, ISMARA analysis also revealed upregulation in the activity of a number of transcription factors that are known to be important for non-neuronal lineages such as T (brachyury), Hnf1a and members of the Myf family (Figure 4j-r). Brachyury is an established mesoderm transcription factor [37, 38], while Hnf1a is critical for liver differentiation [39] and Myf family of transcription factors are known to be crucial for heart development [37]. Target genes of these three transcription factors were significantly upregulated in mutant cells (Figure 4k, n and q). GO enrichment analysis showed that these target genes are involved in the development and function of non-neuronal tissues (Supplementary Table S3). Furthermore, the network for each of these factors mostly consisted of a non-overlapping set of transcription factors (Figure 4l, o and r). Overall, these findings show that Pax6-dependent gene regulatory circuitry induces activity of transcription factors that induce neurogenesis and repress others that promote non-neuronal lineage.

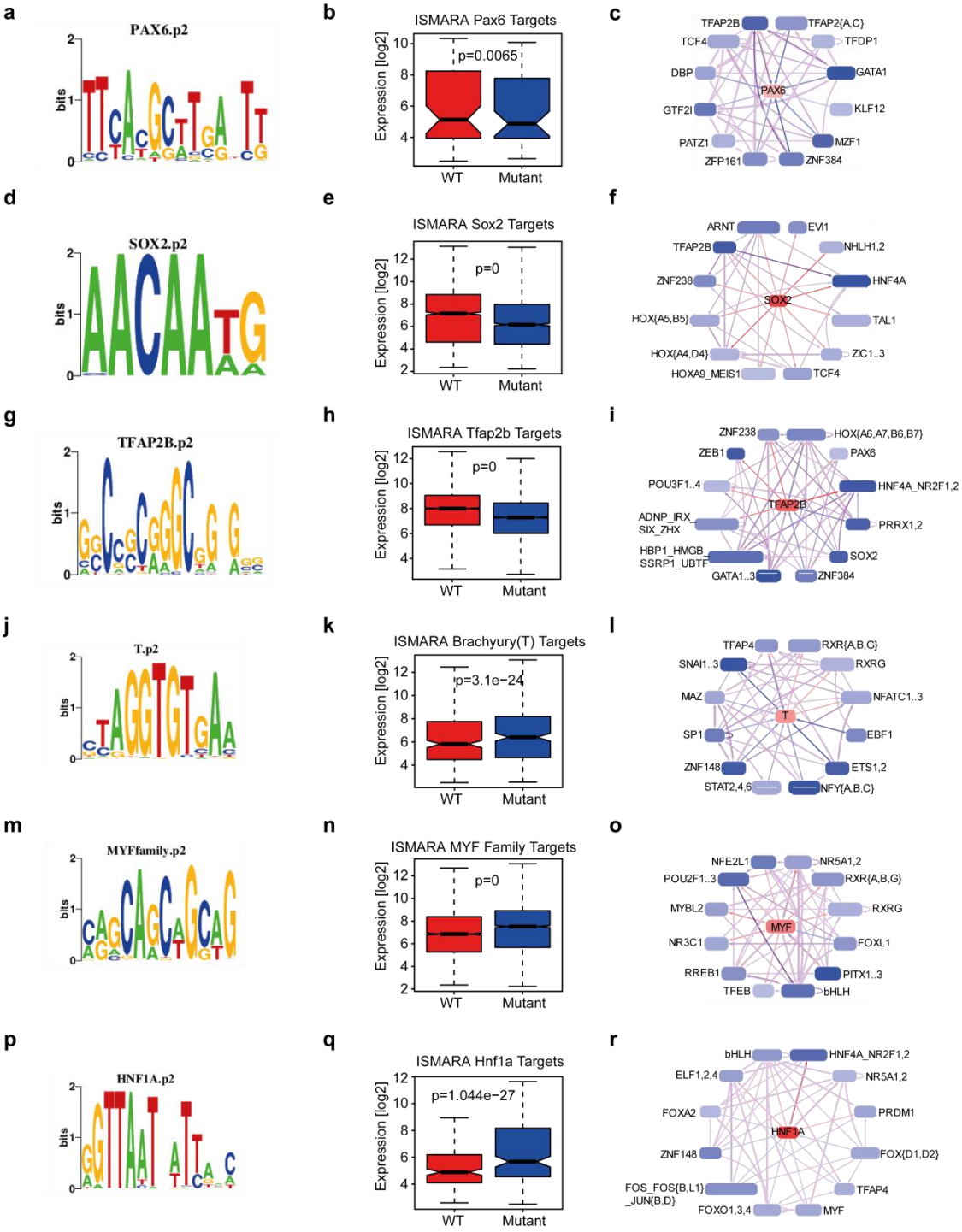



***Figure 4: Pax6 is critical for inducing activity of transcription factors that elicit neurogenesis and repress others that promote non-neuronal lineages.***

(**a**) Pax6 motif identified by ISMARA. (**b**) Box plot showing expression in WT and mutant of Pax6 targets predicted by ISMARA. (**c**) First-level interaction network of Pax6 and its potential targets as predicted by ISMARA. (**d**) Sox2 motif identified by ISMARA. (**e**) Box plot showing expression in WT and mutant of Sox2 targets predicted by ISMARA. (**f**) First-level interaction network of Sox2 and its potential targets as predicted by ISMARA. (**g**) Tfap2b motif identified by ISMARA. (**h**) Box plot showing expression in WT and mutant of Tfap2b targets predicted by ISMARA. (**i**) First-level interaction network of Tfap2B and its potential targets as predicted by ISMARA. (**j**) Brachyury (T) motif identified by ISMARA. (**k**) Box plot showing expression in WT and mutant of Brachyury (T) targets predicted by ISMARA. (**l**) First-level interaction network of Brachyury (T) and its potential targets as predicted by ISMARA. (**m**) Myf family motif identified by ISMARA. (**n**) Box plot showing expression in WT and mutant of Myf family targets predicted by ISMARA. (**o**) First-level interaction network of Myf family and its potential targets as predicted by ISMARA. (**p**) Hnf1a motif identified by ISMARA. (**q**) Box plot showing expression in WT and mutant of Hnf1a targets predicted by ISMARA. (**r**) First-level interaction network of Hnf1a and its potential targets as predicted by ISMARA. All *P*-values are calculated using Wilcoxon test.

**Sox2 targets a large number of Pax6-bound gene promoters**

We were intrigued by our observations that Sox2 activity is significantly reduced in Pax6 mutant NP cells. Both Sox2 and Pax6 are known to be important for the maintenance of the proliferative and developmental potential of NSCs [40]. Although it is known that Pax6 and Sox2 form a complex [41, 42], it remains to be investigated whether they function together in gene regulation at the same targets sites in the genome. We, therefore, compared our list of Pax6 target promoters with that of Sox2-bound promoters in NP cells derived from mouse ES cells in a previous study [43]. This analysis revealed that both Pax6 and Sox2 co-occupy a noticeable set of gene promoters, suggesting a potential cooperativity between these two transcription factors in gene regulation (Figure 5a and Supplementary Figure S4A; hypergeometric *P*-value: 1.28e-65). We next classified the genes encoding transcription factors, which were either expressed or repressed in the NPs *in vivo* (based on the transcriptome analysis of the E14.5 VZ cells) [25] and analyzed their promoter occupancies by Pax6 and Sox2. Pax6 and Sox2 were bound at the promoters of ~40% of the transcription factors expressed in the VZ (Figure 5b; hypergeometric *P*-value: 1.04e-22, Pax6 and 6.92e-27, Sox2). To our surprise, of the transcription factors that were not transcribed in the VZ, Pax6 occupied nearly 3.5-fold more targets compared with Sox2 (~37%, Hypergeometric *P*-value: 1.02e-20 Pax6, versus ~10%, Hypergeometric *P*-value: 1 Sox2; Figure 5c). In line with our previous observations, these results also suggest that Pax6-Sox2 complex preferentially bind to expressed transcription factors while without Sox2, Pax6 acts as a repressor. We were next curious to investigate whether the expression of Pax6 only bound target genes differs with respect to those bound by both Pax6 and Sox2 in WT and Pax6 mutant progenitors. Interestingly, genes bound by both Pax6 and Sox2 were significantly higher expressed in WT as compared with Pax6 alone or a random set of genes (Figure 5d). In line with these observations, these Pax6 and Sox2 common target promoters show higher accessibility as compared with Pax6 only and random promoters (Supplementary Figure S4B). Further supporting these findings, genes bound by both were more severely downregulated in mutant relative to Pax6 alone bound genes (Figure 5e). Together with previous observations, these results argue for an active cooperativity between Pax6 and Sox2 in regulating transcription of distinct set of genes in NP cells.

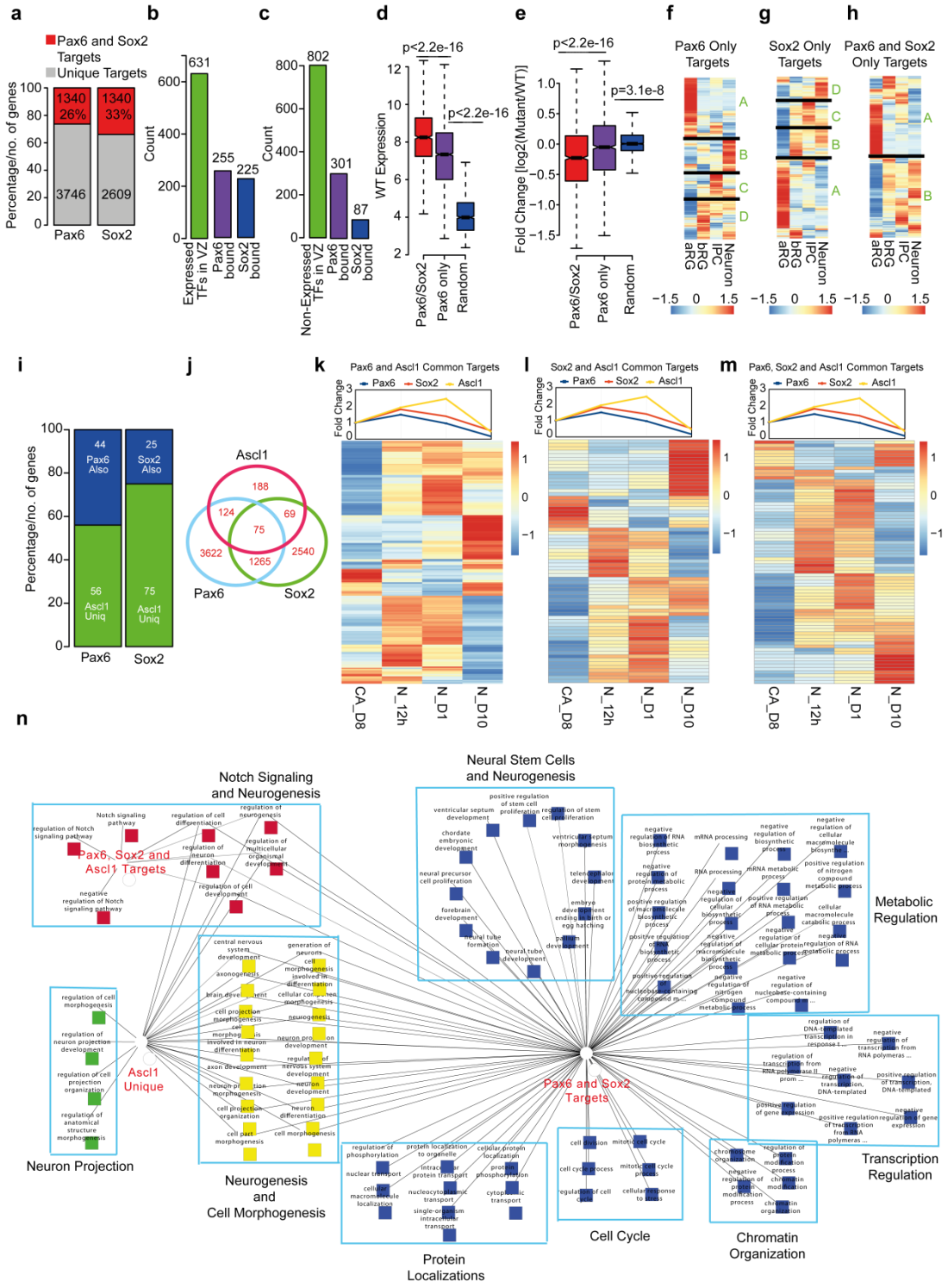



***Figure 5: Pax6 and Sox2 act cooperatively to drive neurogenesis.***

(**a**) Stacked bar plot showing percentage of overlapping Pax6 and Sox2 targets. *y* axis represent percentage of Pax6 and Sox2 targets. (**b**) Bar plot showing expressed transcription factors in VZ as well as the ones targeted by either Pax6 or Sox2. (**c**) Same as in **b** but for not-expressed transcription factors. (**d**, **e**) Box plot showing expression and changes in expression of Pax6 and Sox2 targets, Pax6 only targets and random genes in WT (**d**) and mutant progenitors (**e**). *y* axis in **d** represents expression in WT cells while *y* axis in **e** shows fold change of expression between WT and mutant cells. *P*-value is calculated using Wilcoxon test. (**f-h**) Expression of Pax6 only (**f**), Sox2 only (**g**) and Pax6 and Sox2 targets (**h**) during several stages of neurogenesis. aRG, apical radial glial; bRG, basal radial glial; IPC, intermediate progenitors. (**i**) Stacked bar plot showing the overlap of Ascl1 targets with Pax6 and Sox2. *y* axis represent percentage of Ascl1 targets. (**j**) Venn diagram showing the number of overlapping targets of Pax6, Sox2 or Ascl1. (**k-m**) Heat maps showing expression of Pax6 and Ascl1 common targets (**k**), Sox2 and Ascl1 common targets (**l**) and Pax6, Sox2 and Ascl1 common targets during different steps of *in vitro* neurogenesis. Line plots above heat maps show expression of Pax6, Sox2 and Ascl1 in the same stages. Fold change is with respect to CA day 8 (CA_D8). (**n**) Comparative GO enrichment analysis of different set of genes represented as a network. Each square represent a GO term associated with a particular list, while edges provide information about the list to which the particular GO term is associated. If a GO term was found to be present in all three lists, it got connected with all three list nodes by edges. Pax6 and Sox2 target genes functions were further divided into sub-clusters based on similar GO terms.

To further delineate and substantiate the expression dynamics of Pax6 and Sox2 targets, we explored recently published transcriptome datasets for distinct progenitor subpopulations (aRG, apical radial glial; bRG, basal radial glial; IPC, intermediate progenitors) as well as neurons from developing mouse neocortex [44]. Comparison of NP markers in our *in vitro* neuronal differentiation system and the above datasets showed that our ES-derived progenitors are apical in nature (Supplemental Figure S4C and D). Further comparison revealed an interesting pattern of expression for Pax6 only and Sox2 only bound genes compared with Pax6 and Sox2 co-occupied genes during neurogenesis (Figure 5f and h). The set of genes bound by either Pax6 or Sox2 and expressed in aRG were repressed in the immediate next stage (bRG) and remained repressed throughout neurogenesis (cluster A in Figure 5f and g). However, the genes bound by either Pax6 or Sox2 and repressed in aRG showed transcriptional activation in a stage-specific manner during neurogenesis (cluster B, C and D; Figure 5f and g). In contrast, genes co-occupied by both transcription factors and expressed (Cluster A) or repressed (Cluster B) in aRG were immediately repressed or activated in bRG, respectively, and maintained this state throughout neurogenesis (Figure 5h). Overall, these observations suggest that the gene regulatory function of Pax6 at its target sites may be influenced by co-factors such as Sox2.

We next wondered whether other transcription factors expressed later during neurogenesis could function at Pax6 and Sox2 target sites when Pax6 and Sox2 are no longer available. To test this hypothesis we chose Ascl1, which is shown to be essential for the transition from neuronal progenitors to a neuronal state [45, 46] and neurogenesis is severely impaired in the absence of Ascl1 [47-49]. During neuronal differentiation from ES cells, Pax6 and Sox2 are simultaneously highly expressed in NP cells and following onset of neurogenesis, their levels decrease while Ascl1 levels are further increased (Supplemental Figure S4E). Using a recently published genome-wide binding dataset for Ascl1 during neurogenesis [50], we found that Ascl1 shared 44% of Pax6 (hypergeometric *P*-value, 1.55e-22) and 25% (hypergeometric *P*-value, 3.24e-13) of Sox2 targets (Figure 5i and j). Interestingly, the targets common between Pax6, Sox2 and Ascl1 (*n*=75) included classical Notch pathway (*Id1, Id2, Hey1, Hes6* and *Dll1*) and neuronal (*Tubb2b, Robo1, Mapt* and *Pcdh10*) genes. We next explored how Pax6 and Sox2 targets that are also bound by Ascl1 are expressed during *in vitro* neurogenesis. Heat map visualization of these sets showed that such genes that are bound by Pax6

and/or Sox2 and also by Ascl1 mostly maintain their transcription state as cells exit NP state (higher Pax6/Sox2 and lower Ascl1 levels) towards initiating neurogenesis (lower Pax6/Sox2 and higher Ascl1 levels; Figure 5k and m). Furthermore, most of these genes acquire an opposite expression state in terminally differentiated neurons (no Pax6, Sox2 or Ascl1 expression). This suggests that distinct sets of Pax6/Sox2 target genes might be targeted by other transcription factors in subsequent stages of neurogenesis to facilitate maintenance of their transcription state despite the later absence of Pax6/Sox2 itself.

To further explore the functional differences between the genes occupied by Ascl1 uniquely or Ascl1 along with Pax6 and/or Sox2, we performed a comparative GO term analysis to reveal their possible involvement in any specific biological processes (Figure 5n). The set of genes that were targeted by Pax6 and Sox2 only (blue squares) were enriched for a broad range of functions related to neural precursor or neural tube formation, cell cycle, transcription regulation, protein localization, metabolic processes and chromatin organization (Figure 5n). These genes were also enriched for functions related to neuronal differentiation and maturation (yellow squares). Interestingly, Ascl1 unique target genes were also enriched for these functions (yellow squares) indicating towards a functional takeover of neuronal development by Ascl1 (Figure 5n). The set of genes that were bound by Pax6/Sox2 complex and also by Ascl1 were exclusively enriched for Notch signaling and neuronal differentiation (Figure 5n, red squares). We also observed that the functional class 'neuronal projection' was uniquely attributed to targets that were also targeted by Ascl1 only (green squares; Figure 5n), supporting its known role in early neuronal development. Overall these data indicate that a subset of neurogenesis-related genes that are acted upon by Pax6 and/or Sox2 in NPs may also be targeted by other transcription factors such as Ascl1 for gene regulation during neuronal development.

**Pax6 directly induces expression of many known and novel NP-specific transcription factors**

We next attempted to further investigate the role of Pax6 in regulating the expression of NP-specific genes by performing a series of stepwise analyses. First, we selected the Pax6-bound mutant-downregulated genes that were significantly higher expressed in E14.5 cortical layers compared with other tissues (heart, embryonic fibroblasts, lung and

pancreas). Then, we selected those factors that were at least two-fold upregulated in the VZ compared with the CP (Figure 6a). Interestingly, this final list of 46 genes primarily consisted of transcription factors, including established Pax6 targets and known regulators of NP identity (for example, Nestin, *Neurog1/2, Neurod1/4* and Notch pathway components, such as *Dll1* and *Hes6*; Figure 6b and Supplementary Figure S5A). Pax6 was also bound to its own locus likely for autoregulation as shown previously [12]. Of these 46 Pax6-target gene promoters, 17 were also co-occupied by Sox2 (data not shown). This analysis also identified many novel factors that have not been previously shown to function in regulating progenitor identity (Supplemental Figure S5B). The expression pattern of many of these genes was further validated by their *in situ* hybridization analysis in the embryonic cortex (Figure 6c) [51]. This analysis revealed how Pax6 functions as an upstream regulator of many known critical neurogenesis-related transcription factors, at the same time identified many previously unknown Pax6 targets that are specifically expressed in the cortex and warrant further investigation.

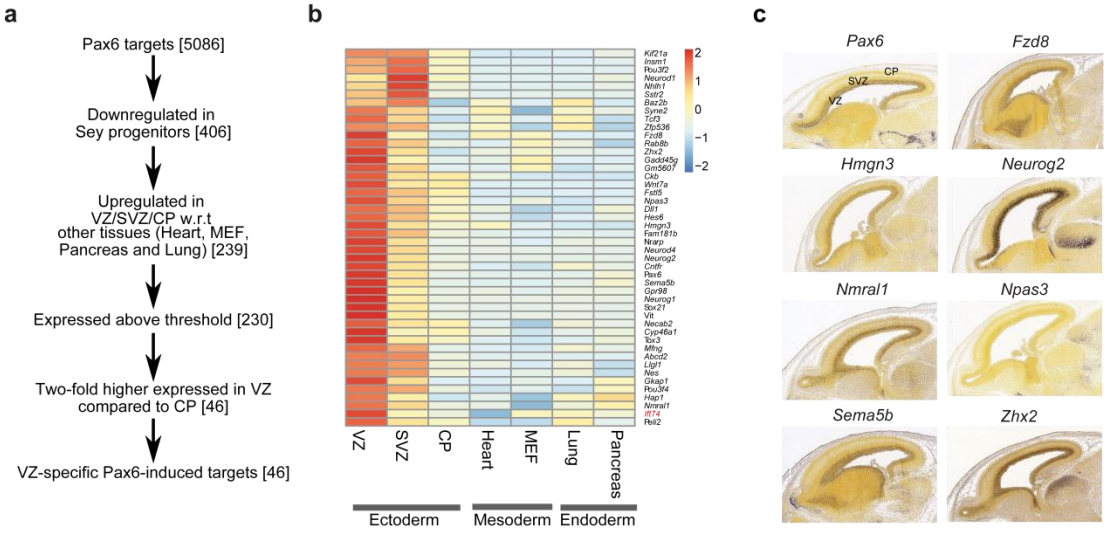



***Figure 6: Pax6 directly induces expression of a large number of known and novel neural progenitor-specific transcription factors***

(**a**) Flow chart showing the identification of genes that are specifically upregulated in neural progenitors *in vivo* and are regulated by Pax6. (**b**) Heat map showing the expression patterns of 46 genes identified in **a**. Red means high expression while blue means lower expression. (MEF, mouse embryonic fibroblast). (**c**) *In situ* hybridization images for known and novel Pax6 targets as derived from the Allen Brain Atlas (http://developingmouse.brain-map.org/).

### *Ift74* is a novel Pax6 target that contributes to neuronal migration

We were next interested to deeply explore the function of novel genes that were directly bound and activated by Pax6 and whose expression was restricted to NPs. We focused on *Ift74* (Intraflagellar transport (IFT) 74 homolog), which is a component of the IFT complex but remains a rather uncharacterized protein in the context of mammalian biology. *Ift74* forms a tubulin-binding module together with IFT81 that specifically mediates transport of tubulin within the cilium required for ciliogenesis [52]. To precisely map the kinetics of its expression with respect to *Pax6*, we analyzed their expression at various time points during the differentiation of ES cells into neurons via a NP state. As expected, this fine time course analysis during neuronal differentiation revealed that *Pax6* was most highly induced upon commitment to NPs and downregulated as soon as neurogenesis progressed (Figure 7a). Interestingly, analysis of *Ift74* at same time points showed that it reached its maximum expression levels few hours after highest *Pax6* expression, a stage that marks the transition of NP cells to neurons, and subsequently its expression was reduced upon neuronal maturation (Figure 7a).

To further substantiate our observations of *Ift74* induction in the context of *Pax6* expression *in vivo*, we next analyzed transcriptome data derived from the three layers of the E14.5 cortex (VZ, SVZ and CP) that showed the prominent expression of *Ift74* in the VZ of the developing mouse brain [25], which is where *Pax6* is also most highly expressed (Figure 7b and Supplementary Figure S1A). In order to confirm the direct binding of Pax6 at the promoter of *Ift74*, we performed ChIP assay in NP cells using Pax6-specific antibody. Real-time PCR analysis confirmed a high enrichment of Pax6 at a region upstream of the transcription start site of *Ift74* gene (Figure 7c). Given that Notch signaling is known to be essential for the self-renewal and identity of NP cells [30-32] and Notch effector transcription factor RBPJ showed occupancy at *Ift74* promoter in NSCs (D. Castro. Personal communication), we studied the effects of blocking Notch signaling on *Ift74* levels. To test whether Ift74 expression is regulated by Notch pathway, we inhibited Notch signaling using two independent inhibitors (LY-411575 and (N-[N-(3,5-Difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester (DAPT)) and analyzed the expression of Ift74. These analyses showed that under both inhibitor treatments, Ift74 was significantly downregulated (Figure 7d). In addition, such

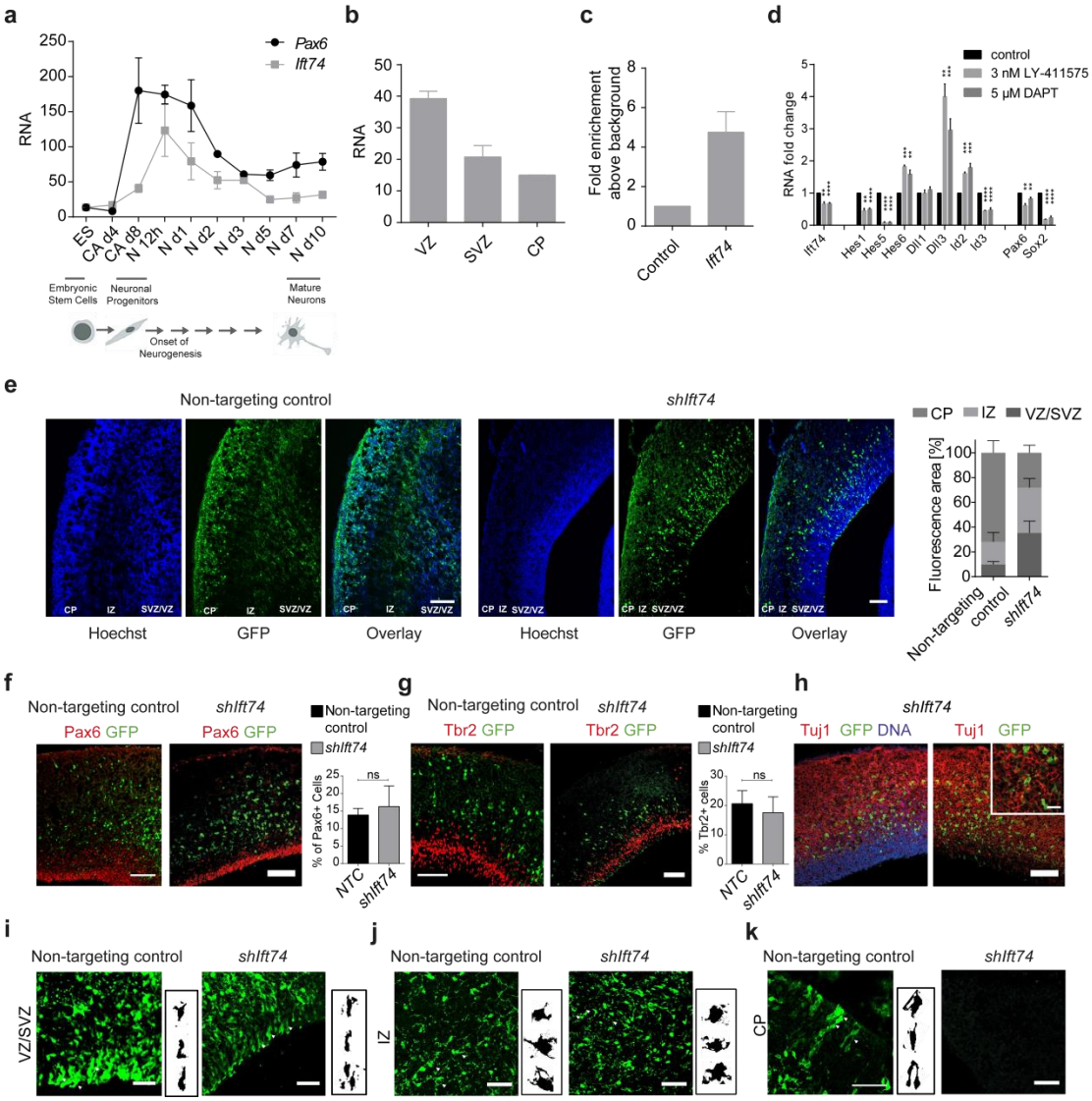



***Figure 7: Ift74 contributes to neuronal migration.***

(**a**) Expression of *Pax6* and *Ift74* during neuronal differentiation of ES cells derived by real time quantitative PCR (RT-qPCR). Expression is shown for various stages of neuronal differentiation (ES cells, CA at day 4 before adding RA (CA d4), CA at day 8 (CA d8)) and various time points during neurogenesis (TN at 12h and day 1, 2, 3, 5, 7 and 10). mRNA expression is normalized to the housekeeping gene *Rpl19* ($n$=3, error bars show s.e.m.). (**b**) Expression of *Ift74* (in RPKM) in VZ, SVZ, and CP dissected from mouse embryos at E14.5 derived by RNA-sequencing (GSE30765). (**c**) Pax6 ChIP-qPCRs to validate Pax6 binding at the promoter (-200±0 bp region) of *Ift74* at CA day 8 (CA_d8) ($n$=3, error bars shown as s.e.m.). Average enrichments are plotted normalized to input and further to an intergenic control region (control). (**d**) Fold change in mRNA levels in CA_d8 cells treated with γ-secretase inhibitor as compared with non-treated cells. ES cells were induced to undergo neuronal differentiation and treated every other day with 5 μM N-[N-(3,5-Difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester (DAPT) ($n$=5), 3 nM LY-411575 ($n$=3) or dimethylsulfoxide as control from CA_d4 stage onwards. Expression of the shown genes were normalized to Rpl19 levels (ΔCT) and fold change with respect to the control is plotted (error bars show s.e.m.). Statistical significance were calculated with an unpaired $t$-test (*$P$<0.05; **$P$<0.01; ***$P$<0.001; ****$P$<0.0001). (**e**) Left panel: representative immunofluorescence images from coronal brain sections at E16.5 stained for DNA using Hoechst (blue). Green fluorescent protein (GFP; green) marks *shIft74* or non-targeting control electroporated cells; brains were electroporated at E12.5 and analyzed after 4 days. Scale bar: 100μm. Right panel: quantifications of fluorescence signal in the GFP channel using ImageJ in the lower Hoechst dense region (VZ/SVZ), the intermediate less DNA dense region (IZ) and the upper Hoechst dense region (CP). Error bars reflect s.e.m. of three representative regions from two independently electroporated brains. (**f**) Representative immunofluorescence images of cortical brain slices electroporated at E12.5 with non-targeting control or *shIft74* and analyzed at E16.5. GFP (green) and PAX6 (red) stain is represented and the scale bar: 100 μm. All GFP-positive cells have been counted and the percentage of cells that also displayed PAX6 signal is plotted on the *y* axis ($n$=2, error bars show s.e.m.). (**g**) Same as in **f** but co-stained with TBR2 (red) and accordingly quantified as above for TBR2-positive electroporated cells. (**h**) Representative immunofluorescence images of cortical brain slices electroporated at E12.5 with *shIft74* and analyzed at E16.5. GFP (green), TUJ1 (red) and DNA (blue) stain is represented and the scale bar: 100 μm. (**i-k**) GFP images to analyze cell shape and polarity for non-targeting control and *shIft74* electroporated brains in VZ/SVZ (**i**), IZ (**j**) and CP (**k**) region. White arrowheads point to multipolar roundish cells in the IZ and bipolar cells in VZ, SVZ and CP of non-targeting control and *shIft74* electroporated brains. Scale bar: 50 μm. On the right side of each image representative thresholded cells of the particular layer are represented.

blockage of Notch pathway also led to expected changes in the expression of Notch signaling components (Figure 7d) and is in agreement to previous studies [53]. Importantly, further in line with a critical role of Notch signaling in regulating NSC self-renewal and identity, we also found that the loss of Notch signaling also led to significant reduction in the expression of NP markers (Pax6 and Sox2; Figure 7d). Given these observations, an alternative explanation for the downregulation of Ift74 by Notch inhibitors is that the NPs differentiate into more mature cell types that do not express Ift74. Since our earlier observations showed a direct induction of Notch signaling components by Pax6, the decrease in Ift74 expression upon Notch inhibition also suggests a potential functional cooperativity between Pax6 and Notch signaling in regulating the downstream gene-expression program.

Given the known function of *Ift74* in the transport of tubulin within the cilium that is required for ciliogenesis [52], we were tempted to investigate whether its induction by Pax6 serves to promote neuronal migration during later stages of neurogenesis. Towards this, we performed *in utero* electroporation using a plasmid encoding a tested shRNA against *Ift74* (Supplementary Figure S6A) at embryonic stage E12.5 and sacrificed the embryos for characterization at E16.5. Depletion of Ift74 via this shRNA *in vitro* does not result in cell death or impaired the cell-cycle progression (Supplementary Figure S6B and C). In the control shRNA electroporated brains, the majority of electroporated cells were detected in the cortical plate and very few such cells were retained in the ventricular zone, reflecting proper cortical migration of newborn neurons (Figure 7e). *Ift74* shRNA electroporated brains, in contrast, showed a distinct phenotype where the majority of Ift74-depleted cells failed to migrate to the cortical plate (Figure 7e). Having observed such mislocalization of *Ift74*-depleted cells we were interested to uncover at which stage of neurogenesis these cells are perturbed. A staining of electroporated brain sections with Pax6 (Figure 7f) and Tbr2 (Figure 7g) showed no defect in the early neuronal maturation processes since the percentage of *shIft74* and non-targeting control electroporated cells showing similar Pax6 and Tbr2 expression. We next assessed whether *Ift74*-depleted cells migrating above the Tbr2 layer, express neuronal markers. A co-staining with Tuj1 revealed a clear overlap with *shIft74* electroporated cells (Figure 7h) indicating that these cells achieve neuronal identity but fail to fully migrate towards upper cortical layers. Although having a closer look into these images, many cells appeared to have multiple small processes and

suggested that in the absence of *Ift74*, cells accumulate in the multipolar phase in SVZ/intermediate zone (IZ) region instead of migrating to CP.

For a detailed look into this phenomenon we took images at higher magnification of non-targeting control and *shIft74* electroporated cells from different cortical layers and further investigated the morphological appearance of such cells (Figure 7i-k). We observed that, in the ventricular/subventricular zone, control as well as *Ift74*-deficient cells displayed a comparable elongated polar appearance (Figure 7i). In the IZ, cells electroporated with either non-targeting control or *shIft74* again looked very similar and displayed a more roundish multipolar morphology (Figure 7j). However, while control cells subsequently transform back into a bipolar shape while migrating towards cortical plate, *Ift74*-depleted cells failed to do so and reside as multipolar cells in the IZ (Figure 7k). These observations suggest that cells deficient for *Ift74* are able to normally differentiate until the stage of immature projection neurons that enter the IZ and become multipolar but are not able to reorient into elongated, bipolar shape essential for proper migration to the cortical layer. Since neuronal migration is driven via epithelial to mesenchymal transition (EMT)-like mechanisms [54], we analyzed the role of Ift74 in cellular migration using mammary epithelial cells as an established cellular model of EMT [54-56]. We induced EMT and at the same time transfected shRNA against Ift74 in the epithelial cells and then measured cellular migration four days later. We find that shRNA-mediated depletion of Ift74 led to a significant reduction in the migration capacity of these cells (Supplementary Figure S6D and E). Taken together, these observations suggest that Pax6 and Notch signaling may cooperate in regulating gene expression (for example, of Ift74) as well as identify *Ift74* as a new Pax6 target that potentially has a role during neurogenesis likely via contributing to the migration of newborn neurons.

# DISCUSSION

Pax6 is a known master regulator of NP identity [8, 9, 15, 16, 57-63]. In this study, we attempted to uncover genes under the transcriptional control of Pax6 in NPs and identified downstream transcription factors that contribute to neurogenesis. We found that in NPs, Pax6 is targeted to many promoters that showed a distinct epigenetic state of open chromatin. Interestingly, many Pax6 sites are also occupied by Sox2, suggesting that they function together in gene regulation. Pax6 deficiency causes defects in the expression of its target genes, linking its binding to a function in transcriptional regulation. Strikingly, our analysis also revealed a dual role for Pax6, in which it activates the neuronal (ectodermal) genes while concurrently represses the mesodermal and endodermal genes. Importantly, Pax6 also directly induces the expression of a number of known as well as novel genes including transcription factors that are specifically expressed in NPs. We further show that one of the novel Pax6 target gene, *Ift74*, may have an important role during neurogenesis, likely via regulating migration of newborn neurons. Furthermore, our results also provide indication that Notch signaling contributes to the transcriptional induction of *Ift74* in NPs. Interestingly, Pax6 also directly binds at the promoters of many Notch signaling components and functions in their activation, suggesting the functional cooperativity between Pax6 and Notch signaling in regulating downstream gene-expression program. Overall, our findings reveal how Pax6 regulates the gene-expression program at multiple levels to ensure proper execution of the neurogenic program, and at the same time ensures the unidirectionality of neuronal differentiation (Figure 8).

The complexity of Pax6 function has been suggested to arise from its interaction with various transcription factors to synergistically regulate target gene expression. In lens development, the transcriptional regulation of several crystallin genes by Pax6 is achieved in coordination with other transcription factors, such as Sox2 and Maf [64, 65]. Pax6 has been shown to form a complex with Sox2 to transcriptionally activate the *δ-crystallin* gene [41]. Sox2 is of further relevance because it is expressed in the developing mouse central nervous system from an early stage [66] and regulates the expression of fibroblast growth factor 4 (*Fgf4*) and Nestin, which are important in maintaining NSCs [67]. Our ISMARA analysis showed that the activity of Sox2 was significantly reduced in Pax6 mutant NP cells. Moreover, Sox2 co-occurred with Pax6 at

many promoters in NP cells. Furthermore, these data also reveal that critical NP genes, such as Nestin, are co-regulated by Pax6 and Sox2. We also observed that the target genes co-occupied by Pax6 and Sox2 are expressed at higher levels, including those encoding important transcription factors, as compared with Pax6 only targets. These findings highlight the importance of the interplay between Pax6 and Sox2 in cooperative transcriptional regulation during neurogenesis, and at the same time also imply that critical neurogenesis genes may require co-activation by more than one stage-specific transcription factor. Furthermore, these observations also indicate that the gene regulatory potential of Pax6 may be determined by its partners and in this specific case, Sox2 occupancy drives it more towards a transcription activating role.

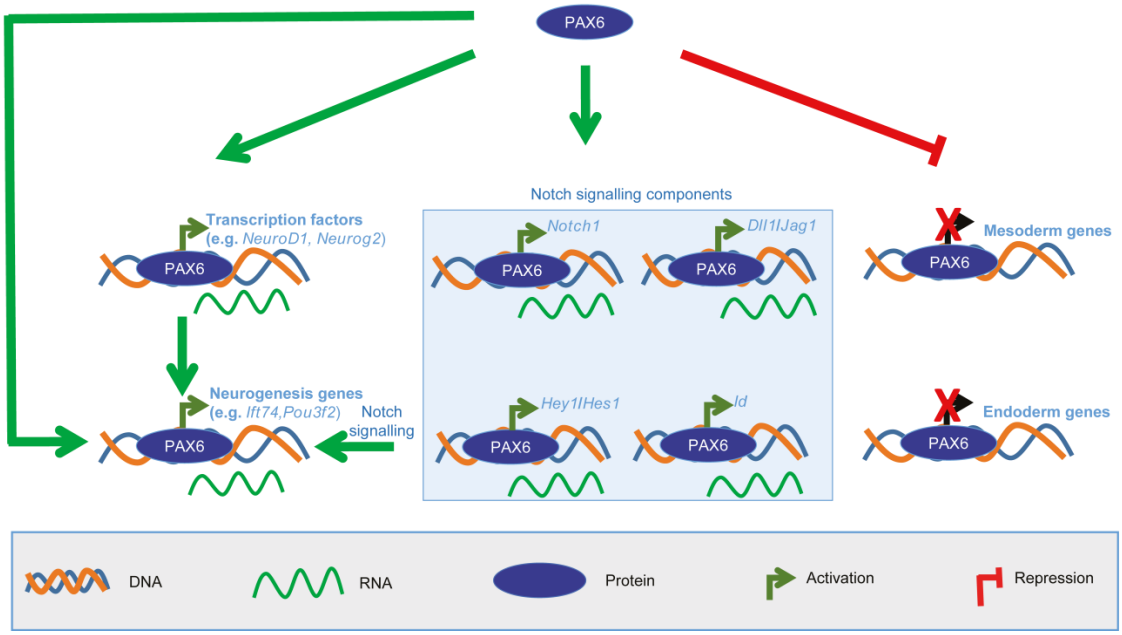



***Figure 8: Pax6 regulates the gene-expression program at multiple levels to promote neuronal differentiation.***

Pax6 mediates the activation of neuronal (ectodermal) genes while concurrently represses the mesodermal and endodermal genes, thereby ensuring the unidirectionality of lineage commitment towards neuronal differentiation. Pax6 directly binds and activates expression of critical transcription factors and components of signaling pathways, all of which then function in concert to orchestrate downstream gene-expression program that drives neurogenesis.

We also observed distinct expression dynamics of Pax6 and Sox2 unique target genes compared with those co-occupied by both factors during subsequent stages of neurogenesis. The genes bound by either Pax6 or Sox2 and expressed in aRG showed transcriptional activation in a stage-specific manner during neurogenesis where different gene-sets were found to be expressed in bRG, IPC and neurons. It may reflect availability of distinct factors or signaling pathways that become active at each of these stages to induce a set of genes critical for that particular stage of neuronal differentiation. On the other hand, all genes co-occupied by both Pax6 and Sox2

acquired changes in their expression state immediately after transition from aRG to bRG and this transcription state was maintained during later stages. This may also imply that genes that are required to be immediately switched on or off during differentiation of NPs may in some way benefit from being targeted by both Pax6 and Sox2.

Pax6 is believed to exert its effects by regulating critical downstream effectors during neurogenesis. A number of such examples have already been described, such as Fabp7, Neurog2, $p27^{kip1}$, cell adhesion molecules (for example, L1 optimedin A, R-cadherin, δ-catenin and tenascin C), patterning molecules (for example, secreted frizzled-related protein 2 (sFRP2) and T-cell factor 4 (Tcf4)), Nkx2.2, Hoxd4, as well as other transcription factors, including Nfia, AP-2γ, NeuroD6, Neurog2, Tbr2, and Bhlhb5 [9, 15, 16, 57-63, 68, 69]. Our study is in line with the previous reports of a direct regulation of neurogenic transcription factors by Pax6 [17]. Importantly, our data also identified several additional genes including transcription factors that are directly induced by Pax6 in NP cells but have not been studied in the context of neuronal development (for example, *Bazb2, Hmgn3, Peli2, Vit*). Furthermore, although the role of Pax6 and Notch signaling in neuronal development is known for long, our data provide the first evidence that Pax6 also promotes Notch signaling by directly inducing the expression of critical components of this pathway (Figure 8).

Our data suggest that although Pax6 activates genes related to neuronal development, it represses the transcription of mesodermal and endodermal genes. Furthermore, ISMARA analysis also showed that Pax6 is required for the induction of transcription factors and their targets that elicit neurogenesis (for example, Sox2 and Tfap2b) and repress others that promote non-neuronal lineages (for example, Brachyury, Hnf1a and Myf family of transcription factors). These observations strongly imply that Pax6-driven gene regulatory program functions to ensure the unidirectionality towards neuronal differentiation.

During neurogenesis, NP cells undergo massive morphological and spatial changes that are tightly linked to cytoskeleton changes. For example, neocortical neurons arise by asymmetric division of radial glia progenitors (RG) in the VZ with a bipolar morphology and gradually become multipolar as they reach SVZ/IZ zone and move erratically. Subsequently, these cells undergo a multipolar to bipolar transition and move rapidly along RGs to the top of CP [70]. Defective ciliogenesis has been shown to accompany

defects in neuronal migration in human ciliopathy phenotypes such as Meckel–Gruber syndrome [71]. We find that one of the Pax6-induced genes, *Ift74,* is highly expressed in NP cells as compared with other cell types. It has been shown in human cells that Ift74 and Ift81 build a tubulin-binding module whose binding to tubulin is important for ciliogenesis [52]. Depletion of *Ift74* by *in utero* electroporation during cortical development led to a retention of migrating cells in the lower layer of the cortex. Furthermore, some cells showed multiple small processes, indicating that the knockdown cells are in the multipolar phase in SVZ/IZ region and fail to migrate to the CP. These observations collectively suggest that the regulation of ciliogenesis and/or axonogenesis via *Ift74* might be essential for cortical development. Moreover, *Ift74*-depleted cells showed significantly reduced migration capacity during EMT *in vitro*. Importantly, as Notch signaling is known to be essential for proper radial migration of cortical neurons [72] and since we also found that Notch signaling is required for proper transcriptional induction of *Ift74*, it is likely that the previously observed defects in neuronal migration in the absence of Notch signaling are, at least in part, contributed by a loss of *Ift74* expression.

Taken together our findings provide novel insights into genomic localization and gene regulatory function of Pax6 during cortical development. Here we show that Pax6 targets a distinct class of epigenetically marked gene promoters, a number of which are co-occupied by other critical transcription factors such as Sox2. Our results suggest a model for a dual function of Pax6 upon neuronal commitment where it mediates the activation of neuronal (ectodermal) genes while concurrently represses the mesodermal and endodermal genes to ensure the unidirectionality towards neuronal differentiation. In addition, Pax6 also induces critical signaling pathways that further work together with Pax6 in guiding critical neurogenic events. Our findings highlight how the gene regulatory circuitry organized by a single factor is able to contribute to neuronal development at multiple levels. In addition to many established downstream effectors, this study has identified many novel targets that are bound and activated by Pax6 and warrant further investigation in cortical development. The *in utero* knockdown of one such gene, *Ift74*, during brain development resulted in impaired neuronal polarity and migration of newborn neurons. Overall, these findings reveal how Pax6 functions in the control of neuronal development at multiple levels to ensure unidirectionality and proper execution of the neurogenic program.

# MATERIALS AND METHODS

### Cell culture

WT and Sey ES cells derived from blastocysts (3.5 PC) of mixed 129-C57Bl/6 background (called 159.2) were cultured and differentiated as previously described [19].

### Quantitative RT-PCR

Total RNA isolation, cDNA synthesis and quantitative RT-PCR were performed according to the manufacturer's (Qiagen, Hilden, Germany) guidelines. Primer sequences will be provided upon request.

### ChIP assay

ChIP experiments were performed as previously described [73]. In brief, crosslinked chromatin was sonicated to achieve an average fragment size of 200 bp. Starting with 70 µg of chromatin and 5 µg of antibodies, 1 µl of ChIP material and 1 µl of input material were used for quantitative real-time PCR using specific primers. Primers covering an intergenic region were used as the control. The efficiencies of the PCR amplifications were normalized to those of the PCR products of the intergenic regions. The following antibodies were used: anti-Pax6 (Covance, Munich, Germany), anti-RNA Pol II: N-20 (Santa Crutz, Heidelberg, Germany), anti-H3K4me2: 07-030 (Millipore, Darmstadt, Germany), and anti-H3K27me3. Primer sequences will be available upon request. The ChIP material for Pax6 was used for ChIP-chip and for H3K4me2 and H3K27me3 was used for ChIP-Seq as described later.

### *In utero* electroporation and imaging

The plasmids containing control shRNA (5'-CAACAAGATGAAGAGCACCAA-3') or shIft74 shRNA (5'-CGAGATCAAATGATTGCAGAA-3') were injected into the lateral ventricle of E12.5 mouse brains, which were given an electrical stimulation (34V with 950mA). Four days later, mice were killed and embryonic brains were fixed in 4% paraformaldehyde, dehydrated in 30% sucrose and embedded in O.C.T. (Tissue-Tek, Staufen, Germany) on dry ice. The brains were frozen-sectioned into 12 µm slices with Leica CS3050S. After DNA staining using Hoechst and marker proteins (TBR2-ab23345 from abcam (Cambridge, UK), PAX6-PRB-278P-100, and TUJ1-D13AF00117 from Covance) confocal

images were achieved through Leica TCS SP5 confocal microscope (Biberach, Germany) and analyzed using ImageJ (NIH, Bethesda, MD, USA).

**Cell viability analysis**

NMuMG cells were transfected with shIft74 green fluorescent protein using lipofectamine according to manufacturer's instructions and trypsinized after 48h. After washing the cells two times with cold PBS, 1.5 million cells were resuspended in 100µl Annexin binding buffer (0.1M hydroxyethyl-piperazineethane-sulfonic acid buffer (HEPES; pH 7.4), 1.4M NaCl and 25 mM $CaCl_2$) supplemented with 5 µl Annexin V antibody labeled with APC (BD Pharmingen, Heidelberg, Germany) and incubated 5 min at room temperature in dark. Then cells were washed twice with 1 ml Annexin binding buffer, taken up in 400 µl Annexin binding buffer supplemented with PI and subsequently measured using the BD LSRFortessa Cell Analyzer with BD FACSDiva software (Heidelberg, Germany).

**Migration assay**

NMuMG cells were transfected using Lipofectamine 2000 following manufacturer's protocol with shControl or shIft74 constructs during TGFβ-induced EMT. Cells were again transfected on the second day and fresh TGFβ was added to the culture. On the third day, wound was created using a 200 µl pipette tip. Light microscope images were taken at time 0 and 23 h and the derived data was further analyzed using ImageJ software to quantify closed area after 23 h compared with 0 h.

**Microarray expression data**

The data from the Affymetrix GeneChip Mouse Gene 1.0 ST Arrays were imported into R (ver. 2.11.1), normalized with RMA [74] and annotated with annotation packages from the Bioconductor repository version 2.6 [75]. A modified version of the *t*-test [76] was used to identify the differentially expressed genes. The obtained *P*-values were corrected for by multiple testing using the Benjamini and Hochberg method. The data is deposited in GEO database with accession number GSE75256.

**ChIP-chip data analysis**

The Nimblegen array intensity files from the GEO data set GSE30204 were imported into R (ver. 2.11.1), and the log2 enrichments (log2 bound/input ratios) for each

individual probe were calculated using the package Ringo [77]. The Arrays were loess-normalized using the normalizeWithinArrays function of the Limma package. For each promoter, we calculated the average log2 enrichment values using data obtained from the overlapping probes. Promoters were defined as 900-bp windows (-700, +200) around the transcription start sites for genes defined in the Ensembl database (version 58_37k, http://www.ensembl.org) and the Refseq db (downloaded on 2010-05-28 from http://genome.ucsc.edu). The data is deposited in GEO database with accession number GSE75256.

**Selection of differentially expressed Pax6 targets**

To derive a list of *bona fide* Pax6 targets, we compared the changes in expression of all Pax6 targets in the Pax6 mutant compared with those of the WT. We considered only those Pax6 targets that were at least two-fold differentially expressed with an FDR cutoff of 0.005, which provided a list of genes that were targeted by Pax6 and differentially expressed in the absence of Pax6.

**RNA-seq analysis**

Tissue-specific data sets were obtained from the Gene Expression Omnibus with the following GEO accession numbers: GSE43194 (Heart E11.5), GSM723775 (mouse embryonic fibroblast E13.5), GSE49581 (Lung E14.5), GSM1150322 (Pancreas E15.5) and GSE30765 (VZ, SVZ and CP; E14.5). The reads were aligned to the mouse genome (mm9) using TopHat [78] with default parameters. The aligned reads were then provided as an input for the HTSeq_count utility from the HTSeq package. The raw read count files obtained from HTSeq-count were then processed for differential expression using the DESeq package [79]. The absolute quantification of the transcripts was performed using Cufflinks with default options. The expression data for apical, radial and intermediate progenitors and neurons were taken from Florio *et. al.* [44].

The data sets of VZ, SVZ and CP were obtained from described reference [25] in which authors have used laser microdissection (LSD) technique to separate out the three main layer of cortex from E14.5 embryos. In brief, dorsolateral and medial pallium areas were dissected to obtained progenitor cells residing in VZ layer. This layer mainly contains apical progenitor cells. Basal and intermediate progenitors reside in the SVZ-IZ regions and laser cuts were consistently performed at the border line of VZ and CP, which

resulted in exclusion of subplate neurons. The CP layer neurons comprise of all differentiated neuronal subtypes present in adjacent layers, for example, Cajal-Retzius layer, VIb layer neurons.

The RNA-seq data for *in vivo* cortical neurogenesis contained well-defined populations of apical, basal and intermediate progenitors [44]. Authors in this study used fluorescence-activated cell sorting to isolate different cell populations from mouse neocortex. Apical radial glial were isolated based on cells that were positive for Dil, Prom1 and negative for Tubb3 while basal radial glial were Dil+, Prom- and Tubb3-. Intermediate progenitors (bIPs) were required to be negative for all three markers and neurons were isolated from Dil+, Tubb3+ but Prom- populations. These pure populations of progenitors and neurons were important for our analysis to analyze expression dynamics of Pax6/Sox2 target genes in NSCs in later stages of neurogenesis.

**ChIP-Seq analysis**

The ChIP-Seq data sets for Sox2 were downloaded from GEO (GSE33059). In this study Sox2 ChIP-seq was performed on NPs derived from mouse ES cells. The reads were mapped to the mouse genome (build mm9) using Bowtie (version 0.12.9) [80] with default parameters. The mapped files were processed using MACS (version 2.0.10.2013071) [81] for peak identification using default parameters. Peaks falling at promoters were used to define Sox2 target promoters. H3K4me2 and H3K27me3 ChIP-Seq data was taken from GSE25533. The genes bound by Ascl1 in the differentiating NPs were obtained from Raposo *et al.* [50]. Briefly, in this study authors performed ChIP-Seq for Ascl1 after 18 h of ectopic expression of Ascl1 in NSCs. Pax6 ChIP-Seq peaks were obtained from a recent study [24]. We assigned each peak to the nearest gene and then shortlisted only those peaks that were found to be within ±1 kb around the transcription start site. These genes were then compared with our Pax6 targets.

**Enrichment analysis**

GO cluster and phenotype enrichment analysis was performed using the ToppGene package [82, 83]. Only the top 20 enriched terms from the GO analysis were plotted. Pathway enrichment was performed using Genomatix.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We would like to thank members of the Tiwari lab for their cooperation and critical feedback during the progress of the project. Support from the core facilities of the Institute of Molecular Biology, Mainz, is gratefully acknowledged, especially the microscopy and bioinformatics core facility. We thank Dirk Schübeler, Tim Roloff and Leslie Hoerner (Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland) for their help with the experiments. We also thank Yves-Alain Barde and Vassiliki Nikoletopoulou (University of Basel, Basel, Switzerland) for providing the Pax6 mutant (Sey) cells. We thank Nikolai Schmarowski (University Medical Center, Mainz) and Jan Baumgart (animal facility at JGU Mainz) for the advice and help with the *in utero* electroporation experiments. The research performed at the laboratory of VKT is supported by the Wilhelm Sander Stiftung 2012.009.2, EpiGeneSys RISE1 program, Marie Curie CIG 322210 and Deutsche Forschungsgemeinschaft (DFG) Grant TI 799/1-1.

## AUTHOR CONTRIBUTIONS

ST performed computational analysis, analyzed data and wrote the manuscript. NT performed experiments (ChIPs, Sey mutant and WT RNA-microarray, migration assay), analyzed data and wrote the manuscript. SS and AG performed experiments (SS: analyses of the Pax6 candidates, Notch inhibitor treatments during neurogenesis, generation of shRNAs, processing of the brains after IUE; AG: quantification of the IUE brain images, cell viability analysis). RI analyzed the data. BB provided suggestions. VT designed the study, analyzed data and wrote the manuscript. All authors read and approved the final manuscript.

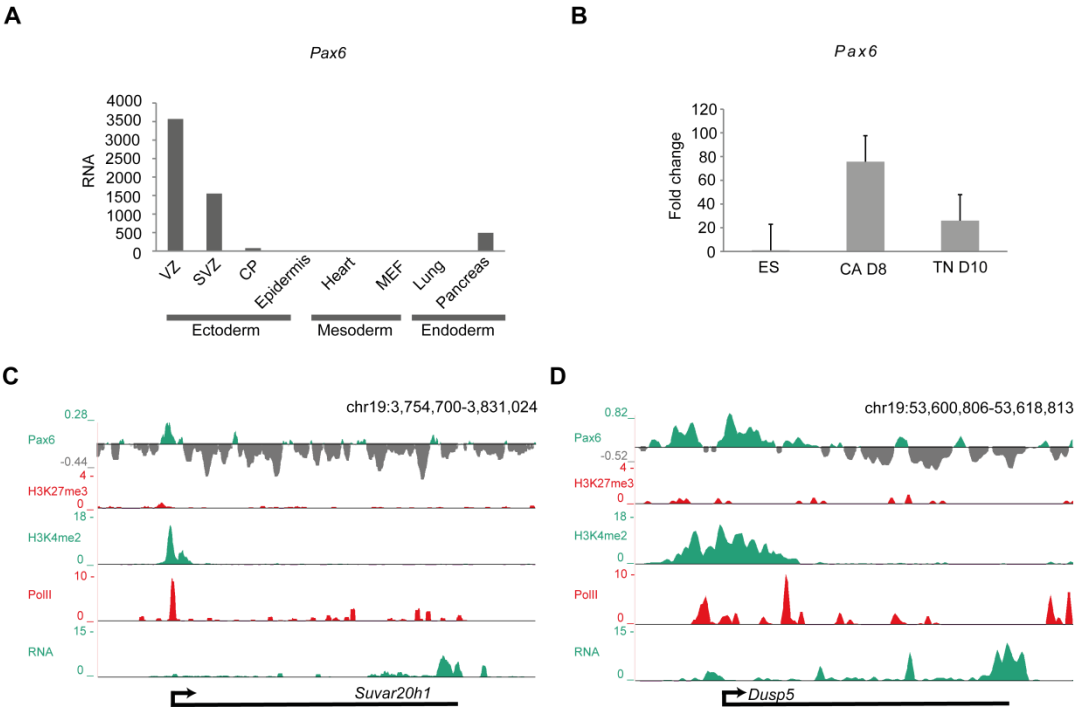
# REFERENCES


1. Glaser, T., D.S. Walton, and R.L. Maas, *Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene.* Nat Genet, 1992. **2**(3): p. 232-9.
2. Duan, D., et al., *Spatiotemporal expression patterns of Pax6 in the brain of embryonic, newborn, and adult mice.* Brain Struct Funct, 2013. **218**(2): p. 353-72.
3. Stapleton, P., et al., *Chromosomal localization of seven PAX genes and cloning of a novel family member, PAX-9.* Nat Genet, 1993. **3**(4): p. 292-8.
4. Walther, C., et al., *Pax: a murine multigene family of paired box-containing genes.* Genomics, 1991. **11**(2): p. 424-34.
5. Gehring, W.J., *The master control gene for morphogenesis and evolution of the eye.* Genes Cells, 1996. **1**(1): p. 11-5.
6. Davis, L.K., et al., *Pax6 3' deletion results in aniridia, autism and mental retardation.* Hum Genet, 2008. **123**(4): p. 371-8.
7. Washington, N.L., et al., *Linking human diseases to animal models using ontology-based phenotype annotation.* PLoS Biol, 2009. **7**(11): p. e1000247.
8. van Heyningen, V. and K.A. Williamson, *PAX6 in sensory development.* Hum Mol Genet, 2002. **11**(10): p. 1161-7.
9. Gotz, M., A. Stoykova, and P. Gruss, *Pax6 controls radial glia differentiation in the cerebral cortex.* Neuron, 1998. **21**(5): p. 1031-44.
10. Sansom, S.N., et al., *The level of the transcription factor Pax6 is essential for controlling the balance between neural stem cell self-renewal and neurogenesis.* PLoS Genet, 2009. **5**(6): p. e1000511.
11. Hill, R.E., et al., *Mouse small eye results from mutations in a paired-like homeobox-containing gene.* Nature, 1991. **354**(6354): p. 522-5.
12. Heins, N., et al., *Glial cells generate neurons: the role of the transcription factor Pax6.* Nat Neurosci, 2002. **5**(4): p. 308-15.
13. Jones, L., et al., *Pax6 is required for the normal development of the forebrain axonal connections.* Development, 2002. **129**(21): p. 5041-52.
14. Nikoletopoulou, V., et al., *Neurotrophin receptor-mediated death of misspecified neurons generated from embryonic stem cells lacking Pax6.* Cell Stem Cell, 2007. **1**(5): p. 529-40.
15. Dohrmann, C., P. Gruss, and L. Lemaire, *Pax genes and the differentiation of hormone-producing endocrine cells in the pancreas.* Mech Dev, 2000. **92**(1): p. 47-54.
16. Kioussi, C., et al., *Pax6 is essential for establishing ventral-dorsal cell boundaries in pituitary gland development.* Proc Natl Acad Sci U S A, 1999. **96**(25): p. 14378-82.
17. Osumi, N., et al., *Concise review: Pax6 transcription factor contributes to both embryonic and adult neurogenesis as a multifunctional regulator.* Stem Cells, 2008. **26**(7): p. 1663-72.
18. Gosmain, Y., et al., *Pax6 controls the expression of critical genes involved in pancreatic {alpha} cell differentiation and function.* J Biol Chem, 2010. **285**(43): p. 33381-93.
19. Bibel, M., et al., *Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells.* Nat Protoc, 2007. **2**(5): p. 1034-43.
20. Bibel, M., et al., *Differentiation of mouse embryonic stem cells into a defined neuronal lineage.* Nat Neurosci, 2004. **7**(9): p. 1003-9.
21. Plachta, N., et al., *Developmental potential of defined neural progenitors derived from mouse embryonic stem cells.* Development, 2004. **131**(21): p. 5449-56.
22. Tiwari, V.K., et al., *Target genes of Topoisomerase IIbeta regulate neuronal survival and are defined by their chromatin state.* Proc Natl Acad Sci U S A, 2012. **109**(16): p. E934-43.
23. Lienert, F., et al., *Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells.* PLoS Genet, 2011. **7**(6): p. e1002090.
24. Sun, J., et al., *Identification of in vivo DNA-binding mechanisms of Pax6 and reconstruction of Pax6-dependent gene regulatory networks during forebrain and lens development.* Nucleic Acids Res, 2015. **43**(14): p. 6827-46.
25. Ayoub, A.E., et al., *Transcriptional programs in transient embryonic zones of the cerebral cortex defined by high-resolution mRNA sequencing.* Proc Natl Acad Sci U S A, 2011. **108**(36): p. 14950-5.
26. DeLaughter, D.M., et al., *Spatial transcriptional profile of the chick and mouse endocardial cushions identify novel regulators of endocardial EMT in vitro.* J Mol Cell Cardiol, 2013. **59**: p. 196-204.
27. Shen, Y., et al., *A map of the cis-regulatory sequences in the mouse genome.* Nature, 2012. **488**(7409): p. 116-20.

28. Sauvageau, M., et al., *Multiple knockout mouse models reveal lincRNAs are required for life and brain development.* Elife, 2013. **2**: p. e01749.
29. Meredith, D.M., et al., *Program specificity for Ptf1a in pancreas versus neural tube development correlates with distinct collaborating cofactors and chromatin accessibility.* Mol Cell Biol, 2013. **33**(16): p. 3166-79.
30. Hitoshi, S., et al., *Notch pathway molecules are essential for the maintenance, but not the generation, of mammalian neural stem cells.* Genes Dev, 2002. **16**(7): p. 846-58.
31. Zhou, Z.D., et al., *Notch as a molecular switch in neural stem cells.* IUBMB Life, 2010. **62**(8): p. 618-23.
32. Aguirre, A., M.E. Rubio, and V. Gallo, *Notch and EGFR pathway interaction regulates neural stem cell number and self-renewal.* Nature, 2010. **467**(7313): p. 323-7.
33. Ribes, V., et al., *Distinct Sonic Hedgehog signaling dynamics specify floor plate and ventral neuronal progenitors in the vertebrate neural tube.* Genes Dev, 2010. **24**(11): p. 1186-200.
34. Warren, N., et al., *The transcription factor, Pax6, is required for cell proliferation and differentiation in the developing cerebral cortex.* Cereb Cortex, 1999. **9**(6): p. 627-35.
35. Xu, X., V.L. Browning, and J.S. Odorico, *Activin, BMP and FGF pathways cooperate to promote endoderm and pancreatic lineage cell differentiation from human embryonic stem cells.* Mech Dev, 2011. **128**(7-10): p. 412-27.
36. Ciruna, B. and J. Rossant, *FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak.* Dev Cell, 2001. **1**(1): p. 37-49.
37. Chen, J.C. and D.J. Goldhamer, *Transcriptional mechanisms regulating MyoD expression in the mouse.* Cell Tissue Res, 1999. **296**(1): p. 213-9.
38. Megeney, L.A. and M.A. Rudnicki, *Determination versus differentiation and the MyoD family of transcription factors.* Biochem Cell Biol, 1995. **73**(9-10): p. 723-32.
39. DeLaForest, A., et al., *HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells.* Development, 2011. **138**(19): p. 4143-53.
40. Gomez-Lopez, S., et al., *Sox2 and Pax6 maintain the proliferative and developmental potential of gliogenic neural stem cells In vitro.* Glia, 2011. **59**(11): p. 1588-99.
41. Kamachi, Y., et al., *Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development.* Genes Dev, 2001. **15**(10): p. 1272-86.
42. Narasimhan, K., et al., *DNA-mediated cooperativity facilitates the co-selection of cryptic enhancer sequences by SOX2 and PAX6 transcription factors.* Nucleic Acids Res, 2015. **43**(3): p. 1513-28.
43. Bergsland, M., et al., *Sequentially acting Sox transcription factors in neural lineage development.* Genes Dev, 2011. **25**(23): p. 2453-64.
44. Florio, M., et al., *Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion.* Science, 2015. **347**(6229): p. 1465-70.
45. Geoffroy, C.G., et al., *Engineering of dominant active basic helix-loop-helix proteins that are resistant to negative regulation by postnatal central nervous system antineurogenic cues.* Stem Cells, 2009. **27**(4): p. 847-56.
46. Berninger, B., F. Guillemot, and M. Gotz, *Directing neurotransmitter identity of neurones derived from expanded adult neural stem cells.* Eur J Neurosci, 2007. **25**(9): p. 2581-90.
47. Yun, K., et al., *Modulation of the notch signaling by Mash1 and Dlx1/2 regulates sequential specification and differentiation of progenitor cell types in the subcortical telencephalon.* Development, 2002. **129**(21): p. 5029-40.
48. Horton, S., et al., *Correct coordination of neuronal differentiation events in ventral forebrain requires the bHLH factor MASH1.* Mol Cell Neurosci, 1999. **14**(4-5): p. 355-69.
49. Casarosa, S., C. Fode, and F. Guillemot, *Mash1 regulates neurogenesis in the ventral telencephalon.* Development, 1999. **126**(3): p. 525-34.
50. Raposo, A.A., et al., *Ascl1 Coordinately Regulates Gene Expression and the Chromatin Landscape during Neurogenesis.* Cell Rep, 2015.
51. Lein, E.S., et al., *Genome-wide atlas of gene expression in the adult mouse brain.* Nature, 2007. **445**(7124): p. 168-76.
52. Bhogaraju, S., et al., *Molecular basis of tubulin transport within the cilium by IFT74 and IFT81.* Science, 2013. **341**(6149): p. 1009-12.
53. Abranches, E., et al., *Neural differentiation of embryonic stem cells in vitro: a road map to neurogenesis in the embryo.* PLoS One, 2009. **4**(7): p. e6286.
54. Sahu, S.K., et al., *JNK-dependent gene regulatory circuitry governs mesenchymal fate.* EMBO J, 2015. **34**(16): p. 2162-81.
55. Maeda, M., K.R. Johnson, and M.J. Wheelock, *Cadherin switching: essential for behavioral but not morphological changes during an epithelium-to-mesenchyme transition.* J Cell Sci, 2005. **118**(Pt 5): p. 873-87.

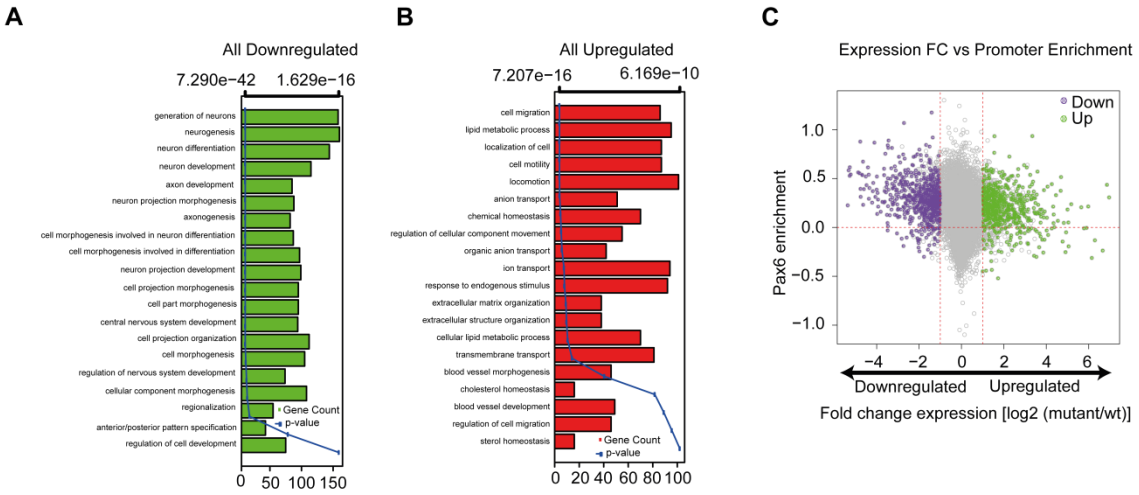
56. Tiwari, N., et al., *Sox4 is a master regulator of epithelial-mesenchymal transition by controlling Ezh2 expression and epigenetic reprogramming.* Cancer Cell, 2013. **23**(6): p. 768-83.
57. Andrews, G.L. and G.S. Mastick, *R-cadherin is a Pax6-regulated, growth-promoting cue for pioneer axons.* J Neurosci, 2003. **23**(30): p. 9873-80.
58. Arai, Y., et al., *Role of Fabp7, a downstream gene of Pax6, in the maintenance of neuroepithelial cells during early embryonic development of the rat cortex.* J Neurosci, 2005. **25**(42): p. 9752-61.
59. O'Flaherty, E. and J. Kaye, *TOX defines a conserved subfamily of HMG-box proteins.* BMC Genomics, 2003. **4**(1): p. 13.
60. Kim, K., et al., *Dendrite-like process formation and cytoskeletal remodeling regulated by delta-catenin expression.* Exp Cell Res, 2002. **275**(2): p. 171-84.
61. Nolte, C., et al., *Stereospecificity and PAX6 function direct Hoxd4 neural enhancer activity along the antero-posterior axis.* Dev Biol, 2006. **299**(2): p. 582-93.
62. Scardigli, R., et al., *Direct and concentration-dependent regulation of the proneural gene Neurogenin2 by Pax6.* Development, 2003. **130**(14): p. 3269-81.
63. Stoykova, A., et al., *Pax6-dependent regulation of adhesive patterning, R-cadherin expression and boundary formation in developing forebrain.* Development, 1997. **124**(19): p. 3765-77.
64. Cvekl, A., et al., *Regulation of gene expression by Pax6 in ocular cells: a case of tissue-preferred expression of crystallins in lens.* Int J Dev Biol, 2004. **48**(8-9): p. 829-44.
65. Kondoh, H., M. Uchikawa, and Y. Kamachi, *Interplay of Pax6 and SOX2 in lens development as a paradigm of genetic switch mechanisms for cell differentiation.* Int J Dev Biol, 2004. **48**(8-9): p. 819-27.
66. Wood, H.B. and V. Episkopou, *Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages.* Mech Dev, 1999. **86**(1-2): p. 197-201.
67. Yuan, H., et al., *Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3.* Genes Dev, 1995. **9**(21): p. 2635-45.
68. Hanson, I. and V. Van Heyningen, *Pax6: more than meets the eye.* Trends Genet, 1995. **11**(7): p. 268-72.
69. Marquardt, T., et al., *Pax6 is required for the multipotent state of retinal progenitor cells.* Cell, 2001. **105**(1): p. 43-55.
70. Cooper, J.A., *Molecules and mechanisms that regulate multipolar migration in the intermediate zone.* Front Cell Neurosci, 2014. **8**: p. 386.
71. Valente, E.M., et al., *Primary cilia in neurodevelopmental disorders.* Nat Rev Neurol, 2014. **10**(1): p. 27-36.
72. Hashimoto-Torii, K., et al., *Interaction between Reelin and Notch signaling regulates neuronal migration in the cerebral cortex.* Neuron, 2008. **60**(2): p. 273-84.
73. Thakurela, S., et al., *Gene regulation and priming by topoisomerase IIalpha in embryonic stem cells.* Nat Commun, 2013. **4**: p. 2478.
74. Carvalho, B.S. and R.A. Irizarry, *A framework for oligonucleotide microarray preprocessing.* Bioinformatics, 2010. **26**(19): p. 2363-7.
75. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics.* Genome Biol, 2004. **5**(10): p. R80.
76. Smyth, G.K., J. Michaud, and H.S. Scott, *Use of within-array replicate spots for assessing differential expression in microarray experiments.* Bioinformatics, 2005. **21**(9): p. 2067-75.
77. Toedling, J., et al., *Ringo--an R/Bioconductor package for analyzing ChIP-chip readouts.* BMC Bioinformatics, 2007. **8**: p. 221.
78. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics, 2009. **25**(9): p. 1105-11.
79. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.
80. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
81. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
82. Kaimal, V., et al., *ToppCluster: a multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W96-102.
83. Chen, J., et al., *ToppGene Suite for gene list enrichment analysis and candidate gene prioritization.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W305-11.

# SUPPLEMENTARY FIGURES



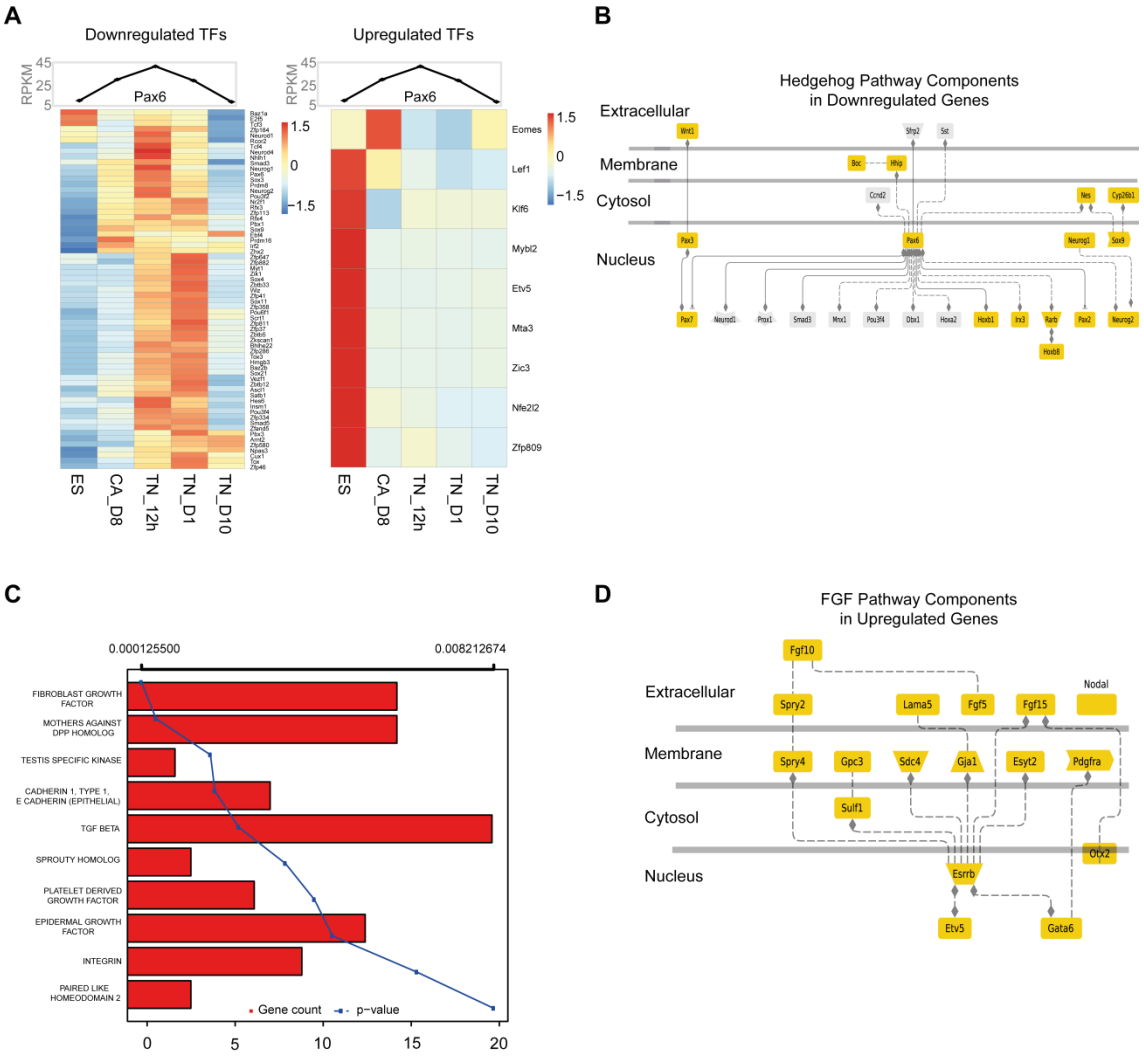

***Supplementary Figure 1. Pax6 is expressed in neuronal progenitors in vivo and in vitro.***

**A.** Bar plot showing normalized RPKM values of Pax6 in several cell types (VZ; Ventricular Zone, SVZ; Sub-Ventricular Zone, CP; Cortical Plate, MEF: mouse embryonic fibroblasts). **B.** qRT-PCRs showing that *Pax6* is specifically upregulated in neuronal progenitors during *in-vitro* differentiation of ES cells into neurons (ES; Embryonic Stem Cells, CA D8; Cluster aggregates or Neuronal Progenitors (NP), TN; Terminal Neurons). Error bars represent SEM. **C-D.** Browser plots showing Pax6, H3K4me2, H3K27me3, Poll II and RNA levels at Suvar20h1 (C) and Dusp5 (D) promoters.



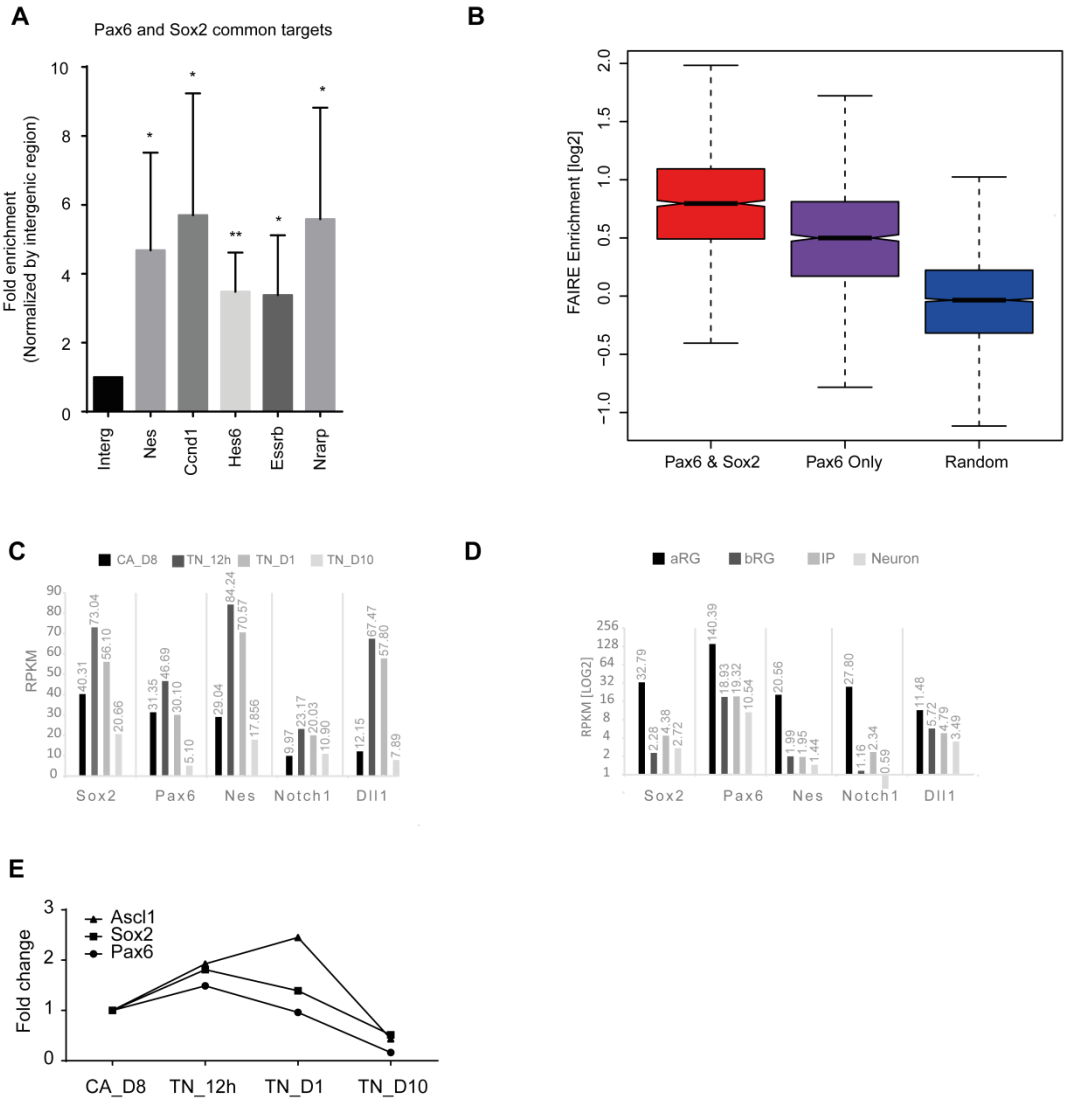

***Supplementary Figure 2.***

**A-B.** Bar and line plots showing GO term enrichment analysis of all downregulated (A) and upregulated (B) genes in Pax6 mutant progenitors. Bar plots show numbers of genes for each enriched GO term (main x-axis), and the line represents the P-value for corresponding GO term (alternate x-axis). **C**. Scatter plot showing expression changes in *Pax6* mutant in comparison with Pax6 enrichment with 0 as the cut-off (red horizontal line). The x-axis represents log2-fold change between WT and *Pax6* mutant cells, and the y-axis represents Pax6 enrichment. Red vertical lines represent 2-fold change in expression.




***Supplementary Figure 3.***

**A.** Heat maps showing expression of transcription factors that were found misregulated (down or up regulated) in Pax6 mutant cells during *in vitro* neurogenesis stages of wild-type ES cells. **B.** Network of Hedgehog pathway components among Pax6 target genes that are downregulated in *Pax6* mutant progenitors derived using Genomatix. **C.** Bar and line plots showing signaling pathway enrichment analysis of upregulated genes in *Pax6* mutant progenitors. Bar plots show number of genes for each enriched GO term (main x-axis), and the line represents the p-value for corresponding GO term (alternate x-axis). **D.** Network of FGF pathway components among Pax6 target genes that are upregulated in *Pax6* mutant progenitors derived from Genomatix.




***Supplementary Figure 4.***

**A.** ChIP-qPCR validations of common Pax6 and Sox2 target genes showing the Pax6 and Sox2 enrichment at their promoters. **B.** Box plot showing accessibility (FAIRE enrichment) of Pax6/Sox2 common, Pax6 only and random target promoters. **C-D.** Expression of selected progenitor genes during *in vitro* (C) and *in vivo* neurogenesis (D). **E.** Line plot showing the normalized RPKM values for Pax6, Sox2 and Ascl1 during *in vitro* neurogenesis. The data is plotted as fold change in relation to CA_D8 stage.

**A**

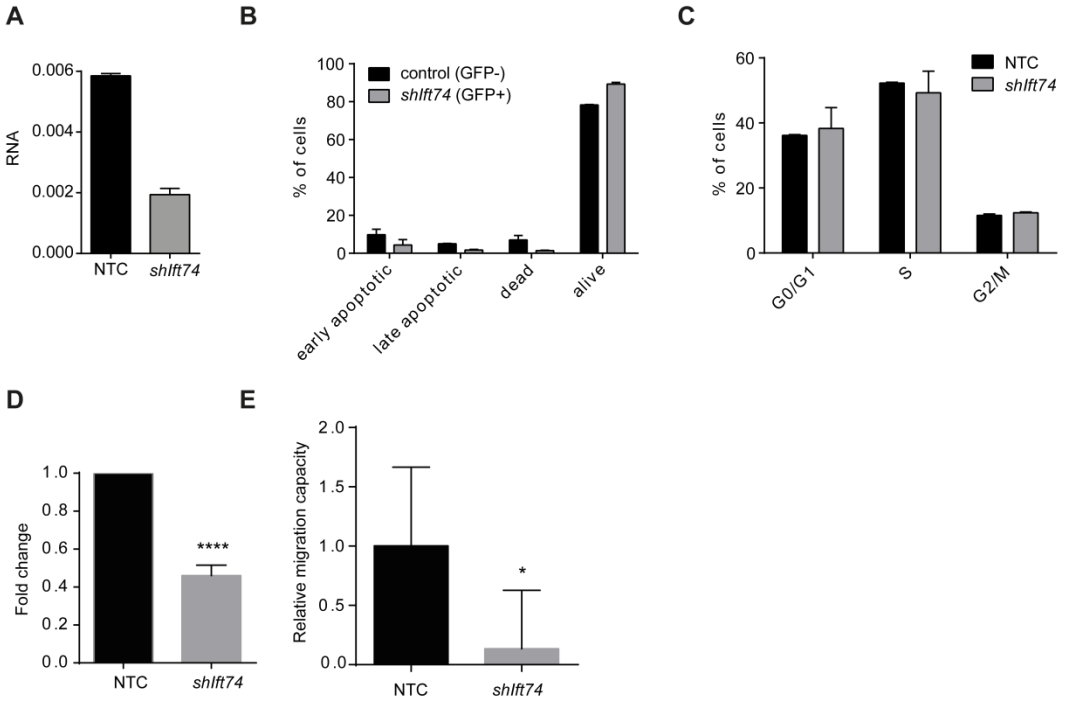
| Function known in progenitors |
|---|
| Cntfr |
| Dll1 |
| Fzd8 |
| Gadd45g |
| Gpr98 |
| Hap1 |
| Hes6 |
| Insm1 |
| Mfng |
| Nes |
| Neurod1 |
| Neurod4 |
| Neurog1 |
| Neurog2 |
| Nhlh1 |
| Npas3 |
| Nrarp |
| Pax6 |
| Pou3f2 |
| Sema5b |
| Sox21 |
| Sstr2 |
| Tcf3 |
| Wnt7a |
| Zhx2 |

**B**

| Function not known in progenitors |
|---|
| Abcd2 |
| Baz2b |
| Ckb |
| Fam181b |
| Fstl5 |
| Gkap1 |
| Gm5607 |
| Hmgn3 |
| Ift74 |
| Kif21a |
| Nmral1 |
| Peli2 |
| Pou3f4 |
| Rab8b |
| Syne2 |
| Vit |
| Cyp46a1 |
| Llgl1 |
| Necab2 |
| Tox3 |
| Zfp536 |

***Supplementary Figure 5.***

**A-B.** List of Pax6 target genes whose function in neuronal progenitors are known (A) or unknown (B). These data were derived using the list from Fig. 4B, followed by a literature survey.




***Supplementary Figure 6.***

**A.** shRNA against *Ift74* significantly depletes its level in NMuMG cells. Knockdown efficiency of the shRNA against *Ift74* was tested using RT-qPCRs. An shRNA against luciferase was used as control. **B.** NMuMG cells were transfected with *shIft74*-GFP and sorted for GFP after 48h. For GFP positive (shIft74 transfected) and negative (non-transfected) populations Annexin V/PI FACS was performed and percentage of early apoptotic (Annexin V positive), late apoptotic (Annexin V and PI positive), dead (only PI positive) and alive cells (double negative) was plotted. Error bars reflect S.E.M. of biological replicates. **C.** NMuMG cells were transfected 48h with NTC or *shIft74* and cell cycle was monitored by measuring BrdU incorporation (1 h pulse treatment with BrdU after 47 h of transfection) and total DNA content determined by Hoechst. Percentage of cells in S-phase, G2/M-phase and G0/G1-phase was accessed by FACS and is plotted on the y-axis. Error bars reflect S.E.M. of two biological replicates. **D.** qPCR is performed to analyze the expression of Ift74 after transient transfection of shControl and shIft74 constructs in NMuMG cells upon TGFb-induce EMT. **E.** Wound healing assay to assess the migration capacity of Ift74 depleted NMuMG cells upon TGFb-induced EMT.

## SUPPLEMENTARY TABLES

For supplemental tables please refer to the Excel-files published together with this manuscript.

**Supplementary Table 1.**

All Pax6 target promoters above enrichment 0.

**Supplementary Table 2.**

All differentially expressed genes.

**Supplementary Table 3.**

GO enrichment analysis for Tfap2b, Tcf4, T, Myf Family and Hnf1a target genes.

# 3 Discussion and Future Perspectives

Cell identity is largely defined by its gene expression patterns that enable all housekeeping functions and, at the same time, ensure the specialized functions of a cell. Tight regulation of gene expression is therefore essential for the development and the maintenance of cell identity. However, gene expression patterns allow certain flexibility to respond to external cues. Often a cell encounters various stimuli and needs to coordinate all possible responses in the most optimal way in order to maintain its function and integrity. To this end, cells have evolved mechanisms to ensure the best adaptation to environmental changes and control the gene expression at multiple layers and by various mechanisms during each step of the involved processes. For instance, a cell specifies which genes are transcribed and at what levels at a given time and also regulates the transcript stability, the translation rate as well as the protein degradation.

Modulation on chromatin level provides an ideal regulatory tool to fine-tune all DNA-templated processes such as transcription. Chromatin is a highly complex structure and its functional properties are determined by its global and local structural organization as well as its composition. It offers immense options for alterations and many mechanisms to establish or erase such modulations have evolved. They include, for example, chromatin remodeling, nucleosome composition variation or histone and DNA modifications. The interplay of these epigenetic mechanisms with other regulators of gene expression, such as transcription factors, results in the final gene expression programs. Therefore, it is important to understand the mechanisms of each of these processes, how they influence each other and how they are coordinated together in order to allow a cell to differentiate, to maintain its attributes or to respond to stimuli.

This thesis provides insights into how gene regulation programs are modulated in response to environmental stimuli that challenge the genome integrity (chapter 1), how they allow adaptation to rhythmic environmental changes (chapter 2) as well as how they enable regulation of cell fate changes during cellular differentiation (chapter 3). The last section of this thesis aims to comprehensively discuss and connect the findings

from these studies and to put them into perspective of the current state of science as well as to highlight the most exciting next directions to follow up on these findings.

## 3.1 Chromatin response to UV irradiation

The genome is continuously confronted by endogenous by-products of chemical reactions inside the cell or by environmental factors, which influence the intactness of the DNA. For instance, skin cells are exposed to UV radiation from the sun that can induce molecular lesions in the DNA. These do not only affect DNA transcription and replication, but can also lead to mutagenesis with consequences such as skin cancer. Cells consequently have evolved a DNA damage response system, which not only ensures DNA repair, but also controls other cellular processes such as transcription, cell proliferation or even cell survival. There is evidence that chromatin modulations, such as chromatin remodeling or posttranslational modifications of histones, also play important roles in this DNA damage response process [1-7]. However, the global changes in chromatin accessibility and epigenetic state in response to UV irradiation have so far not been explored on a genome-wide level and furthermore correlated to the gene expression changes.

The study described in chapter 1 demonstrates that chromatin is an important player in the UV-induced DNA damage response and undergoes genome-wide modulations 6 h after the UV exposure. Previous studies have shown that chromatin is transiently relaxed around the damaged DNA and then restored or even further compacted [2, 5, 8]. A recent study showed that chromatin compaction might be an integral part of the DNA damage response (DDR) by stimulating damage-independent upstream DDR signaling; however, its persistence may inhibit the repair and recovery from the DNA damage [8]. Our analysis revealed that in fibroblast cells, the accessible genome is reduced globally by around one third 6 h after UV-C irradiation of a defined dose. This could be a consequence of local compactions of the many damaged sites. However, also regions that did not change their accessibility state in response to UV irradiation exhibited a gain in the DNA damage mark γ-H2A.X. On the other hand, it may reflect a protection mechanism to inhibit DNA-templated processes globally and to avoid further damage. An argument for this is that the accessibility changes were widely reproducible, while the UV-induced DNA damage is assumed to be random or in part

with sequence dependency [9]. Our study, in accordance with others, indicates that UV-induced DNA damage occurs both in accessible and non-accessible chromatin and therefore a compaction would not protect the DNA from further damage [10, 11]. Nevertheless, a global compaction may reflect a general response to DNA damage, as other types of damages are preferentially induced in open chromatin [12-14]. Moreover, many genes whose promoter or distal regulatory regions lost accessibility also reduced their expression, while a few genes important for the DNA damage response showed an opposite pattern. These observations suggest a controlled selection of the sites changing in response to UV irradiation.

It would further be interesting to investigate, if the observed chromatin compaction is achieved by active chromatin remodeling or by incorporation of new nucleosomes, for example, by exploring the impact of chromatin remodeling factors on the chromatin compaction or by measuring newly incorporated histones with CATCH-IT [15, 16]. As chromatin changes may not only occur locally, examination of the three-dimensional chromatin organization by methods such as high-resolution microscopy and Hi-C before and after UV treatment might add valuable information to the understanding of UV-induced chromatin alterations. Recently, for example, chromatin compaction induced by oxygen and nutrient deprivation in a cardiac muscle cell line has been visualized using high-resolution microscopy [17]. It would be interesting to investigate, if different kinds of stress stimuli lead to similar chromatin reorganizations.

The global chromatin compaction may also be partly responsible for the strong genome-wide reorganization of the histone modification H3K27ac, which marks active promoter and enhancer regions. Our comprehensive analyses disclosed the importance of distal regulatory regions, namely enhancers as well as super-enhancers, in the DNA damage response. These regions were preferentially remodeled in their chromatin features after UV exposure and were to some extent accompanied by expression changes of the neighboring genes. Enhancers and super-enhancers are well known for their function in the spatiotemporal control of gene expression patterns during development and for the regulation of cell identity genes [18-22]. However, their functional spectrum seems to be broader. Recently, NFκB-mediated occurrence and activity regulation of super-enhancers in response to inflammatory stimuli has been reported to drive transcription changes during inflammation [23-25]. Super-enhancers

are also established in cancer cells at genes that promote tumorigenesis and that are sensitive to oncogenic signaling pathways [19]. The study presented in chapter 1 further shows that following UV irradiation super-enhancers are gained next to stress response genes and genes involved in different signaling pathways such as TP53 and MAPK signaling pathways. Latter are well established to be activated in response to UV irradiation [26, 27]. Super-enhancers therefore seem to be responsive to diverse stimuli and to be connected to the signaling pathways.

Our analyses further suggest the involvement of certain DNA-binding factors in the regulation of the observed chromatin changes. For example, CTCF, a multifunctional factor involved in the three-dimensional organization of the genome, was enriched near genomic regions that kept their active chromatin features following UV irradiation. CTCF has previously also been described to act down-stream of the NFκB pathway in response to UV light to protect cells from apoptosis [28]. These results may hint that CTCF plays an important role in the DNA damage response and may guard certain regions, which are possibly important for the three-dimensional organization of the chromatin, from UV-induced chromatin changes and thereby prevent apoptosis.

In contrast to CTCF, the Cellular tumor antigen p53 (TP53) might be involved in the activation of certain enhancers in response to radiation of the cells. TP53 is a factor, whose function in the DNA damage response is very well established [27]. However, its predominant occurrence at and regulation of enhancers in response to DNA damage was only recently reported [29]. The function of TP53 at these enhancers seems not to be regulated by the recruitment of TP53 to these sites in response to DNA damage, since the previous study and ours revealed that TP53 is already bound to these locations in non-stimulated cells [29]. TP53 binds to these regions under standard condition in its inactive form and becomes activated by phosphorylation in response to DNA damage, which might then allow immediate modulation of the chromatin states at these loci [29]. TP53 might mediate the UV-induced chromatin relaxation at these sites via the recruitment of histone acetyltransferases such as EP300 [30]. This elucidates the interplay of several gene regulatory mechanisms in order to coordinate the gene expression changes in response to DNA damage: UV light induces signaling cascades, which activate transcription factors that consequently may contribute to chromatin modulations in order to adjust the gene expression patterns to the requirements of the

DNA damage response. However, the crosstalk between signaling pathways, transcription factors and epigenetic changes needs further experimental validations. Moreover, functional characterizations are required to elucidate the role of the distal regulatory regions in the DNA damage response. For instance, the analysis of their long-range interactions and selected depletion of enhancer regions may provide deeper insights into their functions.

Not only chromatin changes, but also the chromatin status prior to UV exposure seems to impact on the fate of the loci during the DNA damage response. For example, the chromatin state might have influenced whether the loci underwent chromatin accessibility changes following UV exposure. In addition, expressed genes with active promoter features were mainly the genes undergoing expression changes in response to UV irradiation. While these protein-coding genes were mainly down-regulated, several lincRNAs were induced upon UV irradiation. This implies, in line with previous studies, that non-coding RNAs are important regulators of the DNA damage response [31-33]. Their function during the DDR should be further comprehensively investigated, especially since some were located in proximity to stress response genes and were already primed in their chromatin features for immediate up-regulation.

Gene-specific measurements at different time points after UV exposure showed that changes in chromatin accessibility or in H3K27ac levels either occurred simultaneously or preceded the expression changes. This indicates that chromatin alterations regulate the transcriptional changes. In addition, it revealed that chromatin and expression changes of some crucial genes were already established within minutes after the exposure to the DNA damaging agent. These findings should be further followed up on a global level and in high-resolution time courses to better understand the cause-and-consequence relations of the induced alterations. These investigations will offer insights into the dynamics of the DNA damage response, which genes are important at what time and how they are regulated. Ultimately, this will provide information, which could be useful for therapeutic treatment of patients with defects in their DNA damage response [34].

Overall, there is a need for more in-depth studies encompassing different tissues and with modulation of various experimental parameters in order to reveal whether our observations are general features of the DNA damage response or if the observed

changes are, for example, dependent on the UV dose, the DNA damage inducer or the cell type. Moreover, with the improvement of single cell studies, one could investigate the degree of variability of the locations at which DNA damage is induced by UV light or any other DNA damaging agent. In case of reproducible patterns, it would be useful to integrate the chromatin and transcriptome data with genome-wide determined DNA damage sites, as this would allow distinguishing between chromatin and expression changes induced by the DNA damage or by the DNA damage response.

## 3.2 Cellular adaptation to circadian rhythm

Skin cells encounter varying solar UV light exposure during the day-night cycle. Many organisms have evolved a molecular system, called circadian rhythm, to adapt to these daily oscillating environmental conditions. This led to the development of several behaviors and physiological processes occurring with a periodicity of 24 h. These adaptations allow cells to only spend energy when it is required and to anticipate environmental variations in order to maintain cell homeostasis [35]. In terms of DNA damage, for example, various studies have provided evidence that the cellular sensitivity to DNA damage and the DNA damage response, including DNA repair, is controlled in a circadian manner [36-59]. For instance, a recent study showed that UV-induced DNA damage and nucleotide excision repair efficiency in human fibroblast cells depend on the phase of the circadian rhythm at the time of the UV exposure as a consequence of circadianly occurring chromatin condensations [59]. However, other experiments indicated that the functionality of circadian rhythm may not be essential for the DNA damage response and rather a complex interconnection may exist [60-62]. Our studies revealed that several stress response genes showed a cyclical expression pattern in NIH3T3 fibroblast cells synchronized for circadian rhythm. On the other hand, cyclically expressed genes (including Bmal1) were found to be deregulated after UV exposure. This indicates a connection between the circadian rhythm and the DNA damage response.

Besides evidence for circadianly controlled chromatin remodeling and chromatin organization, many histone-modifying enzymes, including Clock itself, which possesses acetyltransferase activity, have been described to play a role in the circadian gene regulation [63-74]. Further reports disclosed genome-wide promoter occupancy

oscillations of the core clock factors, certain histone modifications, such as H3K4me3, H3K9ac and H3K27ac, and RNA polymerase II in accordance with the circadian rhythm [73, 75, 76]. These findings suggest that transcription exhibits temporally separated active and inactive phases within a day.

In the study presented in chapter 2, we identified circadianly expressed genes in NIH3T3 fibroblast cells and then focused on the transcriptional regulators among them as they probably contribute to the diurnal transcription cycles and to the circadian phenotype of a cell. The identification of cyclically expressed genes within a time course experiment is still a challenging task. Our extensive analyses of different data sets have pointed out that standards and optimization for synchronization protocols as well as for the methods to identify cyclically expressed genes are needed to increase the comparability between data sets and to obtain the most realistic outcome of the experiments. For instance, the revealed cyclically expressed genes showed strong dependencies on the experimental sampling rate and on the applied computational identification method. Therefore, we aimed to be rather stringent and considered only genes detected to be cyclically expressed with three out of four applied computational methods.

We could identify 70 potential transcriptional regulators, including transcription factors and epigenetic regulators, which exhibited a cyclical expression pattern in NIH3T3 cells. Most of them have also been found to be expressed in a circadian fashion in at least one other cell or tissue type. Considering the difficulties in the detection of cyclically expressed genes and that the overlap of circadianly expressed genes between different cell and tissue types is described to be low, the results suggest that some may have an important function in the circadian gene expression regulation [77-81].

One of the detected factors was the zinc finger protein Zfp28. It showed a cyclical expression pattern in NIH3T3 cells as well as in different mouse tissues. However, likely due to its low amplitude, it was not always detected by computational approaches to be circadianly expressed. Our analyses suggest that its transcription is under the control of the core clock factor Bmal1. Zfp28 itself also influenced the core clock network, as its knockdown resulted in deregulation of some core clock factors and in a shortening of the circadian period length by one hour. However, further molecular analyses, including the identification of its genomic target sites by ChIP-seq and its interaction partners by

mass spectrometry, are required to elucidate its function in circadian rhythm.

Leo1 is a component of the PAF complex that associates with the RNA polymerase II and with a H3K4 methyltransferase complex [82]. It cycled in phase with Bmal1 in NIH3T3 cells and in diverse mouse tissues. Leo1 is likely under the transcriptional control of classical core clock targets of CLOCK and BMAL1. The knockdown of Leo1 did not influence the circadian rhythmicity of NIH3T3 cells, but led to an expression deregulation of some core clock target genes of CLOCK and BMAL1. These results hint that Leo1 might act together with Bmal1 to promote the transcription of the Bmal1 targets maybe by mediating H3K4 methylation. To investigate that hypothesis, it would be required to investigate if Leo1 and Bmal1 interact and are co-occupied at certain target regions. Also the effects of absent Leo1 at these regions on H3K4me3 levels and on the transcription of the Bmal1 target genes would provide useful information to assess the hypothesis.

Not only the core clock factors are responsible for the pool of cyclically expressed genes, but also their downstream targets can contribute to such expression patterns [83]. Therefore, it would be interesting to also study the function of the other identified factors in regard to circadian rhythm and to expand the circadian gene regulatory network. Mathematical modeling has proven to be helpful in the investigation of circadian gene regulatory networks [84-90]. The proposed mathematical modeling approach of our study can be applied to investigate if a cyclically expressed gene might regulate the circadian expression of another gene. It has been experimentally confirmed that these modelling analyses provide hints for a potential regulatory connection between two genes. By utilizing further data sets, for example experimentally determined RNA and protein stability information of the tested genes, one could more precisely define their potential relationship. These additional input data would also decrease the variables in the model and allow extending it, for example to test the regulation of a gene by multiple factors.

Circadian gene expression is regulated on transcriptional, post-transcriptional and post-translational level [91-105]. Future studies should aim to gain further insights into the interplay of these diverse regulatory mechanisms and into their individual impact on the circadian phenotype of the cell.

Non-coding RNAs often fulfill transcriptional and post-transcriptional regulatory

functions. In other studies and in our analyses, lncRNAs with circadian expression have been identified [81, 106, 107]. However, the function of these lncRNAs during circadian rhythm needs to be explored in more detail. Some lncRNAs could be part of the core clock network or mediate the expression pattern of other genes. The importance of ncRNAs in circadian rhythm is illustrated by their deregulation in diseases. For instance, a spliced lncRNA regulates the diurnal energy expenditure of the brain and its loss causes the Prader-Willi syndrome (PWS) [108]. The lncRNA highly upregulated in liver cancer (HULC) promotes hepatocarcinogenesis by perturbing the circadian rhythm in hepatoma cells [109].

As circadian rhythm is involved in the regulation of many physiological processes and is deregulated in many diseases, the decipherment of the involved gene regulatory networks and the underlying mechanisms will provide useful hints for therapeutic approaches in disease prevention, detection and treatment. This would especially be helpful for shift workers, who are challenging their internal circadian clocks frequently [110].

## 3.3 Regulation of cell fate change during neurogenesis

Similarly, as shown in this thesis for fibroblast cells, researchers have performed *in vitro* luminescence measurements of circadian rhythm in mouse embryonic stem cells. While they were not able to detect any rhythmicity in these cells, they observed an establishment of stable circadian oscillations during neurogenesis starting from the neuronal precursor state and these oscillations were lost again upon reprogramming [111, 112]. Furthermore, circadian rhythm and some of its components have been implicated to be critical regulators of adult neurogenesis and cell fate determination [113-117]. During embryonic development, the expression of some core clock factors in whole mice embryos have been shown to progressively increase during the time embryonic neurogenesis takes place [118]. However, a role of these factors and the occurrence of oscillations during embryonic neurogenesis need further investigations [83, 118].

A well-established factor important for embryonic as well as adult neurogenesis is the homeodomain transcription factor PAX6 [119-122]. This “master regulator of cortical development” is known to regulate the cell fate of neural progenitor cells to proliferate

or to differentiate into glutamatergic excitatory neurons [122]. Our analyses presented in chapter 3 have provided new insights how PAX6 stimulates neuronal differentiation of the progenitor cells. PAX6 activated genes important for neurogenesis either directly through targeting these genes itself or indirectly by activating other transcription factors and maybe also via the induction of components of the Notch signaling pathway. Furthermore, PAX6 seems to act as a gatekeeper by driving the unidirectional differentiation into the ectodermal linage not only by activating genes important for this process, but also by repressing genes necessary for the generation of cells from other linages. In addition, PAX6 has been shown to inhibit the transcription of *Olig2*, a transcription factor required for gliogenesis, and thereby promoting a neuronal fate of the progenitors [123]. Nevertheless, PAX6 also contributes to the spatiotemporal control of gliogenesis, which follows after the completion of neurogenesis during the embryonic development [124-129].

A factor that combines two opposite functions, scilicet an activating as well as repressing role, implies that its function likely depends on the timely and spatially controlled function of additional factors. SOX2, a factor that has previously been described to co-occur at specific genes with PAX6 during lens development and whose function also depends on its co-factors, constitutes a potential candidate [130-133]. Additionally, published work indicates in line with our data that PAX6 enhances the expression of *Sox2* itself [134]. Our genome-wide analyses revealed many targets, which are occupied by PAX6 and SOX2. These common targets were mainly the active genes, while genes that are either not or lowly expressed are only bound by PAX6. Further studies should be designed to investigate the interplay between these factors and to elucidate the functional contributions of each factor. Since many interaction partners are known for PAX6, including chromatin remodelers, histone modifiers and other transcription factors, one should consider to identify all co-factors of PAX6 and investigate their spatially and temporally combinatorial occurrence and function in order to enlarge the knowledge about the gene regulatory network of PAX6 in neural progenitor cells [122, 135].

Our study revealed an interplay of the transcription factor PAX6 with Notch signaling, two components well known for their importance in neuronal development [122, 136-139]. On the one hand, PAX6 was observed to promote Notch signaling by the direct

induction of Notch signaling components. On the other hand, the expression level of *Pax6* and of some PAX6 targets were reduced in Notch-inhibited neural progenitor cells. Future experiments should be applied to elucidate if this is a result of a feed-forward regulation or if this observation is rather an indirect effect due to a more differentiated state of the Notch-inhibited progenitor cells. These analyses will provide deeper insights into the potential cooperativity of PAX6 and Notch signaling in the regulation of important neurogenesis genes and will contribute to a better understanding of their regulatory relationship.

One of the PAX6 target genes that might also be regulated by the Notch signaling is the Intraflagellar transport protein 74 *(Ift74)*. This gene encodes a component of the Ift complex that mediates the transport of tubulin within the cilium during ciliogenesis [140]. The performed experiments show that IFT74 is also required for neurogenesis and for the migration of cells. Our results indicate that IFT74 is probably important for the transition of the neurons from the multipolar to bipolar state and the following migration of neurons to the cortical plate. These processes also rely on microtubule dynamics [141, 142].

The ability of PAX6 to induce transcription factors required for neurogenesis while at the same time inhibiting the expression of genes relevant for other linages ("gatekeeper" properties) may qualify PAX6 as a potential transdifferentiation factor. Indeed, PAX6 not only induces proliferation and neurogenesis in neural progenitor cells, but can also direct neurogenesis in astrocytes [143, 144]. Also mesodermal cells have been reprogrammed into neural-precursor-like cells by applying PAX6 in combination with other factors under special culture conditions [145, 146]. An important proneural factor in progenitor cells of inhibitory interneurons is the Achaete-scute homolog 1 (Ascl1, Ash1, Mash1), which has also been used successfully in combination with the POU domain, class 3, transcription factor 2 (Pou3f2, Brn2) and the Myelin transcription factor 1-like protein (Myt1l) to reprogram fibroblasts into neurons [147]. For therapeutic purposes of neurodegenerative diseases, it would be beneficial if one could enrich different neural precursor cells with different combinations of factors, including PAX6 and ASCL1, which will then give rise to specific types of neurons. For this, an extended understanding of the gene regulatory networks underlying the different neural precursor cells is essential and should therefore be further explored.

Neurodegeneration has been shown to arise, for example, due to defects in DNA repair [148, 149]. Xeroderma pigmentosum (XP) patients with deficiencies specifically in the transcription-coupled nucleotide excision repair (TC-NER) develop neurological defects such as neuron degeneration, mental retardation or microcephaly; this indicates that post-mitotic cells of the nervous system may especially be susceptible to defects in TC-NER and that the utilization of repair mechanisms between dividing and non-dividing cells differs [148, 150-152]. The longevity of neurons entail the requirement for efficient DNA repair programs in neurons and further, mouse models with deletions for different DNA repair factors demonstrate the importance of DNA repair also for the formation of neurons during development [149, 153-155]. As shown here in this thesis for fibroblast cells or in other studies for different cell types, genome integrity and DNA damage response in the nervous system is also regulated widely by chromatin modulations [156-158]. Targeting epigenetic regulators, such as DNA or histone methyltransferases and histone deacetylases, have been proven useful for the treatment of hematological malignancies and cancer [159-165]. Therefore, studies like the ones shown within this thesis and further exploration of the underlying chromatin mechanisms will likely provide information about potential drug targets suitable for the treatment of DNA damage related, neurodevelopmental or neurological diseases as well as for circadian rhythm disorders.

## 3.4 Concluding Remarks

Each biological process is controlled by a highly defined gene expression program. Nevertheless, these gene expression programs have to be plastic to enable a cell to adapt to environmental changes and to integrate several processes proceeding at the same time, while still fulfilling its specialized function.

In the different chapters of this thesis, specific components of such gene regulatory networks were investigated in detail during different biological processes, namely DNA damage response, adaptation to diurnal rhythms as well as cell fate specification during neuronal differentiation. The studies described in this thesis were mainly performed in cell culture systems by applying different stimuli to achieve high synchrony in the particular process among the cells. This allowed a detailed as well as genome-wide investigation of the processes and has, as shown for some parts by *in vivo* validations,

the potential to reveal fundamental gene regulatory mechanisms and insights into their connectivity.

The observations made in this thesis provide evidence for the tight interplay of different gene regulatory mechanisms coordinated by signaling molecules, transcription factors and epigenetic mechanisms that modulate chromatin features (Figure 1). For example, the occurrence and signaling-dependent activity of specific transcription factors as well as the prior chromatin context seem to influence the chromatin alterations, which occur in response to UV exposure and which ultimately lead to gene expression variations. Most cells further contain a set of circadianly expressed transcription factors and epigenetic regulators that are required for the maintenance of circadian rhythm in the cells and also for the circadian adaptation of biological processes to the time of the day. Finally, it was illustrated that a particular transcription factor can imply diverse activities to drive a certain process, such as neurogenesis, and that its function can depend on the interaction with specific co-factors.

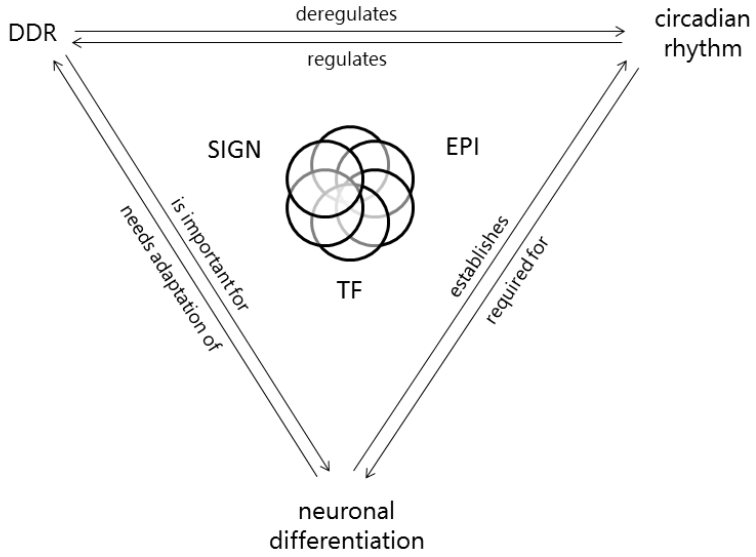



***Figure 1: Different biological processes are regulated and interconnected by diverse gene regulatory mechanisms***

Different gene regulatory mechanisms, such as signaling pathways (SIGN), transcription factors (TF) and epigenetic modifications (EPI), act cooperatively to coordinate cellular processes and their interplay.

Overall, observations in this thesis have revealed several aspects and new connections of the regulatory mechanisms along with the dynamics of the gene regulatory networks during different biological processes. The results presented in this thesis enhance our knowledge about the processes that specify a cell's identity and mediate the cellular response to external stimuli. Ultimately, these findings might be useful for drug development for the diseases where such processes are misregulated.

# References


1. Palomera-Sanchez, Z. and M. Zurita, *Open, repair and close again: chromatin dynamics and the response to UV-induced DNA damage.* DNA Repair (Amst), 2011. **10**(2): p. 119-25.
2. Soria, G., S.E. Polo, and G. Almouzni, *Prime, repair, restore: the active role of chromatin in the DNA damage response.* Molecular Cell, 2012. **46**(6): p. 722-34.
3. Polo, S.E., *Reshaping chromatin after DNA damage: the choreography of histone proteins.* J Mol Biol, 2015. **427**(3): p. 626-36.
4. Adam, S., J. Dabin, and S.E. Polo, *Chromatin plasticity in response to DNA damage: The shape of things to come.* DNA Repair (Amst), 2015. **32**: p. 120-6.
5. Polo, S.E. and G. Almouzni, *Chromatin dynamics after DNA damage: The legacy of the access-repair-restore model.* DNA Repair (Amst), 2015. **36**: p. 114-21.
6. Waters, R., P. van Eijk, and S. Reed, *Histone modification and chromatin remodeling during NER.* DNA Repair (Amst), 2015. **36**: p. 105-13.
7. Dantuma, N.P. and H. van Attikum, *Spatiotemporal regulation of posttranslational modifications in the DNA damage response.* EMBO J, 2016. **35**(1): p. 6-23.
8. Burgess, R.C., et al., *Activation of DNA damage response signaling by condensed chromatin.* Cell Rep, 2014. **9**(5): p. 1703-17.
9. Lee, J., et al., *Visualization of UV-induced damage on single DNA molecules.* Chem Commun (Camb), 2013. **49**(42): p. 4740-2.
10. Zavala, A.G., et al., *High-resolution characterization of CPD hotspot formation in human fibroblasts.* Nucleic Acids Res, 2014. **42**(2): p. 893-905.
11. Yu, W., *Genome-wide analysis of DNA damage and repair.* Technische Universität Darmstadt, 2014. **PhD thesis**.
12. Falk, M., E. Lukasova, and S. Kozubek, *Chromatin structure influences the sensitivity of DNA to gamma-radiation.* Biochim Biophys Acta, 2008. **1783**(12): p. 2398-414.
13. Seo, J., et al., *Genome-wide profiles of H2AX and gamma-H2AX differentiate endogenous and exogenous DNA damage hotspots in human cells.* Nucleic Acids Res, 2012. **40**(13): p. 5965-74.
14. Takata, H., et al., *Chromatin compaction protects genomic DNA from radiation damage.* PLoS One, 2013. **8**(10): p. e75622.
15. Deal, R.B., J.G. Henikoff, and S. Henikoff, *Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones.* Science, 2010. **328**(5982): p. 1161-4.
16. Teves, S.S., R.B. Deal, and S. Henikoff, *Measuring genome-wide nucleosome turnover using CATCH-IT.* Methods Enzymol, 2012. **513**: p. 169-84.
17. Kirmes, I., et al., *A transient ischemic environment induces reversible compaction of chromatin.* Genome Biol, 2015. **16**: p. 246.
18. Kim, T.K. and R. Shiekhattar, *Architectural and Functional Commonalities between Enhancers and Promoters.* Cell, 2015. **162**(5): p. 948-59.
19. Hnisz, D., et al., *Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers.* Molecular Cell, 2015. **58**(2): p. 362-70.
20. Heinz, S., et al., *The selection and function of cell type-specific enhancers.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 144-54.
21. Boland, M.J., K.L. Nazor, and J.F. Loring, *Epigenetic regulation of pluripotency and differentiation.* Circ Res, 2014. **115**(2): p. 311-24.
22. Whyte, W.A., et al., *Master transcription factors and mediator establish super-enhancers at key cell identity genes.* Cell, 2013. **153**(2): p. 307-19.
23. Brown, J.D., et al., *NF-kappaB directs dynamic super enhancer formation in inflammation and atherogenesis.* Molecular Cell, 2014. **56**(2): p. 219-31.
24. Hah, N., et al., *Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs.* Proc Natl Acad Sci U S A, 2015. **112**(3): p. E297-302.
25. Schmidt, S.F., et al., *Acute TNF-induced repression of cell identity genes is mediated by NFkappaB-directed redistribution of cofactors from super-enhancers.* Genome Res, 2015. **25**(9): p. 1281-94.
26. Muthusamy, V. and T.J. Piva, *The UV response of the skin: a review of the MAPK, NFkappaB and TNFalpha signal transduction pathways.* Arch Dermatol Res, 2010. **302**(1): p. 5-17.
27. Latonen, L. and M. Laiho, *Cellular UV damage responses--functions of tumor suppressor p53.* Biochim Biophys Acta, 2005. **1755**(2): p. 71-89.

28. Li, T. and L. Lu, *Functional role of CCCTC binding factor (CTCF) in stress-induced apoptosis.* Exp Cell Res, 2007. **313**(14): p. 3057-65.
29. Younger, S.T., et al., *Integrative genomic analysis reveals widespread enhancer regulation by p53 in response to DNA damage.* Nucleic Acids Res, 2015. **43**(9): p. 4447-62.
30. Rubbi, C.P. and J. Milner, *p53 is a chromatin accessibility factor for nucleotide excision repair of DNA damage.* EMBO J, 2003. **22**(4): p. 975-86.
31. Liu, Y. and X. Lu, *Non-coding RNAs in DNA damage response.* Am J Cancer Res, 2012. **2**(6): p. 658-75.
32. Zhang, C. and G. Peng, *Non-coding RNAs: an emerging player in DNA damage response.* Mutat Res Rev Mutat Res, 2015. **763**: p. 202-11.
33. Francia, S., *Non-Coding RNA: Sequence-Specific Guide for Chromatin Modification and DNA Damage Signaling.* Front Genet, 2015. **6**: p. 320.
34. O'Driscoll, M., *Diseases associated with defective responses to DNA damage.* Cold Spring Harb Perspect Biol, 2012. **4**(12).
35. Gerhart-Hines, Z. and M.A. Lazar, *Circadian metabolism in the light of evolution.* Endocr Rev, 2015. **36**(3): p. 289-304.
36. Fu, L., et al., *The circadian gene Period2 plays an important role in tumor suppression and DNA damage response in vivo.* Cell, 2002. **111**(1): p. 41-50.
37. Antoch, M.P., R.V. Kondratov, and J.S. Takahashi, *Circadian clock genes as modulators of sensitivity to genotoxic stress.* Cell Cycle, 2005. **4**(7): p. 901-7.
38. Kondratov, R.V. and M.P. Antoch, *Circadian proteins in the regulation of cell cycle and genotoxic stress responses.* Trends Cell Biol, 2007. **17**(7): p. 311-7.
39. Kondratov, R.V., V.Y. Gorbacheva, and M.P. Antoch, *The role of mammalian circadian proteins in normal physiology and genotoxic stress responses.* Curr Top Dev Biol, 2007. **78**: p. 173-216.
40. Hoffman, A.E., et al., *The circadian gene NPAS2, a putative tumor suppressor, is involved in DNA damage response.* Mol Cancer Res, 2008. **6**(9): p. 1461-8.
41. Kim, J., et al., *Clock gene mutation modulates the cellular sensitivity to genotoxic stress through altering the expression of N-methylpurine DNA glycosylase gene.* Biochem Pharmacol, 2009. **78**(8): p. 1075-82.
42. Kang, T.H., et al., *Circadian oscillation of nucleotide excision repair in mammalian brain.* Proc Natl Acad Sci U S A, 2009. **106**(8): p. 2864-7.
43. Kang, T.H. and A. Sancar, *Circadian regulation of DNA excision repair: implications for chrono-chemotherapy.* Cell Cycle, 2009. **8**(11): p. 1665-7.
44. Alhopuro, P., et al., *Mutations in the circadian gene CLOCK in colorectal cancer.* Mol Cancer Res, 2010. **8**(7): p. 952-60.
45. Im, J.S., et al., *Per3, a circadian gene, is required for Chk2 activation in human cells.* FEBS Lett, 2010. **584**(23): p. 4731-4.
46. Antoch, M.P. and R.V. Kondratov, *Circadian proteins and genotoxic stress response.* Circ Res, 2010. **106**(1): p. 68-78.
47. Kang, T.H., et al., *Circadian control of XPA and excision repair of cisplatin-DNA damage by cryptochrome and HERC2 ubiquitin ligase.* Proc Natl Acad Sci U S A, 2010. **107**(11): p. 4890-5.
48. Sancar, A., et al., *Circadian clock control of the cellular response to DNA damage.* FEBS Lett, 2010. **584**(12): p. 2618-25.
49. Kim, H., et al., *DNA damage-induced RORalpha is crucial for p53 stabilization and increased apoptosis.* Molecular Cell, 2011. **44**(5): p. 797-810.
50. Gaddameedhi, S., et al., *Control of skin cancer by the circadian rhythm.* Proc Natl Acad Sci U S A, 2011. **108**(46): p. 18790-5.
51. Geyfman, M., et al., *Brain and muscle Arnt-like protein-1 (BMAL1) controls circadian cell proliferation and susceptibility to UVB-induced DNA damage in the epidermis.* Proc Natl Acad Sci U S A, 2012. **109**(29): p. 11758-63.
52. Yang, X., et al., *Mammalian PER2 regulates AKT activation and DNA damage response.* Biochem Cell Biol, 2012. **90**(6): p. 675-82.
53. Liu, R., et al., *Melatonin enhances DNA repair capacity possibly by affecting genes involved in DNA damage responsive pathways.* BMC Cell Biol, 2013. **14**: p. 1.
54. Kang, T.H. and S.H. Leem, *Modulation of ATR-mediated DNA damage checkpoint response by cryptochrome 1.* Nucleic Acids Res, 2014. **42**(7): p. 4427-34.
55. Palombo, P., M. Moreno-Villanueva, and A. Mangerich, *Day and night variations in the repair of ionizing-radiation-induced DNA damage in mouse splenocytes.* DNA Repair (Amst), 2015. **28**: p. 37-47.

56. Wang, F., et al., *The Circadian Gene Clock Plays an Important Role in Cell Apoptosis and the DNA Damage Response In Vitro.* Technol Cancer Res Treat, 2015.
57. Gotoh, T., et al., *Association of the circadian factor Period 2 to p53 influences p53's function in DNA-damage signaling.* Mol Biol Cell, 2015. **26**(2): p. 359-72.
58. Manzella, N., et al., *Circadian Modulation of 8-Oxoguanine DNA Damage Repair.* Sci Rep, 2015. **5**: p. 13752.
59. Bee, L., et al., *Nucleotide excision repair efficiency in quiescent human fibroblasts is modulated by circadian clock.* Nucleic Acids Res, 2015. **43**(4): p. 2126-37.
60. Antoch, M.P., et al., *Disruption of the circadian clock due to the Clock mutation has discrete effects on aging and carcinogenesis.* Cell Cycle, 2008. **7**(9): p. 1197-204.
61. Gauger, M.A. and A. Sancar, *Cryptochrome, circadian cycle, cell cycle checkpoints, and cancer.* Cancer Res, 2005. **65**(15): p. 6828-34.
62. Gaddameedhi, S., et al., *Effect of circadian clock mutations on DNA damage response in mammalian cells.* Cell Cycle, 2012. **11**(18): p. 3481-91.
63. Guillaumond, F., et al., *Chromatin remodeling as a mechanism for circadian prolactin transcription: rhythmic NONO and SFPQ recruitment to HLTF.* FASEB J, 2011. **25**(8): p. 2740-56.
64. Aguilar-Arnal, L., et al., *Cycles in spatial and temporal chromosomal organization driven by the circadian clock.* Nat Struct Mol Biol, 2013. **20**(10): p. 1206-13.
65. Zhu, B., et al., *Coactivator-Dependent Oscillation of Chromatin Accessibility Dictates Circadian Gene Amplitude via REV-ERB Loading.* Molecular Cell, 2015.
66. Menet, J.S., S. Pescatore, and M. Rosbash, *CLOCK:BMAL1 is a pioneer-like transcription factor.* Genes Dev, 2014. **28**(1): p. 8-13.
67. Etchegaray, J.P., et al., *Rhythmic histone acetylation underlies transcription in the mammalian circadian clock.* Nature, 2003. **421**(6919): p. 177-82.
68. Etchegaray, J.P., et al., *The polycomb group protein EZH2 is required for mammalian circadian clock function.* J Biol Chem, 2006. **281**(30): p. 21209-15.
69. Doi, M., J. Hirayama, and P. Sassone-Corsi, *Circadian regulator CLOCK is a histone acetyltransferase.* Cell, 2006. **125**(3): p. 497-508.
70. Nakahata, Y., et al., *The NAD+-dependent deacetylase SIRT1 modulates CLOCK-mediated chromatin remodeling and circadian control.* Cell, 2008. **134**(2): p. 329-40.
71. Alenghat, T., et al., *Nuclear receptor corepressor and histone deacetylase 3 govern circadian metabolic physiology.* Nature, 2008. **456**(7224): p. 997-1000.
72. Katada, S. and P. Sassone-Corsi, *The histone methyltransferase MLL1 permits the oscillation of circadian gene expression.* Nat Struct Mol Biol, 2010. **17**(12): p. 1414-21.
73. Valekunja, U.K., et al., *Histone methyltransferase MLL3 contributes to genome-scale circadian transcription.* Proc Natl Acad Sci U S A, 2013. **110**(4): p. 1554-9.
74. Hosoda, H., et al., *CBP/p300 is a cell type-specific modulator of CLOCK/BMAL1-mediated transcription.* Mol Brain, 2009. **2**: p. 34.
75. Koike, N., et al., *Transcriptional architecture and chromatin landscape of the core circadian clock in mammals.* Science, 2012. **338**(6105): p. 349-54.
76. Le Martelot, G., et al., *Genome-wide RNA polymerase II profiles and RNA accumulation reveal kinetics of transcription and associated epigenetic changes during diurnal cycles.* PLoS Biol, 2012. **10**(11): p. e1001442.
77. Panda, S., et al., *Coordinated transcription of key pathways in the mouse by the circadian clock.* Cell, 2002. **109**(3): p. 307-20.
78. Storch, K.F., et al., *Extensive and divergent circadian gene expression in liver and heart.* Nature, 2002. **417**(6884): p. 78-83.
79. Menger, G.J., et al., *Circadian profiling of the transcriptome in NIH/3T3 fibroblasts: comparison with rhythmic gene expression in SCN2.2 cells and the rat SCN.* Physiol Genomics, 2007. **29**(3): p. 280-9.
80. Miller, B.H., et al., *Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation.* Proc Natl Acad Sci U S A, 2007. **104**(9): p. 3342-7.
81. Zhang, R., et al., *A circadian gene expression atlas in mammals: implications for biology and medicine.* Proc Natl Acad Sci U S A, 2014. **111**(45): p. 16219-24.
82. Rozenblatt-Rosen, O., et al., *The parafibromin tumor suppressor protein is part of a human Paf1 complex.* Mol Cell Biol, 2005. **25**(2): p. 612-20.
83. Brown, S.A., *Circadian clock-mediated control of stem cell division and differentiation: beyond night and day.* Development, 2014. **141**(16): p. 3105-11.
84. Ueda, H.R., et al., *System-level identification of transcriptional circuits underlying mammalian circadian clocks.* Nat Genet, 2005. **37**(2): p. 187-92.

85. Yan, J., et al., *Analysis of gene regulatory networks in the mammalian circadian rhythm.* PLoS Comput Biol, 2008. **4**(10): p. e1000193.

86. Bozek, K., et al., *Regulation of clock-controlled genes in mammals.* PLoS One, 2009. **4**(3): p. e4882.

87. Relogio, A., et al., *Tuning the mammalian circadian clock: robust synergy of two loops.* PLoS Comput Biol, 2011. **7**(12): p. e1002309.

88. Korencic, A., et al., *The interplay of cis-regulatory elements rules circadian rhythms in mouse liver.* PLoS One, 2012. **7**(11): p. e46835.

89. Westermark, P.O. and H. Herzel, *Mechanism for 12 hr rhythm generation by the circadian clock.* Cell Rep, 2013. **3**(4): p. 1228-38.

90. Korencic, A., et al., *Timing of circadian genes in mammalian tissues.* Sci Rep, 2014. **4**: p. 5782.

91. Kwak, E., T.D. Kim, and K.T. Kim, *Essential role of 3'-untranslated region-mediated mRNA decay in circadian oscillations of mouse Period3 mRNA.* J Biol Chem, 2006. **281**(28): p. 19100-6.

92. Woo, K.C., et al., *Mouse period 2 mRNA circadian oscillation is modulated by PTB-mediated rhythmic mRNA degradation.* Nucleic Acids Res, 2009. **37**(1): p. 26-37.

93. Woo, K.C., et al., *Circadian amplitude of cryptochrome 1 is modulated by mRNA stability regulation via cytoplasmic hnRNP D oscillation.* Mol Cell Biol, 2010. **30**(1): p. 197-205.

94. Morf, J., et al., *Cold-inducible RNA-binding protein modulates circadian gene expression posttranscriptionally.* Science, 2012. **338**(6105): p. 379-83.

95. Partch, C.L., C.B. Green, and J.S. Takahashi, *Molecular architecture of the mammalian circadian clock.* Trends Cell Biol, 2014. **24**(2): p. 90-9.

96. Chen, W., et al., *Regulation of Drosophila circadian rhythms by miRNA let-7 is mediated by a regulatory cycle.* Nat Commun, 2014. **5**: p. 5549.

97. Reddy, A.B. and G. Rey, *Metabolic and nontranscriptional circadian clocks: eukaryotes.* Annu Rev Biochem, 2014. **83**: p. 165-89.

98. Mauvoisin, D., et al., *Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver.* Proc Natl Acad Sci U S A, 2014. **111**(1): p. 167-72.

99. Beckwith, E.J. and M.J. Yanovsky, *Circadian regulation of gene expression: at the crossroads of transcriptional and post-transcriptional regulatory networks.* Curr Opin Genet Dev, 2014. **27**: p. 35-42.

100. Luck, S., et al., *Rhythmic degradation explains and unifies circadian transcriptome and proteome data.* Cell Rep, 2014. **9**(2): p. 741-51.

101. Robles, M.S., J. Cox, and M. Mann, *In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism.* PLoS Genet, 2014. **10**(1): p. e1004047.

102. Kim, S.H., et al., *Rhythmic control of mRNA stability modulates circadian amplitude of mouse Period3 mRNA.* J Neurochem, 2015. **132**(6): p. 642-56.

103. Ki, Y., et al., *Warming Up Your Tick-Tock: Temperature-Dependent Regulation of Circadian Clocks.* Neuroscientist, 2015. **21**(5): p. 503-18.

104. Kojima, S. and C.B. Green, *Circadian genomics reveal a role for post-transcriptional regulation in mammals.* Biochemistry, 2015. **54**(2): p. 124-33.

105. Mauvoisin, D., et al., *Proteomics and circadian rhythms: it's all about signaling!* Proteomics, 2015. **15**(2-3): p. 310-7.

106. Coon, S.L., et al., *Circadian changes in long noncoding RNAs in the pineal gland.* Proc Natl Acad Sci U S A, 2012. **109**(33): p. 13319-24.

107. Vollmers, C., et al., *Circadian oscillations of protein-coding and regulatory RNAs in a highly dynamic mammalian liver epigenome.* Cell Metab, 2012. **16**(6): p. 833-45.

108. Powell, W.T., et al., *A Prader-Willi locus lncRNA cloud modulates diurnal genes and energy expenditure.* Hum Mol Genet, 2013. **22**(21): p. 4318-28.

109. Cui, M., et al., *A long noncoding RNA perturbs the circadian rhythm of hepatoma cells to facilitate hepatocarcinogenesis.* Neoplasia, 2015. **17**(1): p. 79-88.

110. Arendt, J., *Shift work: coping with the biological clock.* Occup Med (Lond), 2010. **60**(1): p. 10-20.

111. Kowalska, E., et al., *The circadian clock starts ticking at a developmentally early stage.* J Biol Rhythms, 2010. **25**(6): p. 442-9.

112. Yagita, K., et al., *Development of the circadian oscillator during differentiation of mouse embryonic stem cells in vitro.* Proc Natl Acad Sci U S A, 2010. **107**(8): p. 3846-51.

113. Kimiwada, T., et al., *Clock genes regulate neurogenic transcription factors, including NeuroD1, and the neuronal differentiation of adult neural stem/progenitor cells.* Neurochem Int, 2009. **54**(5-6): p. 277-85.

114. Borgs, L., et al., *Period 2 regulates neural stem/progenitor cell proliferation in the adult hippocampus.* BMC Neurosci, 2009. **10**: p. 30.

115. Bouchard-Cannon, P., et al., *The circadian molecular clock regulates adult hippocampal neurogenesis by controlling the timing of cell-cycle entry and exit.* Cell Rep, 2013. **5**(4): p. 961-73.

116. Malik, A., et al., *Circadian Clock Genes Are Essential for Normal Adult Neurogenesis, Differentiation, and Fate Determination.* PLoS One, 2015. **10**(10): p. e0139655.

117. Malik, A., et al., *Development of circadian oscillators in neurosphere cultures during adult neurogenesis.* PLoS One, 2015. **10**(3): p. e0122937.

118. Dolatshad, H., A.J. Cary, and F.C. Davis, *Differential expression of the circadian clock in maternal and embryonic tissues of mice.* PLoS One, 2010. **5**(3): p. e9855.

119. Gotz, M., A. Stoykova, and P. Gruss, *Pax6 controls radial glia differentiation in the cerebral cortex.* Neuron, 1998. **21**(5): p. 1031-44.

120. Osumi, N., et al., *Concise review: Pax6 transcription factor contributes to both embryonic and adult neurogenesis as a multifunctional regulator.* Stem Cells, 2008. **26**(7): p. 1663-72.

121. Sansom, S.N., et al., *The level of the transcription factor Pax6 is essential for controlling the balance between neural stem cell self-renewal and neurogenesis.* PLoS Genet, 2009. **5**(6): p. e1000511.

122. Ypsilanti, A.R. and J.L. Rubenstein, *Transcriptional and epigenetic mechanisms of early cortical development: An examination of how Pax6 coordinates cortical development.* J Comp Neurol, 2015.

123. Jang, E.S. and J.E. Goldman, *Pax6 expression is sufficient to induce a neurogenic fate in glial progenitors of the neonatal subventricular zone.* PLoS One, 2011. **6**(6): p. e20894.

124. Sugimori, M., et al., *Combinatorial actions of patterning and HLH transcription factors in the spatiotemporal control of neurogenesis and gliogenesis in the developing spinal cord.* Development, 2007. **134**(8): p. 1617-29.

125. Hochstim, C., et al., *Identification of positionally distinct astrocyte subtypes whose identities are specified by a homeodomain code.* Cell, 2008. **133**(3): p. 510-22.

126. Sakurai, K. and N. Osumi, *The neurogenesis-controlling factor, Pax6, inhibits proliferation and promotes maturation in murine astrocytes.* J Neurosci, 2008. **28**(18): p. 4604-12.

127. Genethliou, N., et al., *Spatially distinct functions of PAX6 and NKX2.2 during gliogenesis in the ventral spinal cord.* Biochem Biophys Res Commun, 2009. **382**(1): p. 69-73.

128. Gomez-Lopez, S., et al., *Sox2 and Pax6 maintain the proliferative and developmental potential of gliogenic neural stem cells In vitro.* Glia, 2011. **59**(11): p. 1588-99.

129. Suzuki, T., R. Takayama, and M. Sato, *eyeless/Pax6 controls the production of glial cells in the visual center of Drosophila melanogaster.* Dev Biol, 2015.

130. Kamachi, Y., et al., *Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development.* Genes Dev, 2001. **15**(10): p. 1272-86.

131. Kondoh, H., M. Uchikawa, and Y. Kamachi, *Interplay of Pax6 and SOX2 in lens development as a paradigm of genetic switch mechanisms for cell differentiation.* Int J Dev Biol, 2004. **48**(8-9): p. 819-27.

132. Cvekl, A., et al., *Regulation of gene expression by Pax6 in ocular cells: a case of tissue-preferred expression of crystallins in lens.* Int J Dev Biol, 2004. **48**(8-9): p. 829-44.

133. Narasimhan, K., et al., *DNA-mediated cooperativity facilitates the co-selection of cryptic enhancer sequences by SOX2 and PAX6 transcription factors.* Nucleic Acids Res, 2015. **43**(3): p. 1513-28.

134. Wen, J., et al., *Pax6 directly modulate Sox2 expression in the neural progenitor cells.* Neuroreport, 2008. **19**(4): p. 413-7.

135. Manuel, M.N., et al., *Regulation of cerebral cortical neurogenesis by the Pax6 transcription factor.* Front Cell Neurosci, 2015. **9**: p. 70.

136. Beatus, P. and U. Lendahl, *Notch and neurogenesis.* J Neurosci Res, 1998. **54**(2): p. 125-36.

137. Tokunaga, A., et al., *Mapping spatio-temporal activation of Notch signaling during neurogenesis and gliogenesis in the developing mouse brain.* J Neurochem, 2004. **90**(1): p. 142-54.

138. Vilas-Boas, F., et al., *A novel reporter of notch signalling indicates regulated and random Notch activation during vertebrate neurogenesis.* BMC Biol, 2011. **9**: p. 58.

139. Imayoshi, I., et al., *Genetic visualization of notch signaling in mammalian neurogenesis.* Cell Mol Life Sci, 2013. **70**(12): p. 2045-57.

140. Bhogaraju, S., et al., *Molecular basis of tubulin transport within the cilium by IFT74 and IFT81.* Science, 2013. **341**(6149): p. 1009-12.

141. Chen, T., et al., *NDUFV2 regulates neuronal migration in the developing cerebral cortex through modulation of the multipolar-bipolar transition.* Brain Res, 2015. **1625**: p. 102-10.

142. Tischfield, M.A. and E.C. Engle, *Distinct alpha- and beta-tubulin isotypes are required for the positioning, differentiation and survival of neurons: new support for the 'multi-tubulin' hypothesis.* Biosci Rep, 2010. **30**(5): p. 319-30.

143. Heins, N., et al., *Glial cells generate neurons: the role of the transcription factor Pax6.* Nat Neurosci, 2002. **5**(4): p. 308-15.

144. Berninger, B., et al., *Functional properties of neurons derived from in vitro reprogrammed postnatal astroglia.* J Neurosci, 2007. **27**(32): p. 8654-64.

145. Sheng, C., et al., *Direct reprogramming of Sertoli cells into multipotent neural stem cells by defined factors.* Cell Res, 2012. **22**(1): p. 208-18.

146. Raciti, M., et al., *Reprogramming fibroblasts to neural-precursor-like cells by structured overexpression of pallial patterning genes.* Mol Cell Neurosci, 2013. **57**: p. 42-53.

147. Vierbuchen, T., et al., *Direct conversion of fibroblasts to functional neurons by defined factors.* Nature, 2010. **463**(7284): p. 1035-41.

148. Madabhushi, R., L. Pan, and L.H. Tsai, *DNA damage and its links to neurodegeneration.* Neuron, 2014. **83**(2): p. 266-82.

149. McKinnon, P.J., *Maintaining genome stability in the nervous system.* Nat Neurosci, 2013. **16**(11): p. 1523-9.

150. Mimaki, T., et al., *Neurological manifestations in xeroderma pigmentosum.* Ann Neurol, 1986. **20**(1): p. 70-5.

151. Iyama, T. and D.M. Wilson, 3rd, *DNA repair mechanisms in dividing and non-dividing cells.* DNA Repair (Amst), 2013. **12**(8): p. 620-36.

152. Kraemer, K.H., et al., *Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype-phenotype relationship.* Neuroscience, 2007. **145**(4): p. 1388-96.

153. Sugo, N., et al., *Neonatal lethality with abnormal neurogenesis in mice deficient in DNA polymerase beta.* EMBO J, 2000. **19**(6): p. 1397-404.

154. Orii, K.E., et al., *Selective utilization of nonhomologous end-joining and homologous recombination DNA repair pathways during nervous system development.* Proc Natl Acad Sci U S A, 2006. **103**(26): p. 10017-22.

155. Barzilai, A. and P.J. McKinnon, *Genome maintenance in the nervous system; insight into the role of the DNA damage response in brain development and disease.* DNA Repair (Amst), 2013. **12**(8): p. 541-2.

156. Dobbin, M.M., et al., *SIRT1 collaborates with ATM and HDAC1 to maintain genomic stability in neurons.* Nat Neurosci, 2013. **16**(8): p. 1008-15.

157. Wang, W.Y., et al., *Interaction of FUS and HDAC1 regulates DNA damage response and repair in neurons.* Nat Neurosci, 2013. **16**(10): p. 1383-91.

158. Pan, L., J. Penney, and L.H. Tsai, *Chromatin regulation of DNA damage repair and genome integrity in the central nervous system.* J Mol Biol, 2014. **426**(20): p. 3376-88.

159. Cortez, C.C. and P.A. Jones, *Chromatin, cancer and drug therapies.* Mutat Res, 2008. **647**(1-2): p. 44-51.

160. Saleh, M.H., L. Wang, and M.S. Goldberg, *Improving cancer immunotherapy with DNA methyltransferase inhibitors.* Cancer Immunol Immunother, 2015.

161. Mottamal, M., et al., *Histone deacetylase inhibitors in clinical studies as templates for new anticancer agents.* Molecules, 2015. **20**(3): p. 3898-941.

162. Fang, M., et al., *Enhancement of NAD+-dependent SIRT1 deacetylase activity by methylselenocysteine resets the circadian clock in carcinogen-treated mammary epithelial cells.* Oncotarget, 2015.

163. Benedetti, R., et al., *Epigenetic-based therapy: From single- to multi-target approaches.* Int J Biochem Cell Biol, 2015. **69**: p. 121-31.

164. Ning, B., et al., *Targeting epigenetic regulations in cancer.* Acta Biochim Biophys Sin (Shanghai), 2016. **48**(1): p. 97-109.

165. Mayes, K., et al., *ATP-dependent chromatin remodeling complexes as novel targets for cancer therapy.* Adv Cancer Res, 2014. **121**: p. 183-233.

# CURRICULUM VITAE OF SANDRA SCHICK

CV was deleted for privacy

# Publications

The following three publications were reproduced/adapted with permission within this thesis:

1) Journal of Cell Science; 2015 Dec 1 ; Epub 2015 Oct 7; 128(23); pp. 4380-4394; doi: 10.1242/jcs.173633.

   **Dynamics of chromatin accessibility and epigenetic state in response to UV damage.**

   Schick S[1], Fournier D[1], Thakurela S[1], Sahu SK[1], Garding A[1], Tiwari VK[2].

   [1] Institute of Molecular Biology (IMB), Mainz, Germany.

   [2] Institute of Molecular Biology (IMB), Mainz, Germany v.tiwari@imb-mainz.de.

   Weblink: http://jcs.biologists.org/content/128/23/4380.long

2) Molecular and Cellular Biology; 2015 Dec 7; 36(4); pp. 545-558; doi: 10.1128/MCB.00701-15.

   **Identifying Novel Transcriptional Regulators with Circadian Expression.**

   Schick S[1], Becker K[2], Thakurela S[1], Fournier D[1], Hampel MH[3], Legewie S[4], Tiwari VK[5]

   [1] Epigenetic Regulation of Development and Disease Group, Institute of Molecular Biology, Mainz, Germany.

   [2] Modelling of Biological Networks Group, Institute of Molecular Biology, Mainz, Germany.

   [3] Dr. Senckenbergisches Chronomedizinisches Institut, Frankfurt am Main, Germany.

   [4] Modelling of Biological Networks Group, Institute of Molecular Biology, Mainz, Germany; S.Legewie@imb-mainz.de.

   [5] Epigenetic Regulation of Development and Disease Group, Institute of Molecular Biology, Mainz, Germany; V.Tiwari@imb-mainz.de.

   Weblink: http://mcb.asm.org/content/36/4/545.long

3) Cell Discovery; 2016 Feb 9; 2:15045; doi: 10.1038/celldisc.2015.45. eCollection 2016.

**Mapping gene regulatory circuitry of Pax6 during neurogenesis.**

Thakurela S[1], Tiwari N[2], Schick S[1], Garding A[1], Ivanek R[3], Berninger B[4], Tiwari VK[1].

[1] Institute of Molecular Biology (IMB), Ackermannweg 4 , Mainz, Germany.

[2] Institute of Physiological Chemistry, University Medical Center of the Johannes Gutenberg University Mainz, Hanns-Dieter-Hüsch-Weg 19 , Mainz, Germany.

[3] Department of Biomedicine, University of Basel, Basel, Switzerland.

[4] Institute of Physiological Chemistry, University Medical Center of the Johannes Gutenberg University Mainz, Hanns-Dieter-Hüsch-Weg 19, Mainz, Germany; Focus Program Translational Neuroscience, Johannes Gutenberg University Mainz, Langenbeckstr. 1, Mainz, Germany.

Weblink: https://www.nature.com/articles/celldisc201545