


5-2013

GENOME-WIDE PROFILING UNVEILS CRITICAL FUNCTIONS OF p53 IN HUMAN EMBRYONIC STEM CELLS

Kadir C. Akdemir

Follow this and additional works at: http://digitalcommons.library.tmc.edu/utgsbs_dissertations

 Part of the [Bioinformatics Commons](#), [Medicine and Health Sciences Commons](#), and the [Molecular Biology Commons](#)

Recommended Citation

Akdemir, Kadir C., "GENOME-WIDE PROFILING UNVEILS CRITICAL FUNCTIONS OF p53 IN HUMAN EMBRYONIC STEM CELLS" (2013). *UT GSBS Dissertations and Theses (Open Access)*. Paper 341.

This Dissertation (PhD) is brought to you for free and open access by the Graduate School of Biomedical Sciences at DigitalCommons@The Texas Medical Center. It has been accepted for inclusion in UT GSBS Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@The Texas Medical Center. For more information, please contact laurel.sanders@library.tmc.edu.

**GENOME-WIDE PROFILING UNVEILS CRITICAL FUNCTIONS
OF p53 IN HUMAN EMBRYONIC STEM CELLS**

By

Kadir Caner Akdemir, B.S.

APPROVED:

Michelle C. Barton, Ph.D., Supervisory Professor

Wei Li, Ph.D., Supervisory Professor

Gabor Balazsi, Ph.D.

Nicholas Navin, Ph.D.

Richard Behringer, Ph.D.

APPROVED:

Dean, The University of Texas

Graduate School of Biomedical Sciences at Houston

**GENOME-WIDE PROFILING UNVEILS CRITICAL FUNCTIONS
OF p53 IN HUMAN EMBRYONIC STEM CELLS**

A

DISSERTATION

Presented to the Faculty of

The University of Texas

Health Science Center at Houston

and

The University of Texas

MD Anderson Cancer Center

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Kadir Caner Akdemir, B.S.

Houston, Texas

May 2013

To my parents and my beloved wife for their devotion and endless support...

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the Creator for providing me the opportunity to study his one of the most exquisite and intricate creations, the cell, with the hope to obtain better understanding of his Greatness and contribute to the world, unveiling some unknowns about one of the most interesting players, p53, in the cell.

I am most grateful to my advisor Dr. Michelle C. Barton, who believed in me when I decided to change my field and gave me the chance to work in her lab with the focus on one of the most appealing phenomena in the living organisms, epigenetics. During my training, she helped me to evolve as a young scientist, being patient with me as well as guiding me, encouraging me and expanding my scientific horizons, meanwhile giving me the freedom to pursue my research interests.

On the same line, I also would like to thank the Center for Cancer Epigenetics at M.D. Anderson Cancer Center, especially Director Dr. Sharon Dent for funding my first transition year as an Epigenetic Scholar. I also would like to thank Dr. Dent for being such a supportive faculty during my graduate school.

I would like to extend my gratitude toward my co-mentor, Dr. Wei Li, who along with his lab members truly transformed my perception of high-throughput data analyses from applying the right tools for the undergoing study to representing the obtained data properly. It has been truly an eye-opening

experience and I will always be grateful to him for treating me as a colleague while educating and challenging me to raise my standards.

I also would like to thank my committee members, Dr. Gabor Balazsi for being patient with me, serving on all my graduate school committees, to Dr. Nicholas Navin for being supportive and making me feel like a fellow scientist and Dr. Ralf Krahe for accepting to serve on my supervisory committee as a last minute call. I would like to acknowledge all of my previous committee members, Drs. Pierre McCrea, Shoudan Liang, Xiaobing Shi and Yin Liu for their patience, guidance, scientific discussion, and intuitive comments.

I am also thankful to my project partners, Dr. Abhinav Jain and Kendra Allton, without whose research this project would not be possible, and to the rest of Barton lab members: Zeynep C. Akdemir, Dr. Srikanth Appikonda, Aundrietta Duncan, Dr. Shiming Jiang, Dr. Chunlei Jin, Jing Li, Dr. Zhaoliang Liu, Dr. Ryan McCarthy, Lindsey Minter, Sabrina Stratton, Kaushik Thakkar and Hui Wei for treating me as a part of a big and warm family.

Last but not least, I thank my closest friends and extended family with whom I enjoy this life: Omer Akdemir, Sahin Akdemir, Yildiz Akdemir, Mehmet and Halise Ates, Dursun and Meliha Coban, Avinash Venkatanarayan, Omer F Bayrak, Fatih Demiroz, Turgut Dogruluk, Aundrietta Duncan, Fatih Kazanci, Erkan Ozturan, Serkan Ozturan, Sabrina Stratton, Fatih Semerci, Harun Sencal, Kaushik Thakkar and Tayfun Tuna.

GENOME-WIDE PROFILING UNVEILS CRITICAL FUNCTIONS OF p53 IN HUMAN EMBRYONIC STEM CELLS

Publication No. _____

Kadir C. Akdemir, B.S.

Michelle C. Barton, Ph.D., Supervisory Professor

Embryonic stem cells (ESCs) possess two unique characteristics: infinite self-renewal and the potential to differentiate into almost every cell type (pluripotency). Recently, global expression analyses of metastatic breast and lung cancers revealed an ESC-like expression program or signature, specifically for cancers that are mutant for p53 function. Surprisingly, although p53 is widely recognized as the guardian of the genome, due to its roles in cell cycle checkpoints, programmed cell death or senescence, relatively little is known about p53 functions in normal cells, especially in ESCs. My **hypothesis** is that p53 has specific transcription regulatory functions in human ESCs (hESCs) that a) oppose pluripotency and b) protect the stem cell genome in response to DNA damage and stress signaling. In mouse ESCs, these roles are believed to coincide, as p53 promotes differentiation in response to DNA damage, but this is unexplored in hESCs.

To determine the biological roles of p53, specifically in hESCs, we mapped genome-wide chromatin interactions of p53 by chromatin immunoprecipitation and massively parallel tag sequencing (ChIP-Seq), and did so under three

different conditions of hESC status: pluripotency, differentiation-initiated and DNA-damage-induced. ChIP-Seq showed that p53 is enriched at distinct, induction-specific gene loci during each of these different conditions. Microarray gene expression analysis and functional annotation of the distinct p53-target genes revealed that p53 regulates specific genes encoding developmental regulators, which are expressed in differentiation-initiated but not DNA-damaged hESCs. We further discovered that, in response to differentiation signaling, p53 binds regions of chromatin that are repressed but also poised for rapid activation by core pluripotency factors OCT4 and NANOG in pluripotent hESCs. In response to DNA damage, genes associated with migration and motility are targeted by p53; whereas, the prime targets of p53 in control of cell death are conserved for p53 regulation in both differentiation and DNA damage.

Our genome-wide profiling and bioinformatics analyses show that p53 occupies a special set of developmental regulatory genes during early differentiation of hESCs and functions in an induction-specific manner. In conclusion, our research unveiled previously unknown functions of p53 in ESC biology, which augments our understanding of one of the most deregulated proteins in human cancers.

TABLE OF CONTENTS

APPROVAL SHEET	i
TITLE PAGE.....	ii
DEDICATION	iii
ACKNOWLEDGEMENTS	iv
ABSTRACT	vi
TABLE OF CONTENTS	viii
LIST OF ILLUSTRATIONS	xi
LIST OF TABLES.....	vii
CHAPTER 1 INTRODUCTION	1
1.1 Pluripotent stem cells	1
1.1.1 Human Embryonic Stem Cells.....	1
1.1.2 ESC-specific Transcription factors.....	2
1.1.3 Chromatin regulators.....	3
1.1.4 Non-coding RNAs	8
1.1.5 Signaling mediators of the ESC state	10
1.1.6 Induced Pluripotent Stem Cells	14
1.1.7 ESC-specific gene expression signatures in human cancer	17
1.2 p53 and Pluripotency	19
1.2.1 p53 acts as barrier to somatic cell reprogramming	19
1.2.2 p53-inactivated cancers display plasticity and loss of differentiation.....	20
1.2.3 p53's function in human ES cell differentiation	22
1.3 Genome-wide protein-chromatin interaction studies	23

1.4 Hypothesis, specific aims and rationale.....	25
CHAPTER 2 MATERIAL AND METHODS	27
2.1 ChIP-Seq Analysis.....	27
2.1.1 Sequencing and read alignment	27
2.1.2 Peak calling	27
2.1.3 Conservation of binding sites.....	28
2.1.4 Motif analysis.....	29
2.1.5 Identifying target genes of p53-bound sites	29
2.1.6 Annotation of p53-target genes.....	30
2.1.7 Integration core ES cell transcription factor binding data	31
2.1.8 Bivalent histone modification analysis.....	32
2.2 Gene Expression Analysis.....	32
CHAPTER 3 RESULTS.....	34
3.1 Genome wide mapping of p53 in hESCs reveal distinct functional binding sites.....	34
3.2 p53 binding sites are enriched for p53 and OCT4:SOX motifs in differentiation	38
3.3 p53 targets developmental transcription factors during differentiation...41	41
3.4 p53 binding sites coincide with ESC transcription factors during differentiation	47
3.5 Transcription of development genes is dependent on p53	55
3.6 p53 targets lose repressive histone marks during differentiation	65
CHAPTER 4 DISCUSSION AND FUTURE DIRECTIONS	71
4.1 Discussion.....	71
4.2 Future Directions.....	79

REFERENCES	81
VITA.....	108

LIST OF ILLUSTRATIONS

Figure 1. Bivalent chromatin domains help to establish embryonic stem cell plasticity ..7	
Figure 2. Genome-wide bivalent chromatin modification maps show significant similarities between human iPS and ES cells	16
Figure 3. High-grade human breast cancers display ES-cell-specific gene expression signature	18
Figure 4. Genome-wide mapping demonstrated unique p53 signatures in hESCs after different treatments	35
Figure 5. Condition specific binding sites of p53 are strikingly distant	36
Figure 6. p53 binding regions are evolutionary conserved among vertebrates.	37
Figure 7. p53, OCT4-SOX2 motifs are enriched within p53 enriched sites	39
Figure 8. p53 motif is present in the genomic regions bound by OCT4 and NANOG...40	
Figure 9. p53 targets distinct set of genes during differentiation and DNA-damage in hESCs.....	42
Figure 10. Developmental transcription factor families are targeted by p53 during hESC differentiation.....	46
Figure 11. Binding profiles of p53, OCT4 and NANOG around human HOX loci	48
Figure 12. p53 binding sites coincide with ESC transcription factors during differentiation.....	50
Figure 13. Association of OCT4 and/or NANOG binding sites with p53.....	51
Figure 14. NANOG and OCT4 binding strengths are much higher at differentiation specific sites	53
Figure 15. Integration of gene expression and p53 binding data in differentiating hESCs	56

Figure 16. GO functional classifications of down-regulated p5357

Figure 17. Up-regulated p53 targets are involved in developmental processes58

Figure 18. Transcription of developmental genes during RA-mediated differentiation is p53-dependent60

Figure 19. Enrichment of p53 at developmental genes results in activation.....62

Figure 20. p53's overlapping targets with OCT4 and NANOG are more robustly expressed during differentiation.....66

Figure 21. GO functional classification results down-regulated p53 targets with overlapping OCT4 and/or NANOG sites67

Figure 22. Bivalent chromatin marks around promoter regions of p53 target genes in pluripotent and differentiating hESCs69

Figure 23. Species-specific binding of p53 in different environmental stimuli.....78

LIST OF TABLES

Table 1. Response specific target genes are involved in different biological process ..	43
Table 2. Significant number of p53 targets during differentiation possess homeobox domain	45

CHAPTER 1 INTRODUCTION

1.1 Pluripotent stem cells

1.1.1 Human Embryonic Stem Cells

Human Embryonic Stem Cells (hESCs) are derived from inner cell mass of blastocyst-stage embryo [1] and possess two unique properties together:

- *Pluripotency*: ability to differentiate into any somatic cell type.
- *Self-renewal*: ability to reproduce indefinitely by staying in the same state (without losing pluripotency characteristics).

Unraveling the molecular mechanisms that preserve ESC properties is important for understanding development, how the ground state is maintained and what are the reasons for developmental disorders; studying tissue differentiation, how the genome is regulated for lineage-specific differentiation; and generating the necessary knowledge to manipulate hESCs as an invaluable tool for regenerative medicine. Over the past decade, a global effort has been underway to deconstruct molecular mechanisms that underlie pluripotency in order to realize and harness the full potential of hESCs. The combined results from genetic, biochemical and genomic studies have revealed an intricate regulatory circuitry of pluripotent state, which contains transcription factors, chromatin regulators, non-coding RNAs and signaling molecules [2,3,4].

1.1.2 ESC-specific Transcription factors

Transcription factors can interact with the chromatin through their DNA-binding domains that recognize specific DNA sequences (motifs) [5,6]. These proteins can induce the transcription of coding/non-coding genes while repressing the expression of others and are an important part of the regulatory circuitry. In hESCs, three core (master/key) transcription factors, OCT4 (Pou5f1), SOX2 and NANOG (collectively abbreviated as OSN) act in coherent circuits to maintain the pluripotent state [7]. Functional studies identified Oct4 and Nanog as master regulators by their unique expression patterns: enriched in the pluripotent state and reduced as ESCs undergo differentiation [8,9,10,11]. Oct4 and Sox2 form a heterodimer and bind to the DNA hence Sox2 is placed among the key regulators [12,13], although expression of Sox2 is also observed in some somatic cell types [14].

An “interconnected autoregulatory loop” emerged from genome-wide binding studies whereby the master regulators occupy their own promoters and reciprocally bind to the promoters of other key factors in order to regulate each other [15]. Oct4:Sox2 and Nanog also bind a major portion of coding/non-coding gene promoters along with several hundred intergenic regions, including enhancers for pluripotency related genes. Integration of global gene expression data with OSN binding sites revealed that these factors are involved in transcriptional regulation of both active and repressed genes [15,16]. The ability to affect either repression or activation by the same transcription factors may be

due to context-specific co-factors that are recruited along with these key factors. One subset of actively expressed targets in ESCs is genes that are essential to maintain pluripotency and self-renewal where Oct4, Sox2 and Nanog bind together with co-activators or activating chromatin regulators, e.g. components of the Trithorax Complex. Transcriptionally silent, OSN gene-targets are enriched in developmental and differentiation regulators, as well as several lineage-specific genes. In this case, Oct4, Sox and Nanog repress gene expression by facilitating the binding of chromatin modifiers such as SetDB1 or Polycomb complex proteins that mark the chromatin around the regulatory sequences of silenced genes with repressive histone marks [17].

Several other transcription factors have been shown to play important roles in the regulation of pluripotency, but not all of these are conserved between mouse and human ESCs. Sall4 and Tcf3 are shown to target most of the genes that are bound by the key factors [2,18,19,20]. Other transcription regulators including Smad1, Ronin, Klf4, PRDM14, Tbx3, Esrrb and Trim28 are also implicated in maintaining pluripotency and controlling ESC state [2,20,21,22].

1.1.3 Chromatin regulators

The eukaryotic genome is wrapped around highly conserved histone protein bundles (nucleosomes) to achieve compression of this long string of DNA into the nucleus, creating a higher-order DNA-protein complex called chromatin [23]. Nucleosome structure around a certain region has been shown

to affect the accessibility of underlying genomic elements (promoters, enhancers) thereby influences gene expression, DNA replication, DNA repair and others [24]. Several studies showed that certain sets of chromatin modifying enzymes contribute to the stability of pluripotency: whereas, others influence the establishment of conditions favorable to differentiation [25,26,27]. These chromatin regulators include histone-modifying enzymes, ATP-dependent nucleosome remodeling complexes, DNA (de)methylation complexes and higher order chromatin organizers, such as CTCF and cohesion [4,28].

1.1.3.1 Histone-modifying enzymes

Tails emanating from histone proteins in the nucleosomes are subjected to certain reversible post-translational modifications (PTMs) such as methylation, acetylation, phosphorylation and ubiquitination [29,30]. Combinations of the histone PTMs influence numerous molecular processes; therefore, complexes that are “reading”, “writing” and “erasing” certain modifications have significant roles in ESC biology [31,32,33].

One of the key features of ESCs is the presence of bivalent histone modifications at the regulatory sites of certain genes [34]. Genes encoding developmental and lineage-specific regulators are held in a “poised” state by bivalent histone modifications, defined as concomitant active histone mark histone H3 lysine 4 tri-methylation (H3K4me3) and repressive histone mark histone H3 lysine 27 tri-methylation (H3K27me3) on the same chromatin region.

These poised genes are silent in pluripotent cells but rapidly activated in response to signals that induce differentiation by changing the histone PTM status near promoters [35,36,37,38]. In general, Trithorax group (TrxG) proteins deposit H3K4me3 marks at promoters, and promote the transcription of active genes [39]. On the other hand, polycomb group (PcG) proteins catalyze deposition of H3K27me3 and, when present as a bivalent PTM prevent the transcription of developmental or key signaling genes in order to maintain a pluripotent state [40,41,42]. Several studies have shown that depletion of certain TrxG complex proteins or subunits of the PcG complex, such as PRC1 and PRC2, leads to defects in pluripotency maintenance and proper differentiation, supporting their importance in ESCs [43,44,45]. Genome-wide mapping comparisons revealed high co-localization of core pluripotency factors with PcG proteins [17,42]; moreover, Oct4 is reported to interact with components of TrxG and PcG complexes [46]. Taken together, these findings suggest an interconnection between core transcription factors and histone-modifying enzymes in order to maintain pluripotency. As ESCs differentiate into a certain lineage, specific developmental factors are induced by mechanisms that retain the active histone mark (H3K4me3) while removing the repressive histone mark (H3K27me3). In parallel, non-induced genes, such as regulators of cellular lineages that are not induced, tend to lose their “poising”, active histone modifications and acquire more H3K27me3 mark, which provides a mechanism for how bivalent domains help to establish ESC plasticity [38,47] (Figure1).

In addition to H3K27me3, histone H3 lysine 9 tri-methylation (H3K9me3) is another important repressive histone modification for ESCs [48]. SetDb1, G9a and Suv39h1 are involved in catalysis of H3K9me3, which has been shown to repress diverse developmental regulators in the pluripotent state [49,50,51]. Thus, various histone modifiers are involved in gene silencing of several developmental regulators in ESCs.

Histone acetyltransferases (HATs) are also implicated in the regulation of pluripotency and lineage-specific differentiation. The Tip60-p400 complex, which catalyzes histone H4 acetylation, also targets most of the Nanog binding sites and based on functional screens emerged as an ESC identifier [52]. Another HAT, p300, together with the presence of the histone H3 lysine 4 mono-methylation (H3K4me1) mark, has been associated with enhancer regions and co-localizes significantly with key transcription factors at promoter distal regions of genes in ESCs [53]. Similar to ESC promoters, enhancers may also exist in poised (marked with H3K27me3) or active states (marked with H3K27ac) [54,55]. Although the mechanism remains elusive, during differentiation poised enhancers are converted to active ones, a process that requires HAT enzyme activity to deposit acetylation on histone H3 lysine 27 in a lineage-specific manner, and consequently help to establish tissue-specific gene expression programs [56,57,58].

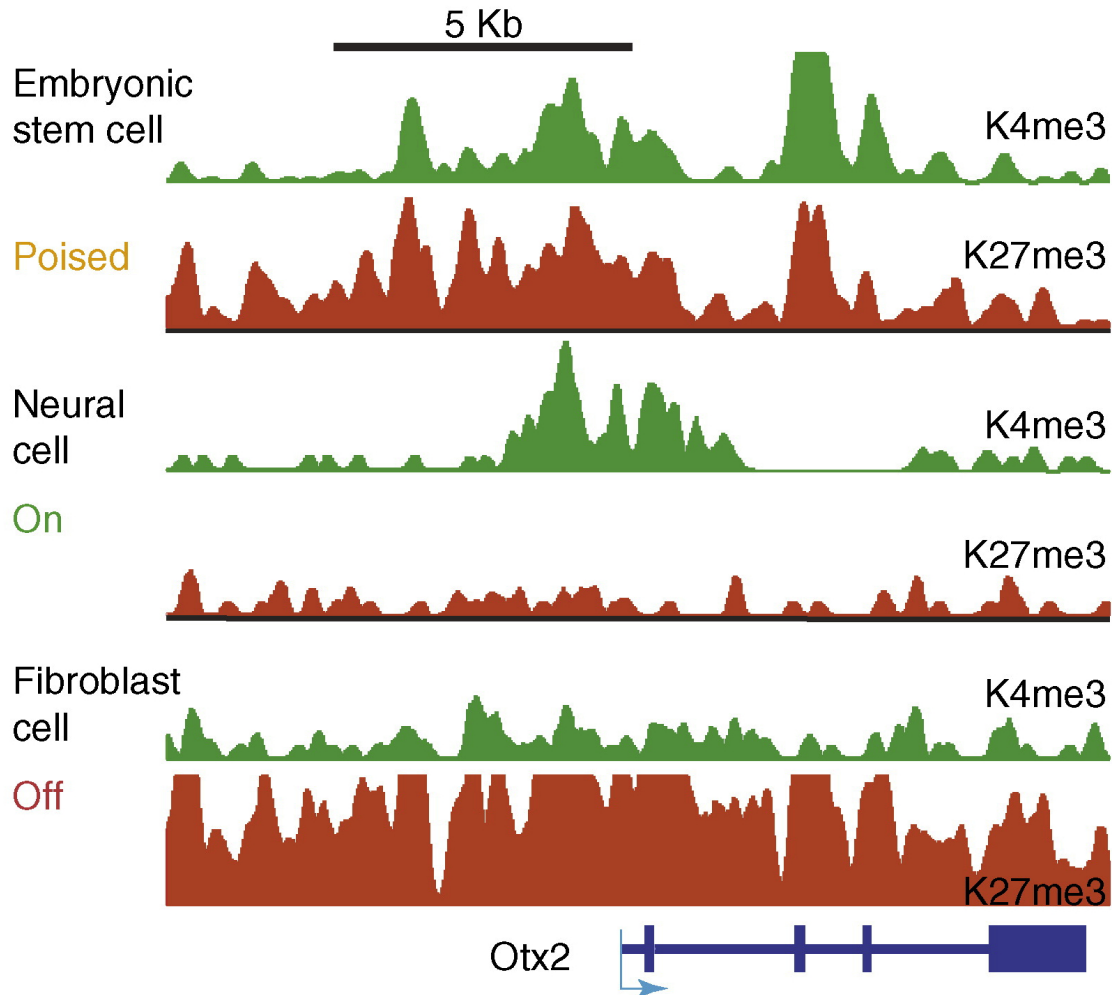


Figure 1. Bivalent chromatin domains help to establish embryonic stem cell plasticity. Reprinted by permission from Elsevier: Current Opinion in Genetics & Development, copyright (2008) [31].

Promoter of developmental transcription factor, Otx2 (neural-specific developmental factor) is marked by bivalent chromatin marks (H3K4me3 and H3K27me3) and transcriptionally poised in ES cells. In neural progenitor cells, Otx2 is transcribed and promoter is only associated with activating mark H4K3me3 while repressive histone mark H3K27me3 is selectively removed. In embryonic fibroblast cells, the expression of Otx2 is permanently repressed as a result of remaining H3K27me3 mark.

1.1.4 Non-coding RNAs

A number of genome-wide transcription studies inferred that the majority of the mammalian genome is transcribed, and many of these transcribed regions do not encode for a protein [59]. Subsequent studies revealed some of the biological functions of these pervasive non-coding transcripts [60]. Regulatory roles of non-coding RNAs (ncRNAs) in bio-molecular processes include repeated elements silencing, X-chromosome inactivation, polycomb repression and regulation of embryogenesis at different stages [61]. A diverse group of ncRNAs transcripts have been postulated to control, in part, the ESC state, including microRNAs (miRNAs) and large intergenic ncRNAs (lncRNAs) [62].

miRNAs are small ncRNAs (~22 nucleotide long) that are involved in post-translational mRNA silencing by base pairing to complementary sequences of their target RNAs in order to regulate a gene-expression program in cells [63,64]. Lack of miRNA biogenesis pathway components (Dicer and DGCR8) in mouse ESC results in defects in differentiation and decreased proliferation rates, which demonstrates the importance of this particular ncRNA family for the regulatory circuitry of pluripotent state [65,66]. Two key themes emerged from a study by Marson et al., which revealed how miRNAs integrate into that regulatory circuitry[67]:

- 1) Key transcription factors induce expression of miRNAs that are critical to fine-tune the mRNA levels of ESC-related genes that maintain pluripotency and those that facilitate the rapid degradation of ESC transcripts during differentiation and establish cell state transitions [68,69]. The cluster of mir-290-

295 constitutes a big portion of such miRNAs [70]. Members of this cluster contain seed sequences that can recognize mRNA of proliferation-related or epigenetic modulator genes; and, therefore, are involved in maintenance of pluripotency.

2) In the same fashion as lineage-dependent gene regulation, with the help of repressive chromatin regulators SetDB1 and PcG complexes, key transcription factors poise the expression of certain miRNA families. These miRNAs are up-regulated during lineage-programming and inhibit several key genes that are required to maintain pluripotency [71]. For example, human miR-145 can target and repress pluripotency specific genes, including OCT4, SOX2, and KLF4. OCT4 binds to up-stream regions of the miR-145 promoter and poises its expression in hESCs to establish an “irreversible positive feedback” loop that helps to control the balance between pluripotency and differentiation [72].

Discovered lncRNAs are defined as intergenic transcripts longer than 200 nucleotides in length with little potential for coding functional proteins and revealed by a specific chromatin signature: a combination of promoter-associated H3K4me3 and RNA Polymerase II (PolII) elongation mark histone H3 lysine 36 tri-methylation (H3K36me3) [73,74]. They can play important roles in numerous cellular processes, including participation in a pluripotency-differentiation balance with some lncRNAs favoring pluripotency and others differentiation [75,76]. An intriguing study by Guttman et al. revealed that, majority of lncRNA regulatory regions are bound by core transcription factors in ESCs [77]. This suggests that, like protein coding genes, lncRNAs are also

regulated by core pluripotency factors to maintain the ES cell state. Additionally, in the same study, the functional relevance of 226 lncRNAs were assessed by RNA interference experiments in mouse ESCs and supported a model where impairment of lncRNA expression influences proper ESC maintenance as well as differentiation. Intriguingly, RNA immunoprecipitation experiments indicated that ~75% of lincRNAs were bound to at least one chromatin regulatory complex, such as PcG and/or LSD1-histone demethylase proteins, substantiating the hypothesis that lncRNAs may function as modular scaffolds to bring different proteins or complexes together and reinforce the recruitment and stabilization of chromatin complexes during development and pluripotency [78,79,80].

1.1.5 Signaling mediators of the ESC state

Signal-transduction pathways are involved in regulation of various cellular processes, and perturbations in a signaling cascade may lead to severe abnormalities, including initiation or progression of cancer [81]. As a part of an effort towards deconstructing regulatory mechanisms of ESCs and development, numerous signaling pathways were scrutinized in detail and divided into intrinsic ones, which maintain an ESC state, and extrinsic signaling, which initiates lineage-specific differentiation [22,82,83].

1.1.5.1 Signaling pathways that maintain pluripotency

Extrinsic signaling pathways that impinge on pluripotency are distinct in human ESCs from those in mouse ESCs [84,85]. LIF and BMP pathways are

related to sustain mouse ESC state; whereas, transforming growth factor- β (TGF- β) signaling is one of the key pathways that maintain pluripotency in hESCs [86]. Activin and nodal proteins are members of the TGF- β family of ligands and suppress hESC differentiation, in part, by blocking BMP4 expression [87]. Additionally, Activin/nodal proteins can activate effector transcription factors (SMAD2/3), which in conjunction with an extracellular protein FGF2, up-regulate expression of core transcription factors NANOG and OCT4 to support hESC self-renewal [88,89,90]. Even though, WNT-mediated signaling has been implied in short-term pluripotency maintenance, the underlying mechanisms remain uncertain [91]. In summary, several extracellular signaling pathways play critical roles in the regulation and maintenance of ESC state.

1.1.5.2 Differentiation-related extrinsic signaling pathways

Pluripotent ES cell can give rise to three primary germ layers: endoderm (pancreas, lung, gut), ectoderm (nerve, skin) and mesoderm (muscle, blood), which are initiated by different extrinsic signaling pathways. Specific small molecules and receptor ligands either alone or in combination cocktails are used to differentiate ESCs into a specific lineage. In this study, we utilized Retinoic Acid (RA) signaling pathway as a model system to study early lineage-specific (neuro-ectoderm in particular) differentiation of hESCs.

1.1.5.2.1 Retinoic Acid signaling

Active metabolites of Vitamin A are collectively called retinoids, and they have been implicated in regulation of various biological processes [92]. For animals, dietary intake is the only source of retinoids since de novo synthesis mechanisms for these molecules do not exist. Several enzymes are involved in retinoid uptake regulation in mammalian systems. Retinoids are first converted into retinaldehyde by oxidization enzymes called alcohol dehydrogenases (ADHs). Retinaldehyde dehydrogenase (RALDHs) enzymes catalyze the second step (oxidization of retinaldehyde), from which Retinoic Acids (RAs) are produced [93]. RALDH2 is the sole enzyme responsible for embryonic uptake of RAs that, when deleted in mice, results in lethality, which signifies the importance of RAs during mammalian embryogenesis [94]. Given their significance in development, distribution patterns of RAs are strictly controlled by cytochrome P450 26 subfamily proteins that convert RA into less stable byproducts which are rapidly degraded in tissues that should not receive RA signaling [95].

Once transported inside the cell, RAs are shuttled to the nucleus with the help of specialized-proteins, such as CRABP2 (cellular RA-binding proteins). In the nucleus, RAs form a new complex by binding to retinoic acid receptors (RARs) and retinoid X receptors (RXRs), which when activated by RA-binding form heterodimers and bind to specific DNA motifs known as RA-response elements (RAREs). Following DNA binding of the RXR/RAR complex, a number of co-activators, e.g. NF1, together with ATP-dependent chromatin remodeling

complexes are recruited to RAREs in order to facilitate transcription of lineage-specific RA-responsive genes [96].

Early studies demonstrated that RA-treated ESCs undergo neuro-ectodermal lineage differentiation, which leads to the formation of neural progenitor cells [97]. Numerous RA target genes have been identified so far, including developmental transcription factor HoxA1, suggesting that activation of RA signaling drives ESCs towards neural-lineage development by inducing expression of a particular set of lineage-specific developmental factors [93,98].

1.1.6 Induced Pluripotent Stem Cells

A groundbreaking experiment in 2006 performed by Yamanaka's lab – for which Dr. Yamanaka eventually was awarded the 2012 Nobel Prize in Physiology or Medicine - demonstrated that retroviral-mediated transfer of four transcription factors (Oct4, Sox2, Klf4 and c-Myc) can reprogram differentiated mouse embryonic fibroblasts to an ESC-like state, known as induced pluripotent stem cells (iPSCs) [99]. Successive studies showed that similar reprogramming could be achieved by the transduction of the same or a modified set, e.g. Lin28 as substitute for Klf4 and c-Myc as a dispensable factor, of transcription factors in human differentiated cells [100,101,102]. Similarly, some ncRNAs, such as lincRNA-regulator of programming (linc-RoR) [103], or miRNAs, miR-294 and miR-295 [104], can also be used to enhance reprogramming efficiencies.

Notably, *in vivo* studies elucidated the striking morphological and biological similarities between ESCs and iPSCs, including the most stringent tests of pluripotency: differentiation into multiple germ layers and formation of teratomas [99]. Comparison of the genome-wide binding of core transcriptional factors demonstrated that localization of these factors significantly overlaps between hESCs and hiPSCs, except at some heterochromatin regions marked by H3K9me3 (named as OSKM-DBRs) [105,106]. Although some studies indicate that reprogramming fails to completely erase the epigenetic memory of the cell of origin [106], limited but consistent genome-wide transcriptional and chromatin-based variations, mainly bivalent modifications, H3K4me3 and H3K27me3, are observed between hESCs and hiPSCs [107] (Figure2). Taken

together, shared similarities by ES and iPS cells increase the hopes that human iPS cells could one day be used as therapeutic agents in immune-matched patient-specific regenerative medicine practices [108].

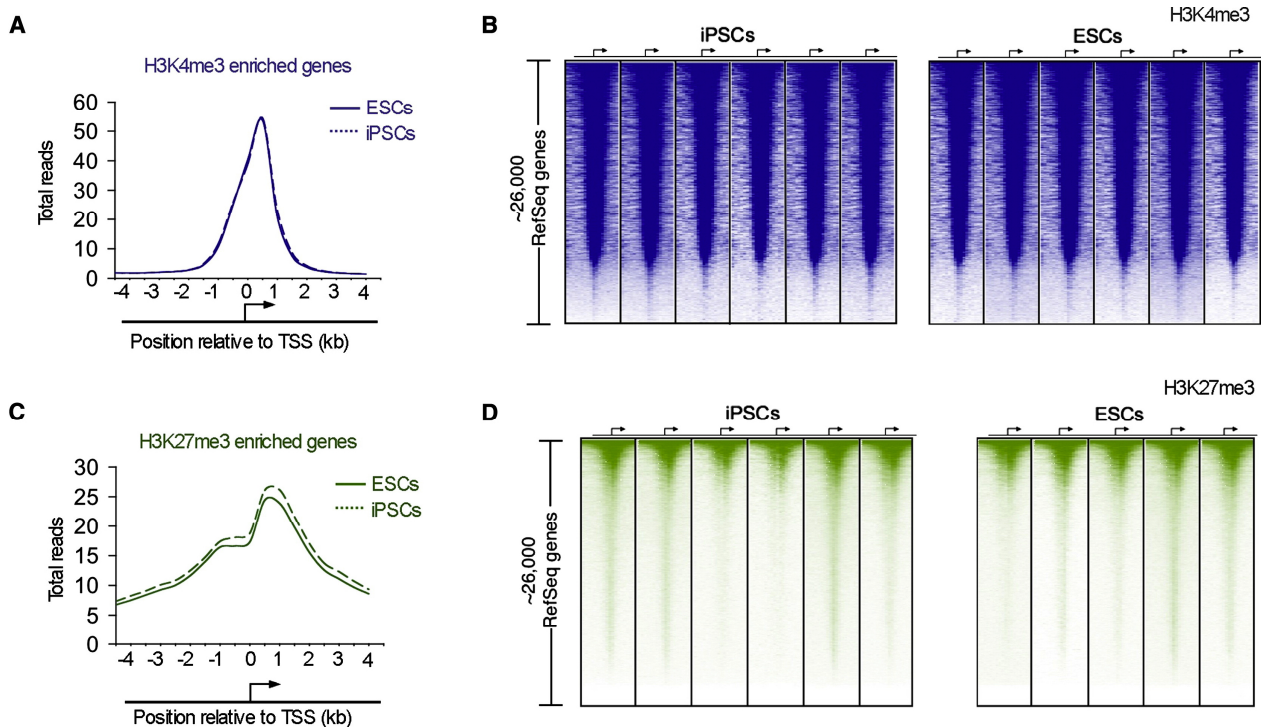


Figure 2. Genome-wide bivalent chromatin modification maps show significant similarities between human iPS and ES cells. Reprinted by permission from Elsevier: Cell Stem Cell, copyright (2010) [107].

A. Aggregate plot show H3K4me3 enrichment profile for all RefSeq genes in ES cells (solid blue) and iPS cells (dashed blue). The arrow indicates transcription start site (TSS) and direction of transcription of the average.

B. Heatmap depicts the density of H3K4me3 mark (blue) around all Refseq genes promoters – genomic region from -4.5kb to +4.5kb relative to the TSS is shown. Gene order was determined by highest average ChIP-Seq density in ES cells and arranged from highest to lowest density.

C. Aggregate plot show H3K4me3 enrichment profile for all RefSeq genes in ES cells (solid blue) and iPS cells (dashed blue).

D. Heatmap depicts the density of H3K27me3 mark (green) around all Refseq genes promoters – genomic region from -4.5kb to +4.5kb relative to the TSS is shown.

1.1.7 ESC-specific gene expression signatures in human cancer

Cancer cells exhibit molecular and biological traits that resemble some hallmarks of stem cells, including high proliferation rate, self-renewal and even lack of differentiation since some aggressive tumors are present in an undifferentiated state [109]. Recent studies showed that ES cell-like gene expression signatures are shared among different human cancers, which could account for some of the reported similarities between cancer and ES cells [110]. One of the earliest studies that compared the underlying gene expression programs of ESCs and epithelial cancer cells revealed an evolutionary conserved (between mouse and man) ESC-like transcriptional signature, which is activated in various human epithelial cancers yet suppressed in normal cells [111]. Furthermore, Weinberg and colleagues have shown that poorly differentiated human tumors exhibit transcription of ES-cell-specific genes along with repression of PcG complex (PRC2, Eed and Suz12) target genes [112] (Figure3). In contrast, a more recent study argued that recapitulated ESC-like gene expression signatures in cancers are mainly due to activation of pro-proliferation factor c-Myc in human tumors but not the core transcription factors [113]. Although, the idea is compelling, since c-Myc locus amplification is one of the most frequent copy-number alterations in human cancers [114], it is unclear how c-Myc can be solely responsible for the activation of a core ESC program during tumor initiation, considering that c-Myc is not strictly required for iPS cell generation or reprogramming [102]. Overall, accumulated evidence indicates that an ES cell-like gene expression program is positively correlated with poorly differentiated tumors (histologically graded), increased risk of metastasis and

decreased survival rate in human patients.

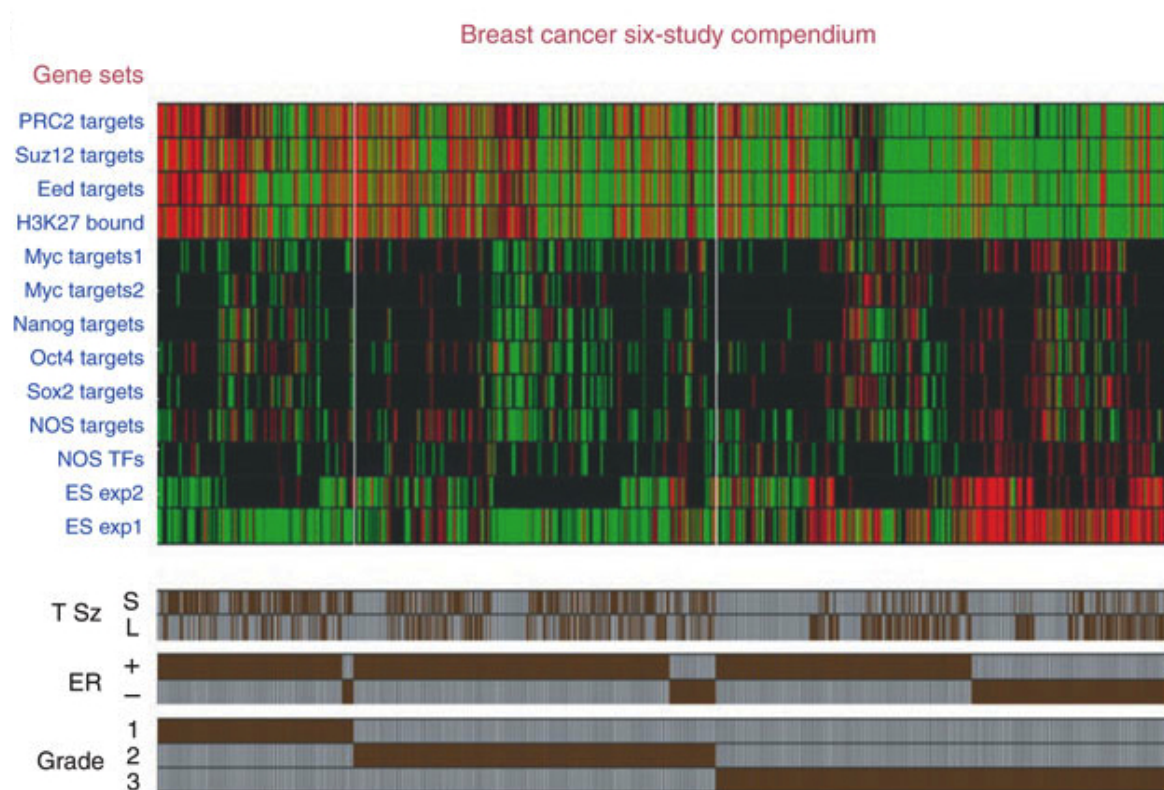


Figure 3. High-grade human breast cancers display ES-cell-specific gene expression signature. Reprinted by permission from Macmillan Publishers Ltd: Nature Genetics, copyright (2008) [112].

1,211 breast cancer samples have been investigated (columns). Red/green colors indicate significantly over- or under expressed gene sets, respectively. Bottom bars (brown) indicate individual tumor annotations - where available - for ER status (positive or negative), grade (1,2 or 3), and tumor size (S – tumor smaller than 2cm, L – tumor larger than 2cm).

1.2 p53 and Pluripotency

Transcription factor p53 drives expression of an array of target genes in a cellular-context and stress-stimuli specific manner. p53's function as a tumor suppressor has long been recognized, hence it is aptly named as the “guardian of the genome”. It functions as a tumor suppressor by promoting apoptosis and regulating cell proliferation, primarily by cell-cycle arrest, in response to various stress signals, such as oncogenic activation, tumor-suppressor gene inactivation, genotoxic damage exposure and loss of normal cell-cell contacts. Thus, p53 prevents an accumulation of genomic instability, which is one of the major causes of cancer formation [115]. However, p53's contribution to numerous other cellular processes has only recently been appreciated, including its functions in development and differentiation [116].

1.2.1 p53 acts as barrier to somatic cell reprogramming

The seminal study by Takahashi and Yamanaka on nuclear reprogramming offers great possibilities for regenerative medicine, as generation of patient-specific iPS cells becomes feasible, in addition to the ability to study mechanisms of development and disease in these cell systems. However, the shortcomings of the original method, namely, an inefficient reprogramming rate (1-3%) and slow kinetics of iPSC generation (as long as several weeks), are major drawbacks to the clinical use of reprogrammed cells. These challenges led researchers to consider whether proteins that acted as barriers and limited somatic cell reprogramming were expressed in

differentiated cells.

Notably, five simultaneous reports showed, by various experimental approaches, that depleting p53 or inhibiting p53-dependent pathways to disrupt p53 functions dramatically increases the reprogramming rate (as much as 80%) and accelerates the kinetics (as early as 3 to 5 days) of iPSCs generation [117,118,119,120,121]. Although the obtained results were exciting and encouraging, several concerns have arisen regarding inhibition of a crucial tumor suppressor during reprogramming [122,123]. One of the previously mentioned five studies, Hong et al., observed that mice generated by partially using p53-deficient iPS cells were viable but these mice eventually developed tumors [117]. In addition, Marion et al. reported increased genome instability and abnormal telomere shortening in p53-deficient iPS cells [120]. Although, the use of oncogenic reprogramming factors, such as c-Myc and Klf4 or retroviral-mediated infections may be potential explanations for the induction of p53 and its activity as a barrier to reprogramming, less oncogenic reprogramming techniques, which exclude oncogenic factors from the reprogramming cocktail or using different transfection methods, still lead to a p53-mediated cell-cycle arrest of a majority of cells during reprogramming. This suggests that p53's function during creation of iPS cells could extend beyond its responsibility to safeguard the genomic integrity during oncogenic stress [124].

1.2.2 p53-inactivated cancers display plasticity and loss of differentiation

Although cancerous cells exhibit striking differences between individuals

or due to the tissue origin of the disease, most of them share one general deficiency: p53 loss-of-function, which underscores the importance of p53 in maintaining cellular integrity. Given p53's prominent role to restrain cellular reprogramming and the gene expression signatures shared between cancers and ES cells, it is reasonable to ask whether there is a positive correlation between p53 inactivation and acquisition of a stem-like state.

In two separate studies, Levine and associates surveyed global gene expression in metastatic breast and lung cancers [125], or prostate tumors [126], and demonstrated that cancers that are mutant for p53 function exhibit an ESC-like expression program that correlated with worse overall survival rates for patients. A similar association was previously observed at a molecular level in poorly differentiated thyroid cancers [127], in lung cancers [128] or in acute myeloid leukemia progenitors [129]. Consistent with these findings, it has also been shown that expression of p53 induces differentiation of leukemia-derived cells K562 cells [130].

Taken together, a better understanding regarding the pathways that drive dedifferentiation in p53-inactivated cells or the precise mechanism of how p53 can function to favor differentiation is required to enhance efficiency of iPS cell production without jeopardizing genomic stability of those cells. Additionally, a better understanding of how tumor cells acquire cellular plasticity after p53-inactivation may lead to development of more potent and targeted therapeutic treatments.

1.2.3 p53's function in human ES cell differentiation

Tumor suppressor p53 has been implicated in limiting the self-renewal of stem cells, specifically in mouse ES cells by suppressing core pluripotency factor Nanog [131] or by activating developmental Wnt-signaling [132]. These findings led to the hypothesis that p53 imposes differentiation of mouse ESC as a tumor-suppressive mechanism in response to DNA damage [133]. In addition to being regulated by distinct extrinsic signaling pathways, multiple studies suggest there are fundamental differences between mouse and human ES cells at the basic mechanisms of transcription factor function. As an example, even the core transcription factor binding sites show significant differences: only 5% of the most enriched OCT4 and NANOG binding sites in hESC are present at homologous regions in mice [134,135]. Additionally, hESCs contain one inactivated X-chromosome thereby present in a “primed” state for differentiation, while mESCs are in a more primitive, “naïve” state, which maintain two active X-chromosomes [136,137]. Further understanding of the earliest stages of human embryonic development is needed to resolve such controversies [138].

Unlike differentiation in mouse ES cells, p53-dependent cell cycle arrest is observed in human ES cells in response to DNA-damage [139], which suggests that different stress-specific functions of p53 exist between mouse and man. Recent work from our laboratory revealed that p53 plays a significant role during retinoic acid-mediated differentiation of human ESCs. Depletion of p53 results in inefficiencies during differentiation since the majority of the cells maintain higher levels of OCT4 and NANOG expression even after several

days of RA treatment. This suggests that p53 is an important factor for efficient differentiation of hESCs [140]. Human ESCs stably expressing wild-type p53 under TET-inducible promoter underwent differentiation even in absence of Retinoic Acid. However, the same effect was not observed when a mutated form of p53, p53R175H, which is incapable of binding to DNA, is ectopically expressed. This suggests that p53 promotes hESC differentiation by binding to DNA and functioning as a transcription factor to activate or repress target gene expression.

Further analyses revealed that, in response to RA, p53 is enriched at the promoter of one of the key p53-effector genes, p21 or *CDKN1A* and induces its expression. This is significant since higher levels of p21 results in the accumulation of hESCs in G1-phase of cell cycle, which promotes differentiation. These actions of p53 in hESCs are in complete contrast to its roles in mouse ESC differentiation, where it represses Nanog expression by directly binding to its promoter [131]. Lengthening of the hES cell cycle and impeding rapid cell divisions not only limit self-renewal but also facilitate the programs that induce differentiation [141]. Additionally, p53 also activates expression of specific micro-RNAs, miR-145 and miR-34a, which repress expression levels of core pluripotency factors OCT4, SOX2 and KLF4 and thus prevent partially differentiated hESCs from backsliding to pluripotency.

1.3 Genome-wide protein-chromatin interaction studies

Cell fate and development are established through an intricate network that regulates gene expression programs in a certain tissue at a given time.

Understanding the nature of DNA-protein interactions and epigenetic modifications is crucial for deciphering the codes of the underlying gene regulatory networks [5]. Several approaches have been devised to identify genome-wide locations of transcription factor binding and histone modifications [142]. Chromatin immunoprecipitation (ChIP) is a powerful method to purify DNA fragments that are associated with a particular transcription factor (TF) or a post-transcriptionally modified histone. Initial high-throughput screens used ChIP and predesigned microarrays, a method known as ChIP-chip, by hybridizing fluorescent-labeled, ChIP-antibody precipitated fragments of DNA to homologous oligomers of DNA fixed to substrates [143]. Although, whole genome tiling arrays can be used to screen entire genome in a ChIP-chip study, this method requires several chips per condition, therefore is infeasible and not cost-effective for mammalian genome studies [144].

Advancements in next-generation sequencing technology, where the antibody-bound chromatin fragments obtained from a ChIP experiment are directly subjected to deep sequencing of DNA, made identification of DNA-protein interactions more comprehensive [145]. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is advantageous over the ChIP-chip method in several ways: it provides better resolution and unbiased genome coverage, obtained results contain fewer artifacts and it requires smaller amounts of starting material [146,147]. Numerous computational tools have been developed to pinpoint the precise location of a protein of interest's binding site within the genome of the studied organism and to annotate or compare the obtained data for downstream analyses [147].

Some common steps of ChIP-Seq data analysis pipeline can be listed as:

- Read mapping – As a first step, obtained sequenced ChIP fragments (tags) are aligned onto the genome with the help of any available short-read mappers (i.e. Bowtie, BWA or Illumina’s ELAND software).
- Identification of significantly enriched regions (Peak calling) – Once alignment is done, the next step is to identify genomic sites where the obtained reads are enriched significantly higher than expected by chance. Although ChIP-seq offers less technical artifacts, it is still subject to some inherent biases due to the experimental protocols (antibody specificity), sequencing technology (non-specific noise) or the genomic structure (regional GC bias, open chromatin regions tend to precipitate more easily). Thus generating input control data is a vital step for augmenting this identification step.
- Down-stream analysis – Several subsequent analyses can be performed based on the purpose of the study, such as identifying location of the enriched regions on the genome relative to any known genomic features, motif discovery or incorporating gene expression data to identify potential function of studied transcription factor.

1.4 Hypothesis, specific aims and rationale

My **hypothesis** is that p53 regulates transcription of a signal-specific subset of genomic targets in hESCs that a) oppose pluripotency and b) protect the stem cell genome in response to differentiation and DNA damage.

Specific Aims

I tested this hypothesis by the following specific aims:

Specific Aim 1) To characterize p53's genome-wide binding profiles in DNA-damaged induced hESCs.

Specific Aim 2) To characterize p53's genome-wide binding profiles in differentiating hESCs and their potential functions.

Specific Aim 3) To compare p53-enriched sites with ES cell landmark signatures.

Rationale: p53 protein levels are elevated to comparable levels in DNA damage-induced hESCs and differentiation-initiated hESCs. Although similar abundance of p53 is observed under these conditions, cellular outcomes are strikingly different where DNA damage causes cells to arrest or undergo apoptosis and RA induces cells to differentiate and change their molecular signature. Our previous data showed that p53's DNA-binding ability is essential for its role of promoting hESCs differentiation. Thereby, p53's binding preferences could be the dictating factor for the different readouts and identification of those p53 binding sites may reveal which subset of target genes are responsible for each specific response.

CHAPTER 2 MATERIAL AND METHODS

2.1 ChIP-Seq Analysis

2.1.1 Sequencing and read alignment

Sequencing of p53-bound DNA was performed at the Bioinformatics Core of the Cincinnati Children's Hospital Medical Center, Cincinnati, OH. p53-bound DNA (~10 ng) was purified by PAGE to obtain 100–300 bp fragments and sequenced on an Illumina Solexa GAII sequencer. Sequencing of chromatin marks H3K4me3 and H3K27me3 ChIP DNA was performed at the MD Anderson DNA Analysis Facility. DNA associated with modified histones (~10 ng) was purified by PAGE to obtain 100–300 bp fragments and sequenced on an Illumina HiSeq2000 sequencer. Sequence reads (36 base pair long) derived from Illumina sequencers were aligned to the NCBI Build 36 (UCSC hg18) human genome using ELAND software (Illumina) to produce uniquely matched reads with up to two mismatches per read allowance.

2.1.2 Peak calling

Enriched regions for each condition were normalized to input DNA and detected by MACS version 1.4.0 (Model based analysis of ChIP, <http://liulab.dfci.harvard.edu/MACS/>) [148] with a p-value threshold of enrichment of $P < 1.00 \text{ E-}8$ for damage and differentiation datasets; however, a higher cut-off was used for untreated dataset because of the low throughput and high signal-to-noise ratio in this experiment - $P < 1.00 \text{ E-}10$. Non-default shift and bandwidth sizes were used for each dataset based on average

precipitated DNA fragments length in each case. Wiggle files (<http://genome.ucsc.edu/goldenPath/help/wiggle.html>) were generated using the same sequence files and density of reads per base pair was calculated in a 25bp window and later normalized to 10 million reads per sample.

Peaks share at least one base under their enriched regions called as overlapped between different conditions. BEDTools functions (`intersectBed` or `windowBed`) were used to perform overlapping sites analyses (<http://code.google.com/p/bedtools/>) [149].

The distance between unique peaks in each condition was measured using a gradually increasing window and determining the unique peaks summits coinciding in the same window. Obtained numbers were plotted and pie charts were generated by ratios of overlapping versus non-overlapping summits for a certain window length.

2.1.3 Conservation of binding sites

PhastCons conservation scores for 44 vertebrate species were downloaded from UCSC website (which contains base-by-base conservation scores based on a statistical model called phylogenetic hidden Markov model [150]) and individual chromosome files were merged into a single wiggle file (<http://hgdownload.cse.ucsc.edu/goldenPath/hg18/phastCons44way/vertebrate/>). Aggregate plots for conservation scores across (-3kb to +3kb) enriched sites were generated using the Sitepro version 0.6.6 program under CEAS (<http://liulab.dfci.harvard.edu/CEAS/>) [151] with 100bp resolution.

2.1.4 Motif analysis

Both de-novo motif discovery and known motif matching were performed using the MEME software suit. The sequences of the p53-peak regions were extracted in FASTA format and used as input for the MEME-ChIP pipeline, which is specifically designed to discover associated motifs in large sets of DNA sequences (http://meme.nbcr.net/meme4_6_1/memechip-intro.html) [152]. The pipeline runs MEME (good for long motifs) [153], DREME (good for short motifs) [154] for over-represented DNA-sequences in input, and AME (Analysis of motif enrichment) to search and compare the motifs that are discovered by MEME and DREME in the existing motif databases [155]. Briefly, zero or one motif per sequence was searched with the motif lengths between 6-30 base pairs, around 600bp of the peak summits and outputs for each dataset are shown with a p-value cut-off less than 1.00 E-10.

SeqPos motif discovery program in Cistrome analysis pipeline (<http://cistrome.org/ap/>) [156] was also performed for motif discovery underneath enriched sites (around 400bp of the peak's center) in each condition by using cistrome's curated motif database.

2.1.5 Identifying target genes of p53-bound sites

Human RefSeq gene information was obtained from UCSC table browser for human genome hg18 assembly (<http://genome.ucsc.edu/cgi-bin/hgTables?command=start>) [157]. Fold enrichment analysis over the randomized binding sites was performed as previously described [158]. Genes

with a nearby p53 peak 10Kb up/down-stream of transcription start sites were designated as targets.

2.1.6 Annotation of p53-target genes

Gene Ontology analyses for each set of target genes were performed using DAVID (<http://david.abcc.ncifcrf.gov/>) [159]. Developmental transcription factors were obtained from the HUGO Gene Nomenclature Committee at the European Bioinformatics Institute's website (<http://www.genenames.org/genefamily.html>) [160], previously published study annotations (Supplementary table S11 in Lee et al. <http://www.sciencedirect.com/science/article/pii/S0092867406003849#mmc12>) [17] and NCBI's Gene database (<http://www.ncbi.nlm.nih.gov/gene>). Each dot shown represents a member of a particular family only if the gene's ontology terms (GO - Biological Process and Molecular Function) entail transcription or DNA binding and also development or differentiation. Gephi (<http://gephi.org/>) graphic visualization software was used to generate network graph.

INTERPRO protein domain analysis was performed using Genomic Regions Enrichment of Annotations Tool or GREAT (great.stanford.edu) [161]. Peak files (Differentiation-specific, Damage-specific and conserved p53 bindings sites) were imported into GREAT by setting a gene association rule as a single gene within 10 kb ranges of binding sites. The top five categories by binomial p-value scores are shown.

2.1.7 Integration core ES cell transcription factor binding data

ChIP-Seq datasets of OCT4 (GSM518373) and NANOG (GSM518374) were obtained from GEO database (<http://www.ncbi.nlm.nih.gov/geo/>) [162]. Raw sequences were re-analyzed with MACS version 1.4.0. Obtained peaks were used for overlap analysis and circular plot. Circos (<http://circos.ca/>) [163] was used to visualize p53, OCT4, NANOG and H3K27me3 around four HOX clusters. H3K27me3 ChIP-Seq data was obtained from UCSC genome browsers' ENCODE project website (<http://hgdownload.cse.ucsc.edu/goldenPath/hg18/encodeDCC/wgEncodeBroadChIPSeq/>) [164].

Wiggle files were generated by using the obtained sequence files and density of reads per base pair was calculated in a 25bp window and later normalized to 10 million reads per sample and used for aggregate plots which were generated by using Sitepro program in CEAS toolkit. Normalized wiggle files were used to generate a density plot, using the heatmap tool in the Cistrome analysis pipeline (<http://cistrome.org/ap/>) [165]. K-means clustering (5 cluster) was applied to the intensity signals of p53-Damage, p53-Differentiation, OCT4 and NANOG that were extracted around (-500 to +500bp) the p53-condition-specific genomic regions.

Peaks share at least one base under their enriched regions called as overlapped between different datasets (OCT4, NANOG, p53-Damage, p53-Differentiation). BEDTools functions (intersectBed or windowBed) were used to perform overlapping sites analyses (<http://code.google.com/p/bedtools/>).

In order to test if observed differences in the association of OCT4 and

NANOG with p53-Differentiation is significant, randomized binding sites showing similar distribution in each chromosome were generated 10000 times and used for determining statistical significance.

2.1.8 Bivalent histone modification analysis

Normalized wiggle files were used to generate histone aggregate plots. Transcription start site (TSS) of p53 target genes (up or down-regulated based on microarray data results) was used as the center of the window and each window was divided into 40 bins of 25bp resolution. Average ratios were plotted for each category.

2.2 Gene Expression Analysis

Affymetrix U133 Plus2.0 microarrays were performed for each condition (Pluripotent, +Adr and +RA) in triplicates. Robust multi-array average (RMA) method was used with default options (with background correction, quantile normalization, and log transformation) to normalize raw data from batches using R/Bioconductor's `affy` package (<http://www.bioconductor.org/>) [166]. EntrezGene IDs were assigned to the probe-sets using Affymetrix annotation package (`hgu133plus2.db`) in Bioconductor. For genes, which are represented by multiple probes on the array, maximum expression value was retained for further analyses. A gene is called as differentially expressed if FDR corrected p-value is less than 0.05, which is calculated with empirical Bayes method by `eBayes` function in Bioconductor's `limma` package [167]. Gene Ontology

analysis of differentially expressed gene was performed using DAVID (<http://david.abcc.ncifcrf.gov/>). Volcano plot is generated by using R's plot function, whereas the bar plots were generated by using ggplot2 (<http://ggplot2.org/>) package.

CHAPTER 3 RESULTS

3.1 Genome wide mapping of p53 in hESCs reveal distinct functional binding sites

We mapped p53 occupancy throughout the genome using ChIP-Seq method by deep sequencing of p53-bound chromatin fragments isolated from hESCs in a pluripotent state (untreated), undergoing differentiation (+RA) or after DNA damage (+Adr) in order to determine the molecular basis for these signal-specific responses and define a landscape of p53-chromatin interactions in hESCs. In pluripotent hESCs, p53 is enriched at 4509 genomic sites, compared to 8282 and 4941 in hESCs undergoing differentiation or damage, respectively (Figure 4). We found that p53 is enriched at distinct loci during each of these different conditions, since intersection of obtained enriched sites demonstrated that only a fraction of p53-bound peaks (26.5%) overlapped in between differentiation and damage induction (Figure 4). Comparison of unique sites in a gradually increasing genomic window revealed that only 44% of unique sites in differentiation and damage overlapped in a 100kb window, suggesting highly diverse p53 functions in these two states (Figure 5).

We investigated the evolutionary importance of identified p53-binding sites by profiling PhastCons score around those sites. Comparing genomic regions within 4kb of each p53-peak summit in 44 vertebrate species, revealed high evolutionary conservation of p53 binding regions suggesting potential regulatory functions of obtained genomic regions in each condition (Figure 6).

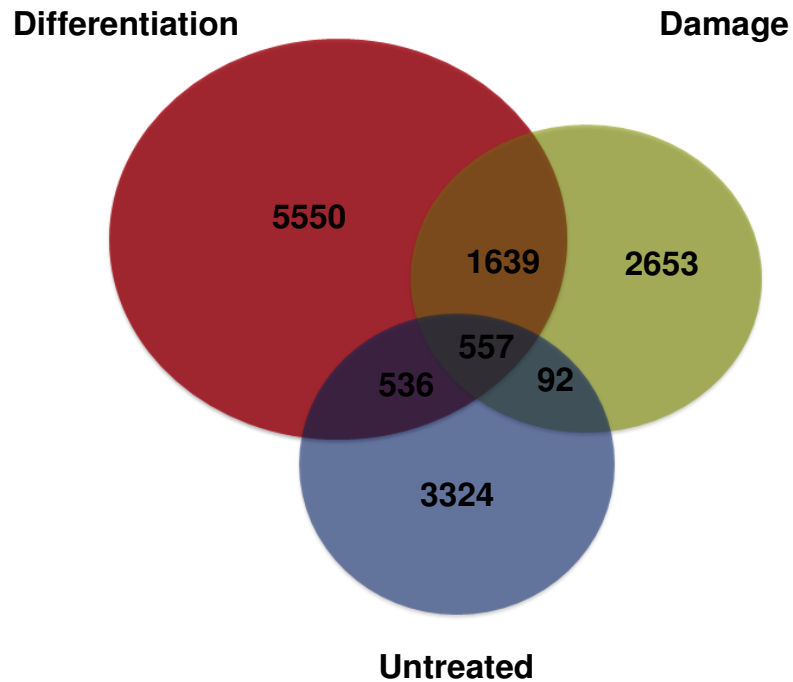


Figure 4. Genome-wide mapping demonstrated unique p53 signatures in hESCs after different treatments

Comparison of genome occupancy of p53 in untreated, differentiation (RA 2days) and damage (Adriamycin: Adr 6h) induced hESCs. p53 binding sites identified by peak calling program MACS with p-value 10^{-8} .

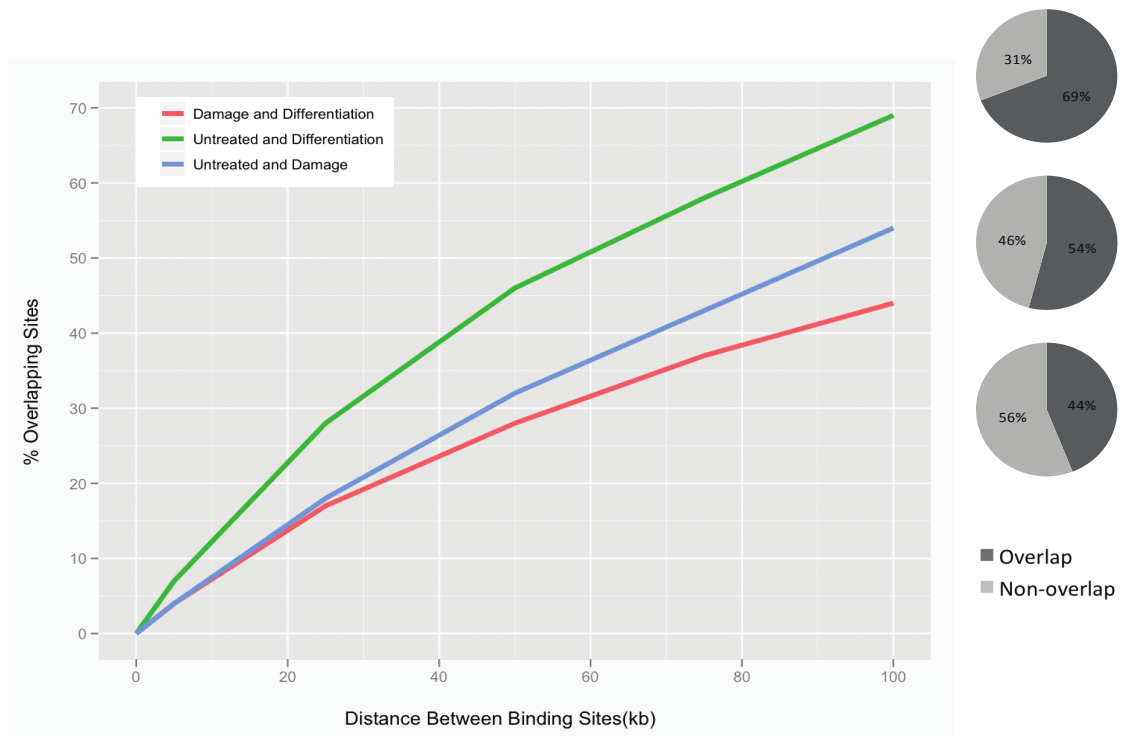


Figure 5. Condition specific binding sites of p53 are strikingly distant

Frequency of overlap between unique sites is shown as a function of distance between binding sites. Pie charts show percent overlap between unique sites in 100kb distance. Poor overlap of unique sites in differentiation and damage was observed (44%) even in a 100kb window.

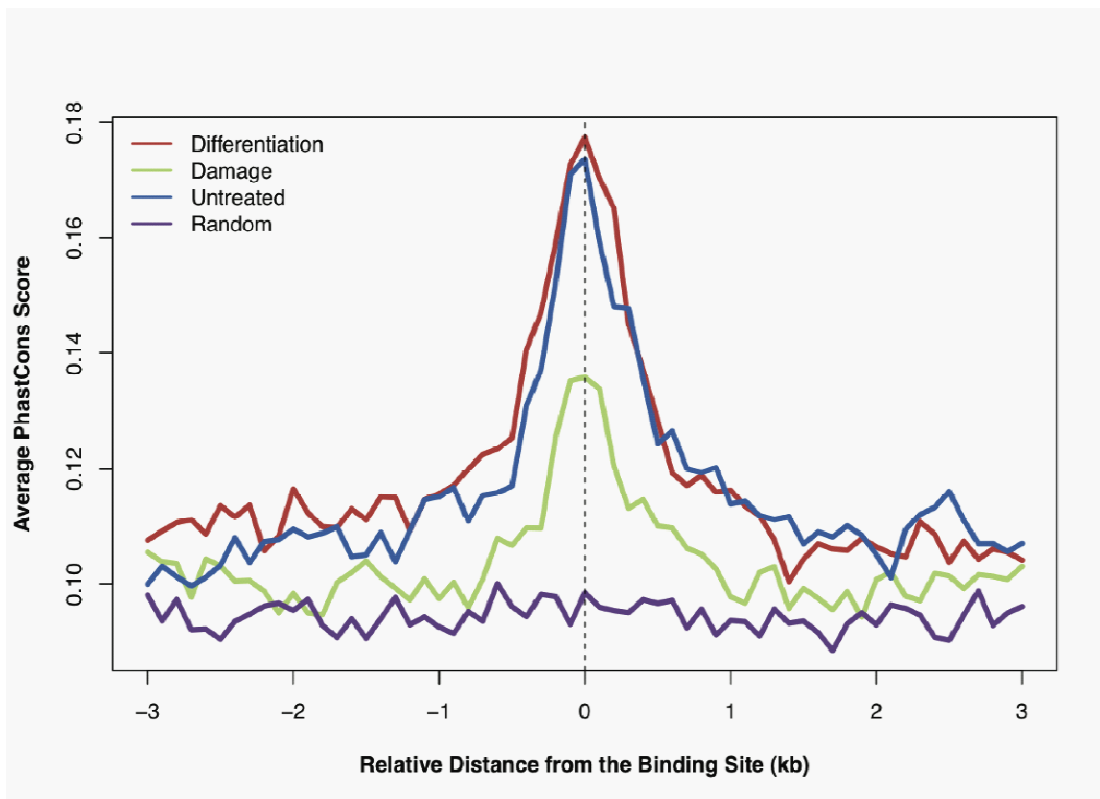


Figure 6. p53 binding regions are evolutionary conserved among vertebrates.

Average PhastCons score profiles depicting conservation in the vicinity of p53 binding sites and randomly generated genomic loci (purple).

3.2 p53 binding sites are enriched for p53 and OCT4:SOX motifs in differentiation

Motif analysis revealed that p53-bound regions were significantly enriched with consensus p53 binding sites (p53-motif) in both differentiation and damage ($P < 10^{-35}$ and $P < 10^{-235}$, respectively), a motif that is similar to the p53 consensus obtained from TRANSFAC database (Figure 7A). However, sequences bound by p53 in pluripotent hESCs (untreated) did not match the consensus p53-motif significantly ($P > 10^{-5}$), suggesting signals that activate p53 in hESCs stabilize p53-chromatin interactions, as a result precipitating precise p53-bound regions is challenging and yielding an ambiguous signal across the genome. These results support proposed models of p53 scanning along DNA, prior to inductive signaling, in a gene-specific manner that determines downstream response [168].

Intriguingly, p53-bound regions in hESCs undergoing differentiation were significantly enriched in core transcription factors OCT4 and SOX2 binding motifs ($P < 10^{-16}$ and $P < 10^{-12}$ respectively) (Figures 7A-B), whereas no OCT4-SOX2 motifs were found in p53-bound genomic regions from pluripotent hESCs or those exposed to damage (Figures 7A-B). We performed a reciprocal analysis to detect any p53-motif within OCT4-SOX2 and NANOG enriched sites, using previously published ChIP-Seq datasets [134]. Our analysis revealed overlapping p53 response elements (p53REs) in both OCT4-SOX2 and NANOG datasets (Figure 8). The presence of consensus binding motifs for OCT4 and SOX2 in p53-bound regions suggests a possible interplay between these transcription factors in determination of specific stem cell states.

A

Differentiation	
Motif Name	Corrected P-value
p53	4.8 E-37
Pou5f1(OCT4)	5.9 E-16
SOX2	4.6 E-12

Damage	
Motif Name	Corrected P-value
p53	9.3 E-235

Untreated	
Motif Name	Corrected P-value
p53	1.19 E-5

B

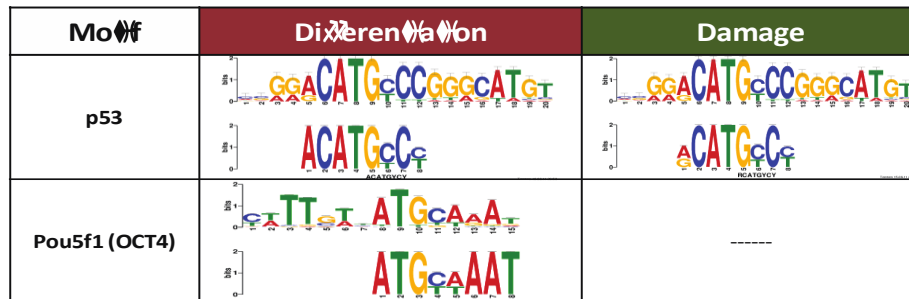


Figure 7. p53, OCT4-SOX2 motifs are enriched within p53 enriched sites

A) p53 and OCT4 consensus motif sequence from TRANSFAC database [top], and matching enriched motifs under p53 peaks [bottom].

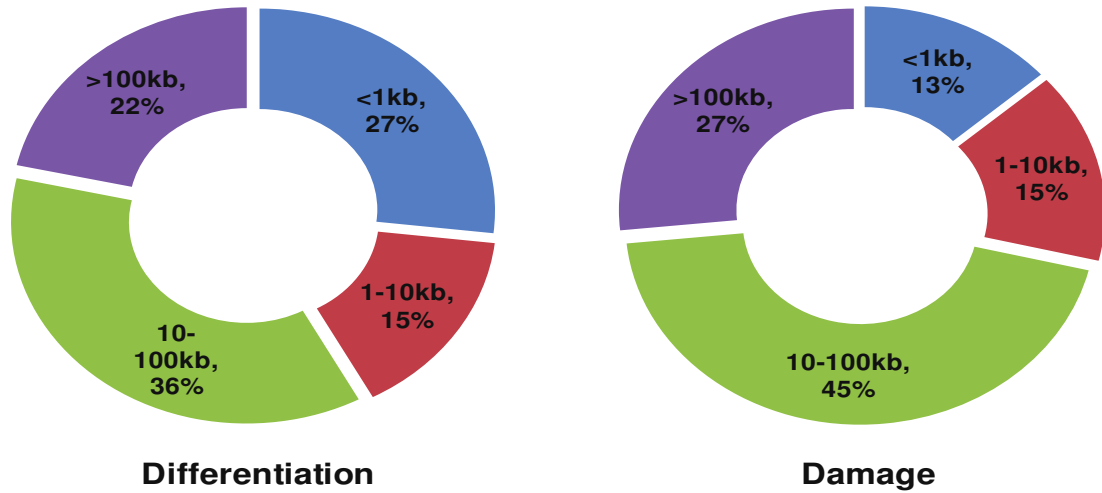
B) The OCT4 motif is enriched in p53-bound regions in cells undergoing differentiation, but not in response to damage.

3.3 p53 targets developmental transcription factors during differentiation

Across the genome, a significant portion of p53 binding sites (42% for +RA and 28% for +Adr) are enriched (0.68 fold for +RA and 0.61 fold for +Adr over randomized binding sites) within 10kb of the nearest annotated transcription start site (TSS) (Figure 9A). Therefore, we used a 10kb window of distance from the p53-peak summit to the nearest gene TSS to call a p53 target gene (Figure 9B).

Similar to the identified binding sites between conditions, target-gene comparison analysis revealed only 22% overlap in identity (717 genes) between damage (1326 genes) and differentiation (3172 genes) (Figure 9B), suggesting distinct roles for p53 dependent on cellular environment. Gene-ontology (GO) analysis revealed a startling distinction between genes regulated by p53 during differentiation versus damage (Table 2). While most of the p53-targets during differentiation are categorized primarily as genes involved in development (particularly in neuronal development, a pathway which is triggered by the RA signaling) and transcription regulation ($P < 10^{-6}$), damage-specific p53-targets are associated with cell migration and motility ($P < 10^{-4}$) (Table 2). Highly studied p53 targets, e.g., *CDKN1A*, *MDM2*, are significantly ($P < 10^{-6}$) represented in genes common to both differentiation and damage (Table 2).

A



B

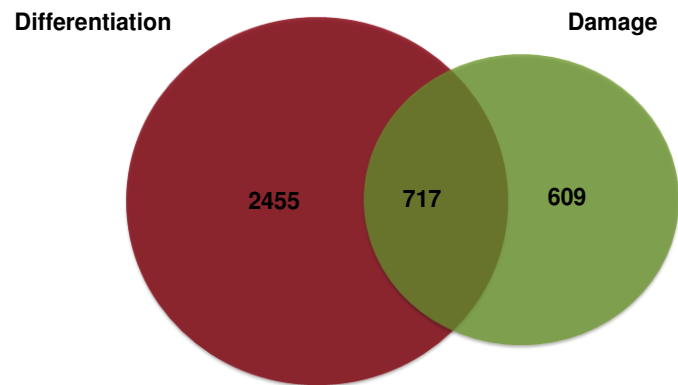


Figure 9. p53 targets distinct set of genes during differentiation and DNA-damage in hESCs

A) Distribution of p53 occupied regions relative to the nearest annotated TSS in hESCs undergoing differentiation or damage.

B) Numbers of distinct and overlapping p53-target genes in hESCs undergoing differentiation and DNA damage.

Differentiation Specific Target Genes				
Identifier	GO Term	Genes in the List	Total Genes	P-value
GO:0045449	regulation of transcription	452	2601	1.56E-12
GO:0048598	embryonic morphogenesis	73	307	3.49E-07
GO:0030182	neuron differentiation	94	438	9.40E-07
GO:0007389	pattern specification process	64	267	1.50E-06
GO:0030900	forebrain development	42	152	3.16E-06

Damage Specific Target Genes				
Identifier	GO Term	Genes in the List	Total Genes	P-value
GO:0016477	cell migration	21	276	3.10E-04
GO:0051674	localization of cell	22	307	4.78E-04
GO:0048870	cell motility	22	307	4.78E-04
GO:0006928	cell motion	29	475	6.52E-04
GO:0007266	Rho protein signal transduction	7	38	9.29E-04

Overlapping Target Genes				
Identifier	GO Term	Genes in the List	Total Genes	P-value
GO:0042981	regulation of apoptosis	61	804	4.67E-07
GO:0043067	regulation of programmed cell death	61	812	6.42E-07
GO:0008629	induction of apoptosis by intracellular signals	13	54	6.91E-07
GO:0010941	regulation of cell death	61	815	7.29E-07
GO:0043065	positive regulation of apoptosis	39	430	1.39E-06

Table 1. Response specific target genes are involved in different biological process

GO term analysis revealed significant and diverse functions of p53 downstream target genes that are specific or shared in response to each treatment (differentiation and damage).

Next, we determined enrichment of protein domains encoded by p53-target genes in each condition using InterPro terms of the GREAT functional annotation tool. Homeobox domains were revealed as differentiation targets ($P < 10^{-13}$). This finding is consistent with the GO-term analysis results since the proteins encode Homeobox domains are evolutionary conserved and developmentally important transcription factors with the ability to bind DNA through their Homeobox domains. On the other hand, EGF-type domains were targeted in damage ($P < 10^{-6}$) (Table 3), currently this domain's significance remains to be known because of its presence in protein families what seems to be unrelated.

Several transcription factor families that regulate specification and development are highly represented as differentiation targets (Figure 10). These include members of the Homeodomain-box (HOX) gene family, which are activated as a first response to RA and regulate pattern formation during embryogenesis [96]; LIM homeobox (LHX) genes, which are involved in embryonic development and specifically neuronal differentiation [169]; the forkhead box (FOX) family of genes, which are involved in axial patterning and tissue development from all three germ layers [170]; the sex determining region-Y box (SOX) gene family that regulates cell-fate specification [171]; and, Zic family members (ZIC) that are important during neuronal development, mutations of which cause a wide variety of congenital malformations [172] (Figure 10). These findings suggest that, during differentiation of hESCs, the regulatory influence of p53 is extensive and amplified by targeting transcription factors that promote a committed cellular state.

Enriched protein domains in Differentiation targets				
Identifier	INTERPRO Term-Name	Genes in the List	Total Genes	Binomial FDR Q-value
IPR009057	Homeodomain-like	109	314	7.62E-19
IPR012287	Homeodomain-related	106	304	8.81E-19
IPR001356	Homeobox	93	237	2.26E-17
IPR017970	Homeobox, conserved site	79	183	8.72E-16
IPR020479	Homeobox, region	40	87	1.65E-13

Enriched protein domains in Damage targets				
Identifier	INTERPRO Term-Name	Genes in the List	Total Genes	Binomial FDR Q-value
IPR001881	EGF-like calcium-binding	30	108	7.56E-07
IPR013091	EGF calcium-binding	25	87	1.29E-06
IPR013032	EGF-like region, conserved site	41	197	3.23E-06
IPR000152	EGF-type aspartate/asparagine hydroxylation site	28	102	6.64E-06
IPR018097	EGF-like calcium-binding, conserved site	27	99	8.63E-06

Enriched protein domains in Overlapping targets				
Identifier	INTERPRO Term-Name	Genes in the List	Total Genes	Binomial FDR Q-value
IPR020465	Tumour necrosis factor receptor 10	4	4	1.11E-03

Table 2. Significant number of p53 targets during differentiation possess homeobox domain

Enrichment analysis of protein domains encoded by p53 downstream target genes that are specific or common in differentiation and DNA-damage. Top categories from each dataset are listed.

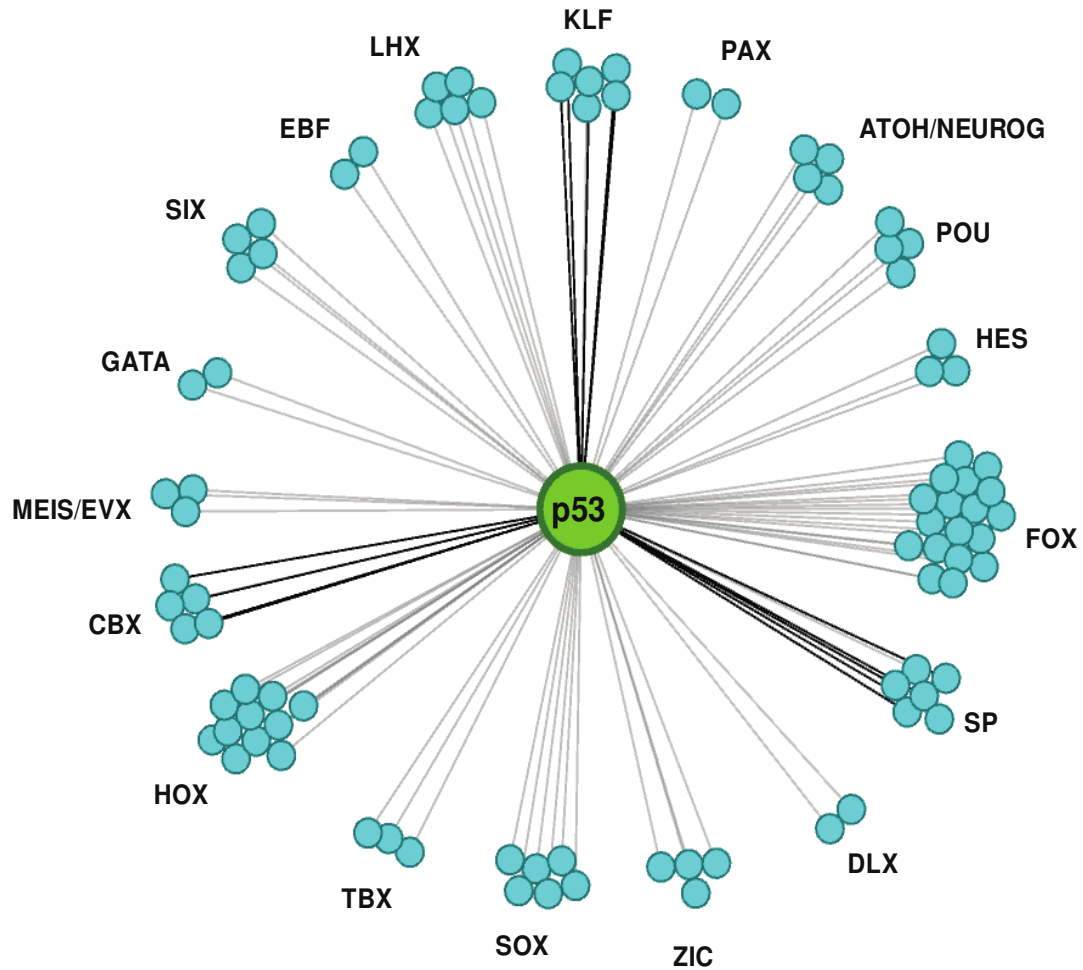


Figure 10. Developmental transcription factor families are targeted by p53 during hESC differentiation

Gene families of developmental transcription factors are targets of p53 during differentiation. p53 (green circle) regulation is linked to individual transcription factors (cyan circles), shown grouped by family.

3.4 p53 binding sites coincide with ESC transcription factors during differentiation

Developmental genes are often poised in ESCs by core pluripotency factors and bivalent histone modifications [35,36,37,38]. In addition, our motif analysis revealed that OCT4 and NANOG motifs are enriched at differentiation-induced p53 binding sites but not in DNA damage binding sites. Therefore, we analyzed the distribution of p53 binding sites, across four representative HOX loci of the human genome and compared them to OCT4, NANOG and H3K27me3 enrichment sites (Figure 11). A circular plot of human chromosomes 2, 7, 12 and 17, representing a ~100 Kb region of each HOX cluster, illustrates enrichment of OCT4, NANOG and H3K27me3 in pluripotent hESCs (Figure 11). During differentiation p53 binds (21 binding sites to 11 identified target genes) in and around these HOX clusters. In contrast, there is only one intergenic p53-bound site induced by DNA damage at these loci. These findings suggest that, during differentiation of hESCs, the regulatory influence of p53 is extensive and amplified by targeting transcription factors that promote a committed cellular state.

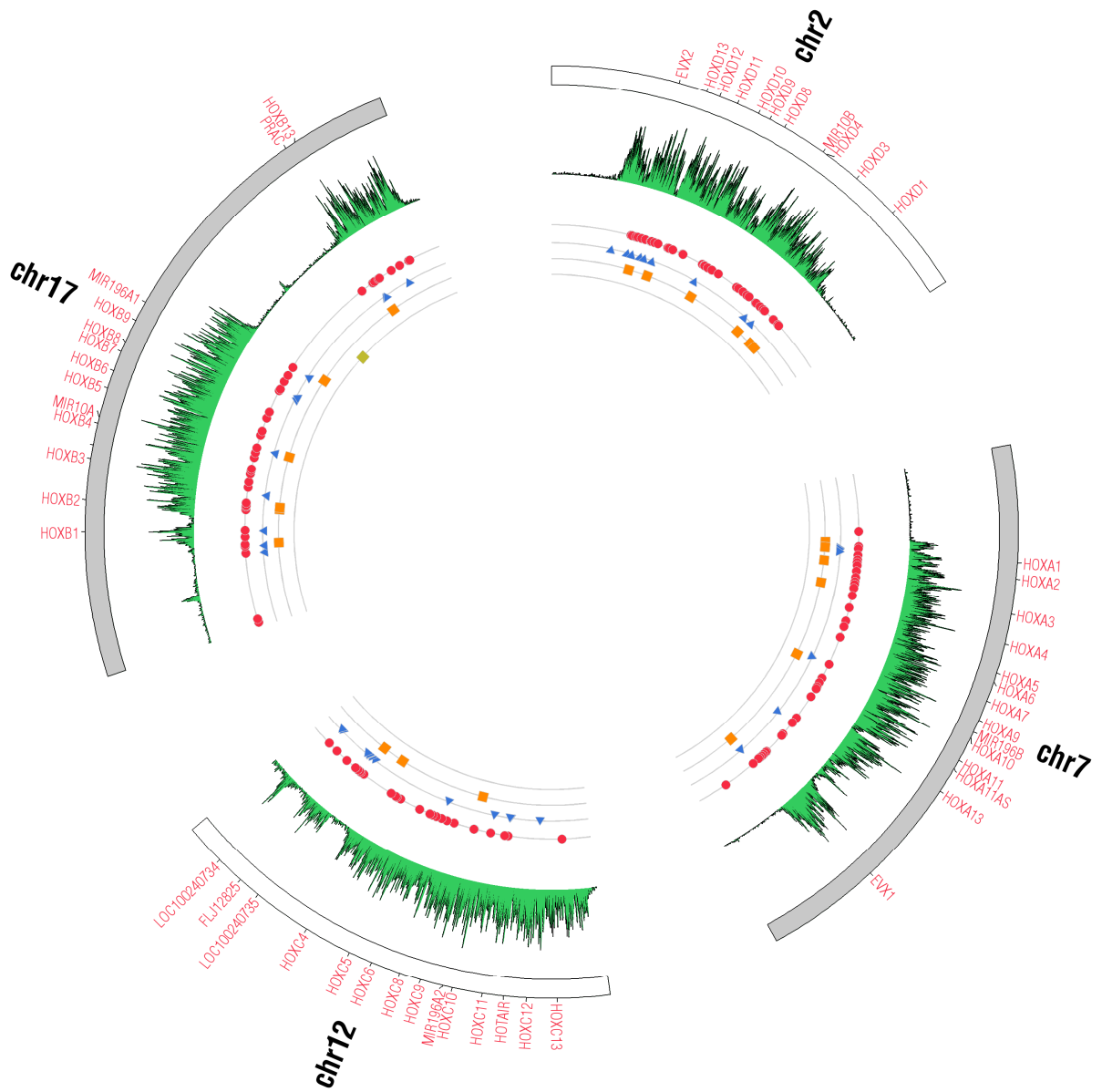


Figure 11. Binding profiles of p53, OCT4 and NANOG around human HOX loci

Circos plot of four human HOX gene clusters showing differential binding patterns of OCT4 (blue), NANOG (red), H3K27me3 (green) in pluripotent hESCs and p53 (damage:yellow, differentiation:orange).

Overlap between core transcription factors and differentiation-induced p53 binding sites around the HOX clusters lead us to investigate whether binding sites of mentioned factors overlap in a region specific or genome-wide fashion. Obtained results indicated that overlap between p53, OCT4 and NANOG binding sites is widespread across the genome, as ~50% of the 1000 highest confidence, differentiation-bound p53 sites are occupied by OCT4, NANOG or both in pluripotent hESCs; only a small fraction (~12%) overlap with damage-specific p53 sites (Figure 12A). Randomization tests demonstrated that percentage of differentiation-induced p53 binding sites that overlap with OCT4 and/or NANOG sites is significantly higher than those observed with randomly generated genomic sites, whereas overlap between damage-specific p53 sites and OCT4 or OCT4:NANOG binding sites are within random range (Figure 12B). We extended co-occupancy analysis to genome-wide by ranking each set of p53-binding sites (differentiation- and damage-induced) based on their enrichment scores and performed the intersection analysis for each segment. Results showed a significantly higher ratio of p53:OCT4:NANOG overlap and stronger p53-peaks at differentiation- versus damage-induced binding sites (Figures 13-14).

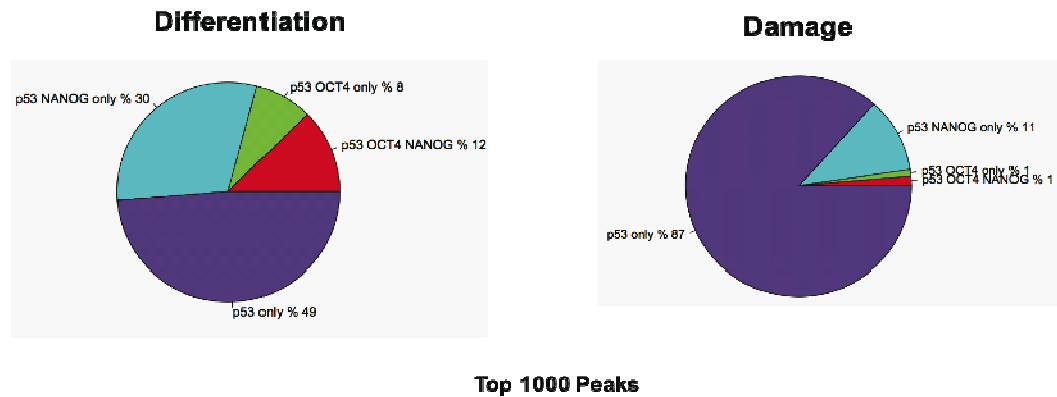
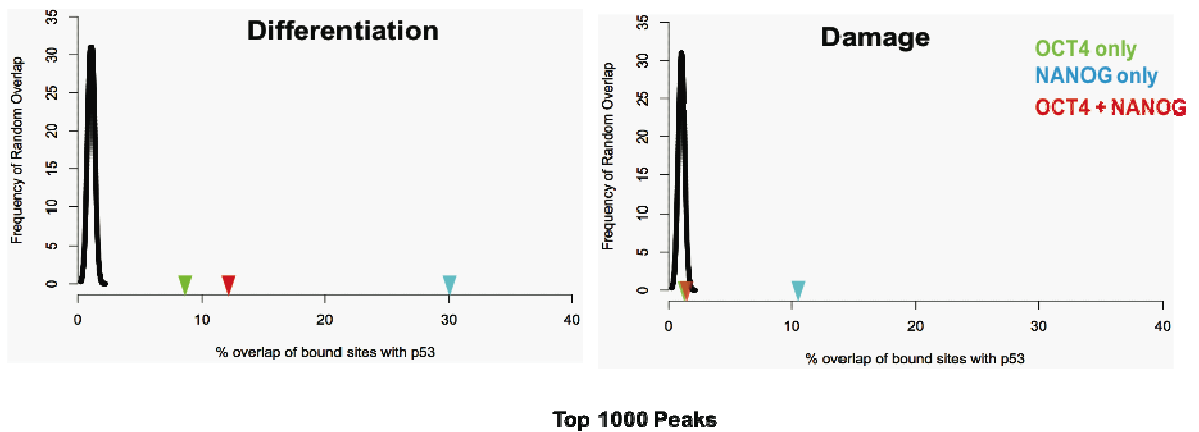
A**B**

Figure 12. p53 binding sites coincide with ESC transcription factors during differentiation

A) Overlap of top p53 binding sites with OCT4 and NANOG in hESCs undergoing differentiation or damage.

B) Plots indicate percent overlaps along the x-axis, solid curve represents expected overlap with random data.

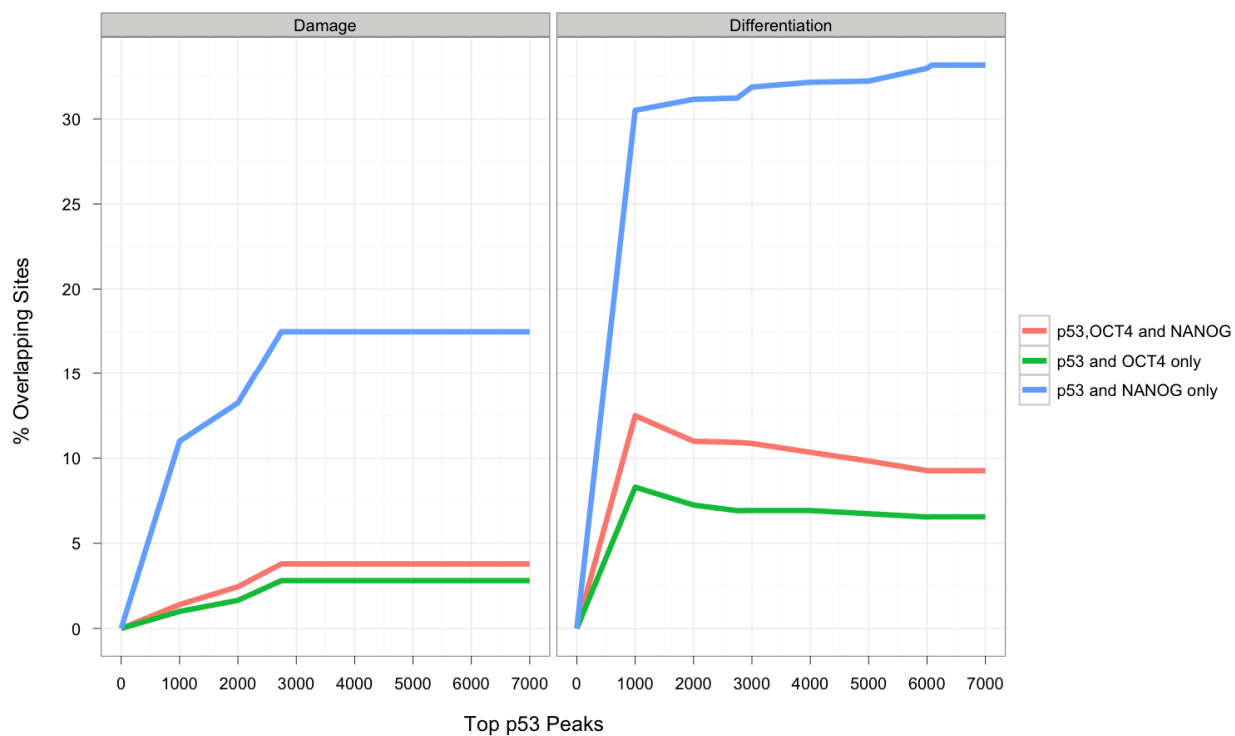


Figure 13. Association of OCT4 and/or NANOG binding sites with p53

Percent overlap among OCT4, NANOG and enrichment based top ranked p53 bound regions in hESCs undergoing damage (left) or differentiation (right).

In order to compare the raw signal intensities we performed heat map analysis, which revealed that ChIP-Seq signal intensity of OCT4 and NANOG at genomic sites bound by p53 exclusively during hESC differentiation is notably higher than their signals around p53-damage specific sites (Figure 14A). This suggests that a specific subset of genes (mostly developmental transcription factors) is kept in a repressed state by OCT4/NANOG during pluripotency and, in response to RA, p53 occupies nearby to regulatory regions of these genes to promote hESC differentiation.

Binding profiles and comparison of p53 and NANOG peaks reveal that OCT4 enrichment at p53 peaks, established during differentiation, is of the same magnitude as at NANOG sites (Figure 14B). However, NANOG enrichment is stronger at OCT4 binding sites than p53 (Figure 14C). The absence of OCT4 or NANOG at damage-induced p53 sites suggests that p53 plays distinct regulatory roles in hESCs, which are dictated by external stimuli.

A

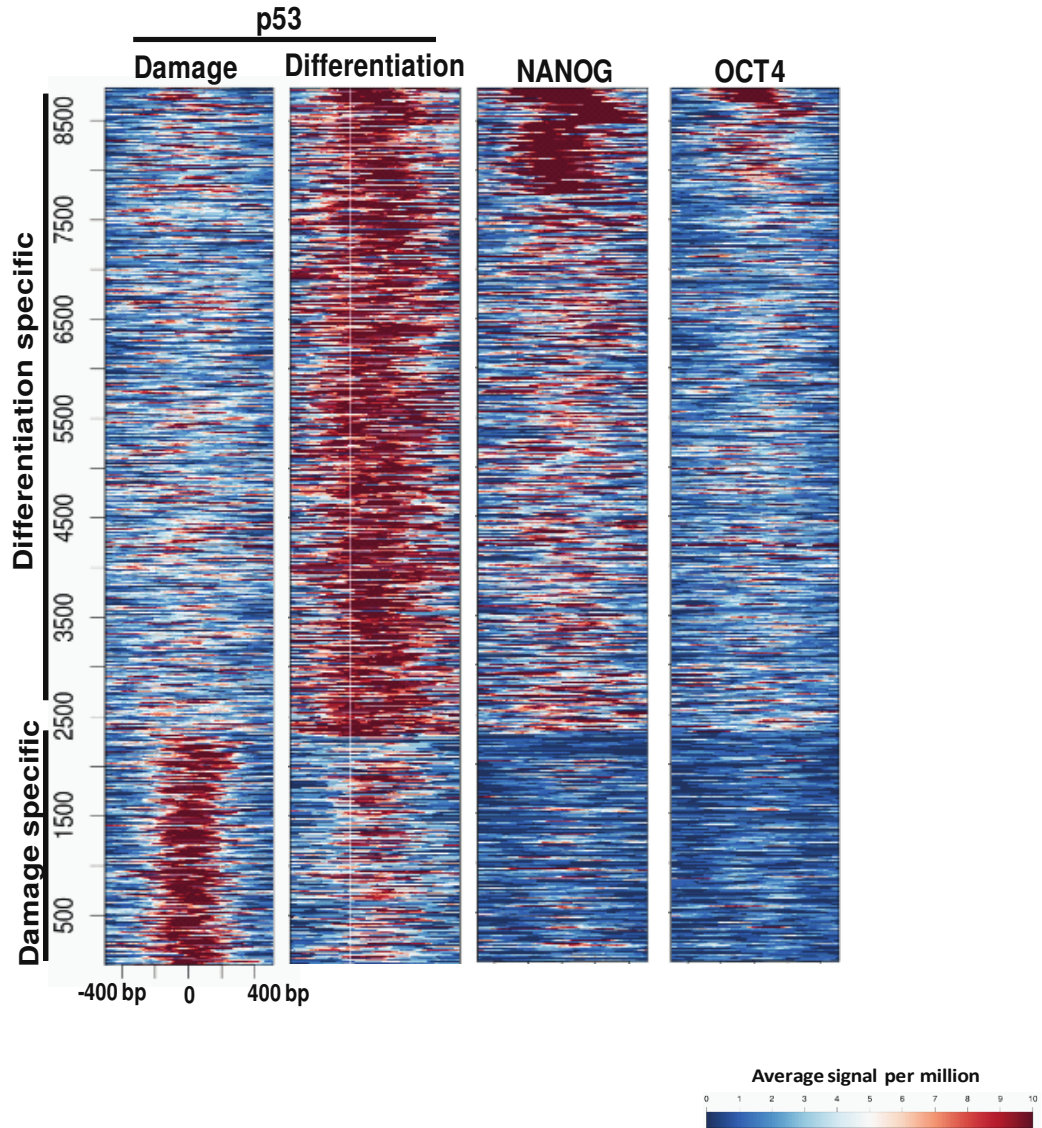


Figure 14. NANOG and OCT4 binding strengths are much higher at differentiation specific sites

Heat map of binding signals of p53 (damage and differentiation), OCT4 and NANOG within -500bp to +500 bp of p53 condition-specific peak summits.

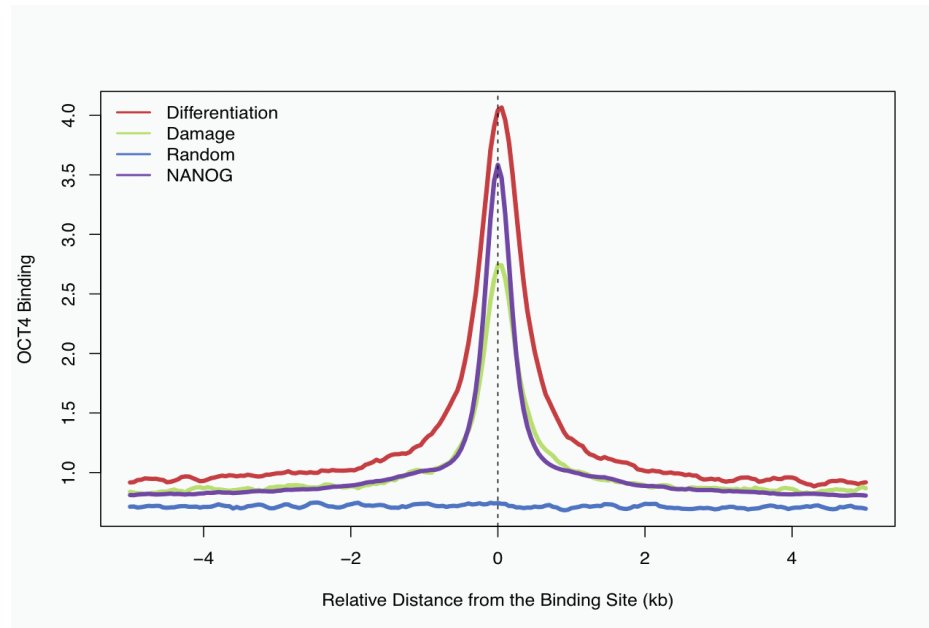
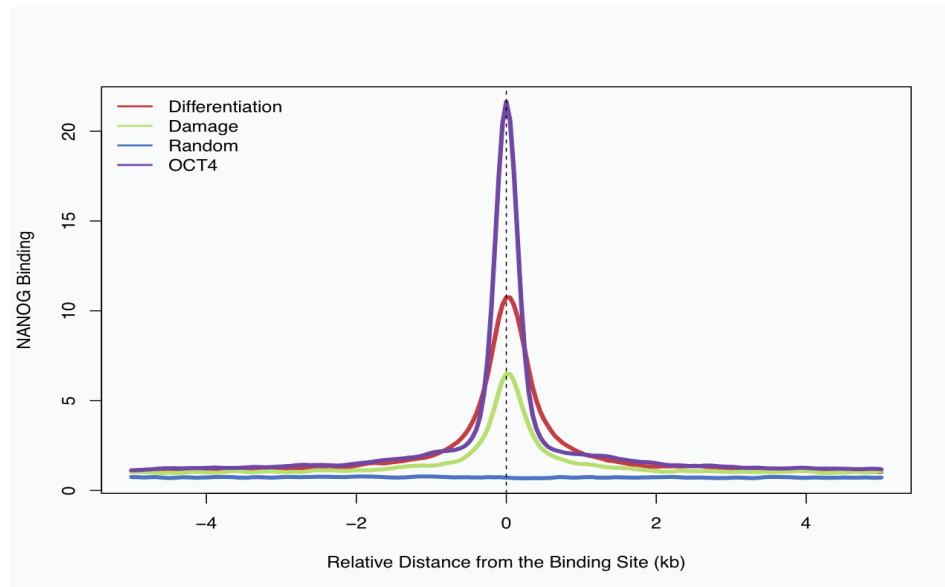
B**C**

Figure 14. NANOG and OCT4 binding strengths are much higher at differentiation specific sites

Aggregate plots shows average OCT4 (B) and NANOG (C) enrichment profiles around central position of p53 (Damage:green, Differentiation:Red) and NANOG/OCT4 (Purple) binding regions.

3.5 Transcription of development genes is dependent on p53

To uncover the functional consequences of p53 interactions with chromatin, we performed microarray-based gene expression analysis of hESCs undergoing differentiation and integrated these data with our p53 ChIP-Seq dataset (Figure 15). Expression analysis revealed a total of 1220 up- and 1221 down-regulated genes (with FDR-corrected p-value less than 0.05) during differentiation of hESCs compared to pluripotent state. Intersection with our p53 ChIP-Seq data revealed that more than 25% of genes regulated during differentiation (262 down- and 361 up-regulated) are bound by p53. We next sought to identify differentiation-specific p53 targets by eliminating genes that are targeted by p53 during DNA damage, as a result 198 down- and 271 up-regulated genes were assigned as p53's differentiation-specific targets and further analyses performed on this set of genes (Figure 15).

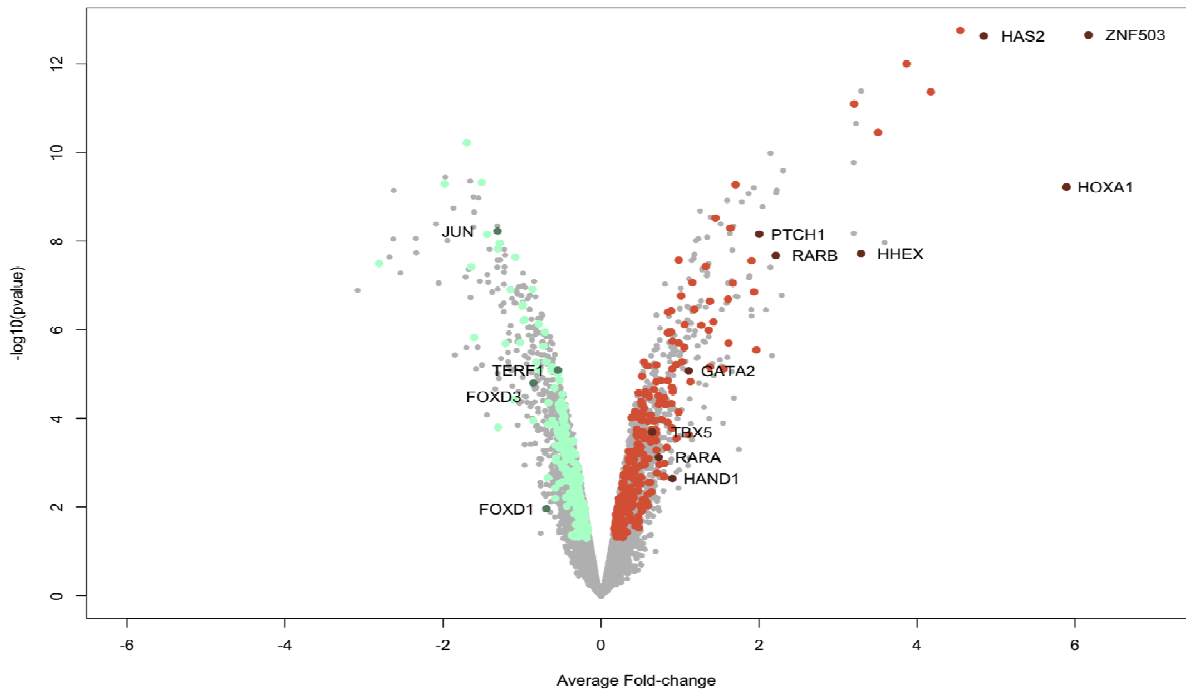


Figure 15. Integration of gene expression and p53 binding data in differentiating hESCs

Volcano plot of microarray gene-expression data. Each point corresponds to RefSeq gene; in RA treated samples with average log2 fold change compared to pluripotent hESCs and negative log10 p- value scores. Colored points correspond to genes bound by p53: significantly up- (red) or down- regulated (green) p53 targets are highlighted. Target genes overlapping with damage datasets are discarded.

GO-term analysis of RA-down-regulated p53 targets revealed that these genes are enriched for cell motion and mesodermal differentiation (Figure 16). These genes include FOXO3: essential activator of mesodermal marker Brachyury [173]; KLF6: associated with hematopoiesis [174]; chromatin modifiers HDAC5 and HDAC9: class II HDACs with critical functions in heart development [175]; and, telomere repeat binding factor TERF1: a telomere maintenance factor associated with pluripotency [176] (highlighted in Figure 15).

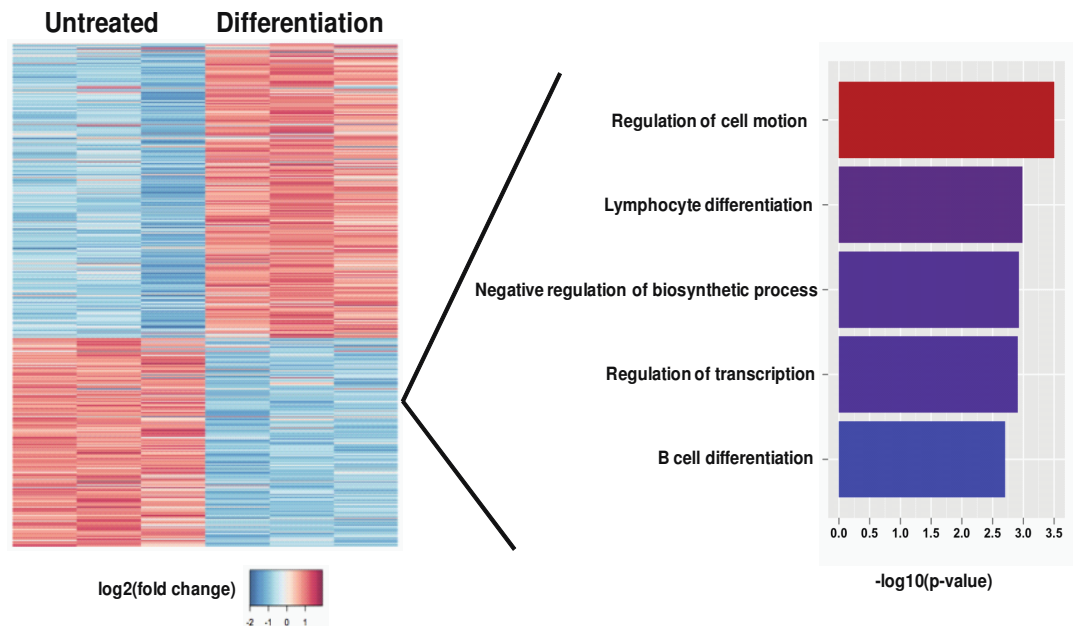


Figure 16. GO functional classifications of down-regulated p53

Heat map, generated for differentiation-specific p53 target genes, reveals up- or down-regulated targets during differentiation compared to pluripotent hESCs. The GO-term analysis of down-regulated p53-target genes is shown.

RA-up-regulated p53 targets revealed significant ($P < 10^{-5}$) representation in neuro-ectodermal development, embryonic morphogenesis and pattern specification categories (Figure 17). These genes include homeobox domain genes (HOXA1, HOXA3, HHEX and HOXB1), developmental transcription factors (GATA2, LHX8, ZIC1 and TCF7L2) and RA nuclear receptors (RARA and RARB) (highlighted in Figure 15). Several of these genes are repressed by Polycomb complexes and poised by core pluripotency factors in pluripotent hESCs [17], but a role for p53 in their activation during differentiation has not previously been reported.

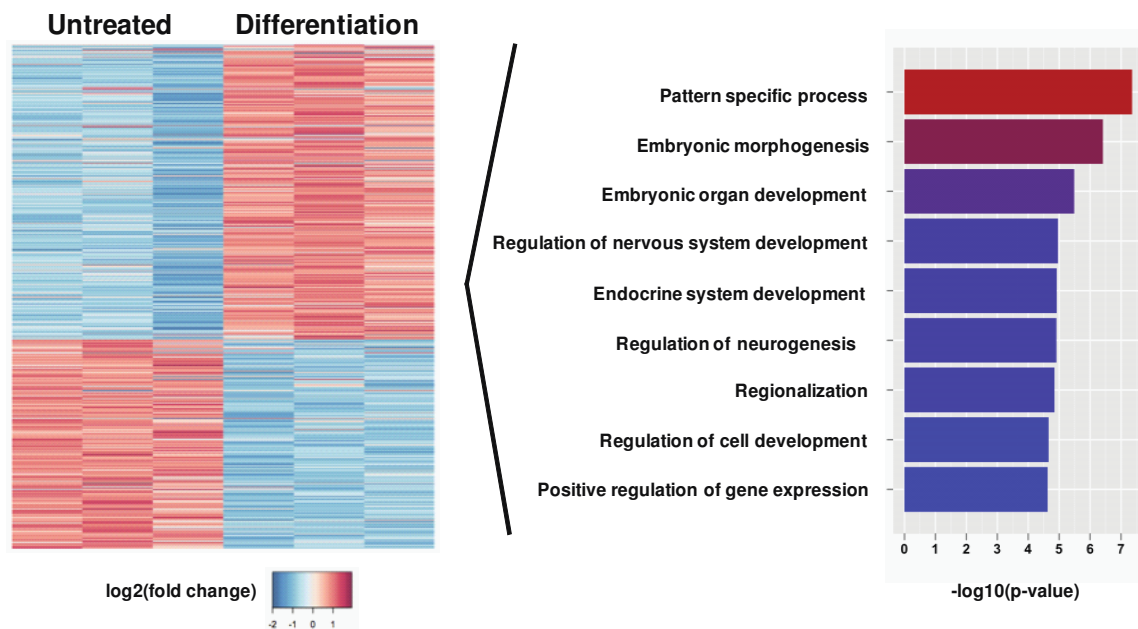


Figure 17. Up-regulated p53 targets are involved in developmental processes

Heat map, generated for differentiation-specific p53 target genes, reveals up- or down-regulated targets during differentiation compared to pluripotent hESCs. The GO-term analysis of up-regulated p53-target genes is shown.

We performed quantitative RNA and p53 ChIP-PCR analyses of selected genes (Figure 15), to assess the impact of p53 binding and to validate the outputs of our genome-wide assays (Figures 18-19). RA treatment for 2 and 4 days resulted in significant activation of genes belonging to HOX and GATA families (Figure 18A). Four days of RA increased expression of these genes, as well as developmental transcription factors: TBX5, homeobox genes MSX2 and GBX2, hedgehog receptor PTCH1, Notch co-repressor TLE3, polycomb protein BMI1 and histone H3K36 demethylase KDM2B (Figure 18B). Observed differences in the timing of target gene inductions may be due to a cascade of transcriptional events, where certain genes are activated as early as two days during RA-mediated hESCs differentiation, whereas it takes others longer to be induced.

RA-mediated transcriptional activation of selected genes is dependent on p53, since hESCs transfected with si*TP53* showed no significant activation of these genes with RA-treatment. In contrast, p53 induction by DNA damage had no significant effect on these genes (Figures 18A-B). Expression of well-known p53 pathway genes *CDKN1A* and *MDM2* was induced during both differentiation and damage in a p53-dependent manner, confirming the GO analysis results (Table 2) which indicated that p53-pathway genes are enriched in the shared targets under these two conditions. (Figure 18C).

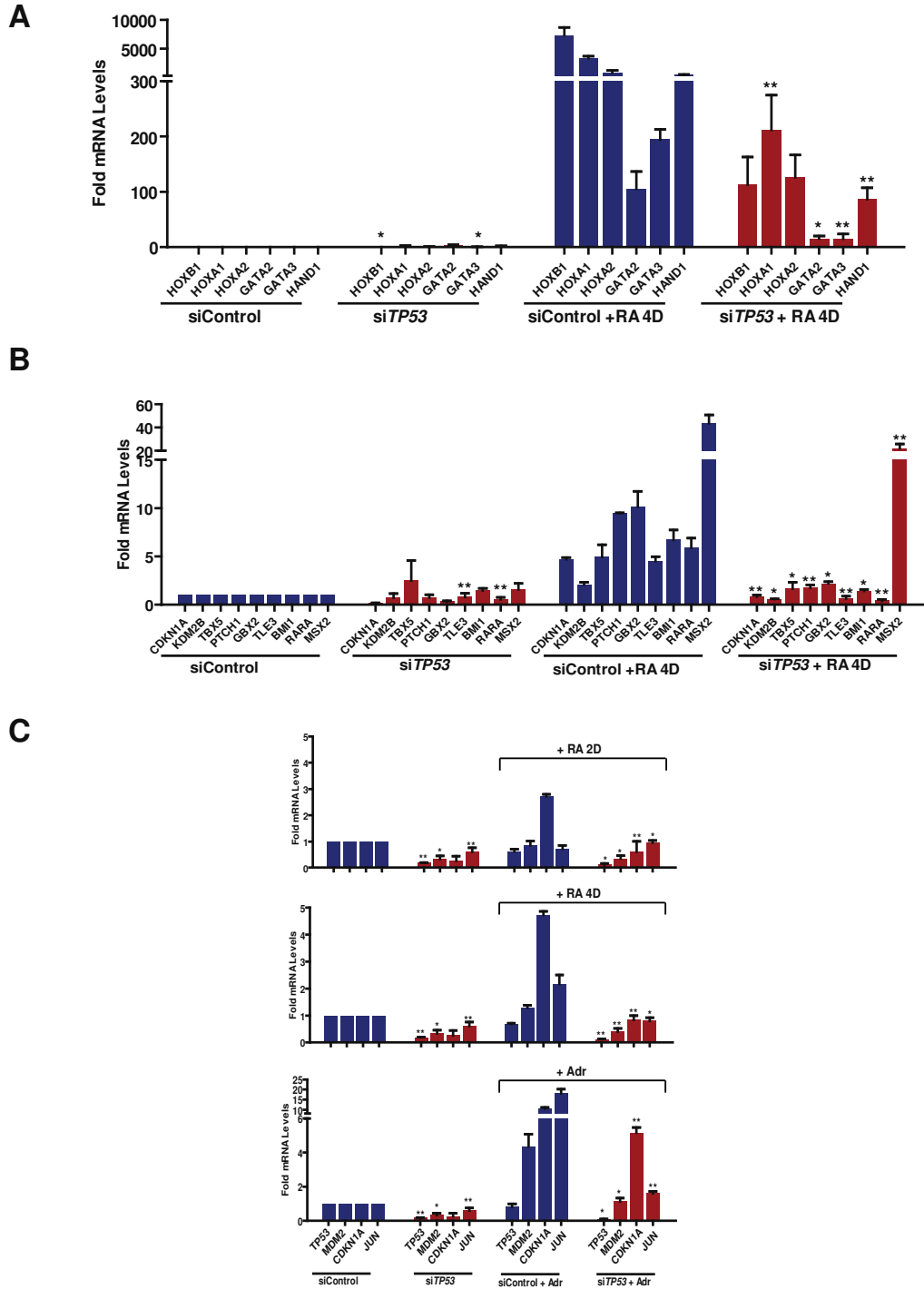


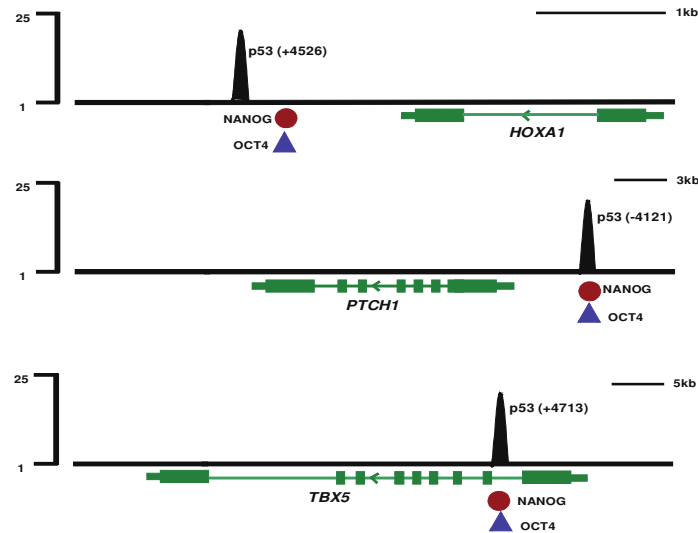
Figure 18. Transcription of developmental genes during RA-mediated differentiation is p53-dependent

RT-qPCR analyses of selected genes in hESCs after 4 d of RA-treatment with *TP53* or control non-targeting siRNA. Error bars represent standard deviation from three replicates (* <0.05, ** <0.01). [data contributed by Abhinav Jain]

We used positional weight matrixes (PWMs) obtained from transcription factor motif analysis (Figures 7-8) of p53-enriched genomic regions to map OCT4, NANOG and p53 binding elements at specific developmental genes: *HOXA1*, *PTCH1* and *TBX5* (Figure 19A). ChIP-qPCR analyses revealed robust enrichment of p53 binding, within two days of RA exposure, at the p53REs of *PTCH1*, *HOXA1*, *TBX5* and *CDKN1A* (Figure 19B). Importantly, p53-enrichment at these sites (*PTCH1*, *HOXA1* and *TBX5*) is RA-specific, since no significant changes observed in response to DNA-damage. On the other hand, in both conditions p53 enriched around the *CDKN1A* promoter, this suggests that developmental gene targeting is specific to p53's role in hESC differentiation (Figure 19B).

To assess whether OCT4 and p53 co-occupy the overlapping binding sites, we performed sequential ChIPs (re-ChIP) on OCT4-enriched chromatin fragments from hESCs treated with RA for 2 days (Figure 19C). RA robustly induced p53 enrichment and co-occupancy at OCT4-associated regions of *PTCH1* and *TBX5*, roughly equivalent to the increase in p53 association induced by RA (Figure 19C). The OCT4-OCT4 re-ChIP indicates equal efficiency of OCT4 binding to chromatin sites in both untreated and 2-day RA-treated hESCs. However, the distance between p53 and OCT4 binding sites on *HOXA1* (> 500bp) is greater than the vast majority of our fragmented chromatin length (Figure 19A) that's why re-ChIP experiments was not feasible for this genomic locus.

A



B

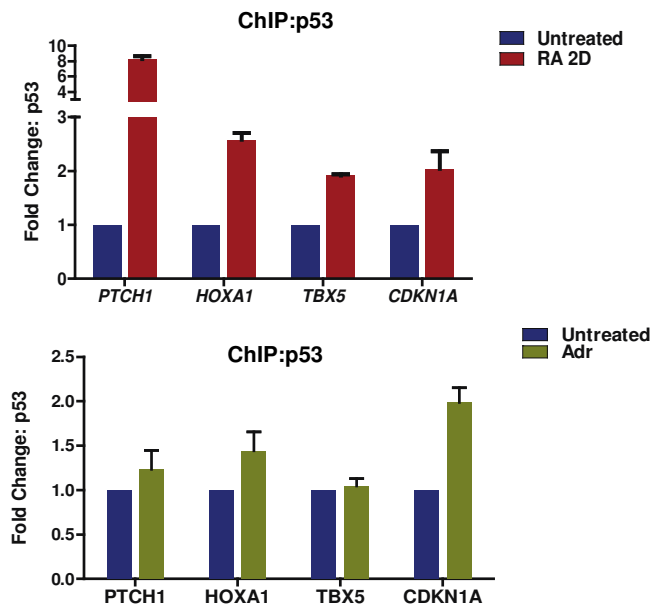


Figure 19. Enrichment of p53 at developmental genes results in activation

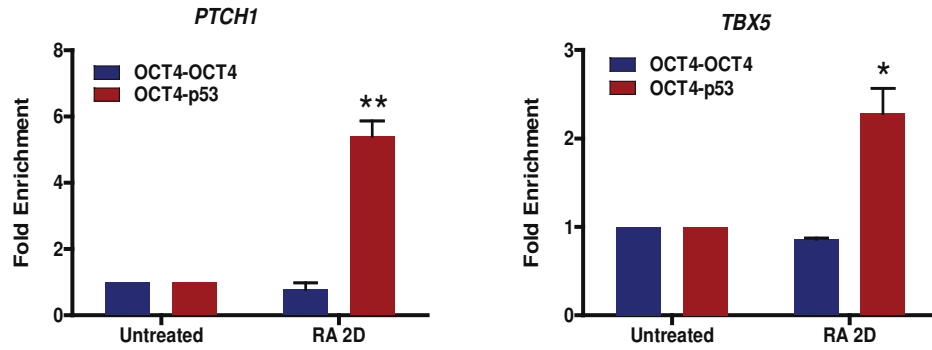
A) Tracks represent normalized p53 sequence tag enrichments (numbers indicate distance from TSS). Binding location of NANOG (red) and OCT4 (blue) are shown at the bottom of the tracks.

B) ChIP-qPCR analysis of p53 occupancy at select target genes during differentiation [top] or DNA damage [bottom]. [data contributed by Kendra Alton]

Developmental genes are held poised in ESCs by repressive histone marks (H3K27me3), which are lost upon differentiation [2]. We generated hESCs stably expressing non-target (shControl) or shRNA against p53 (sh*TP53*) to determine whether RA-activated p53 had an impact on levels of H3K27me3 at the promoters and/or p53-response elements (p53RE) of *PTCH1*, *TBX5* where p53 co-localizes with OCT4 at 2 days of RA treatment (Figure 19C). Stable integration of sh*TP53* resulted in a significant knockdown of p53 protein and failure to elicit an RA-response, since no reduction in AP-staining and OCT4 protein was observed in sh*TP53*-hESCs as compared to control (data not shown). In response to RA, H3K27me3 levels are significantly reduced at *PTCH1* and *TBX5* in shControl cells, whereas no change in H3K27me3 levels were observed in hESCs stably depleted of p53 (sh*TP53*) (Figure 19D).

Together, these results suggest that RA-induced signals of differentiation mobilize p53 to bind and activate a number of chromosomal locations around the developmentally important transcription factors that are poised by OCT4/NANOG in pluripotent hESCs by altering the chromatin status.

C



D

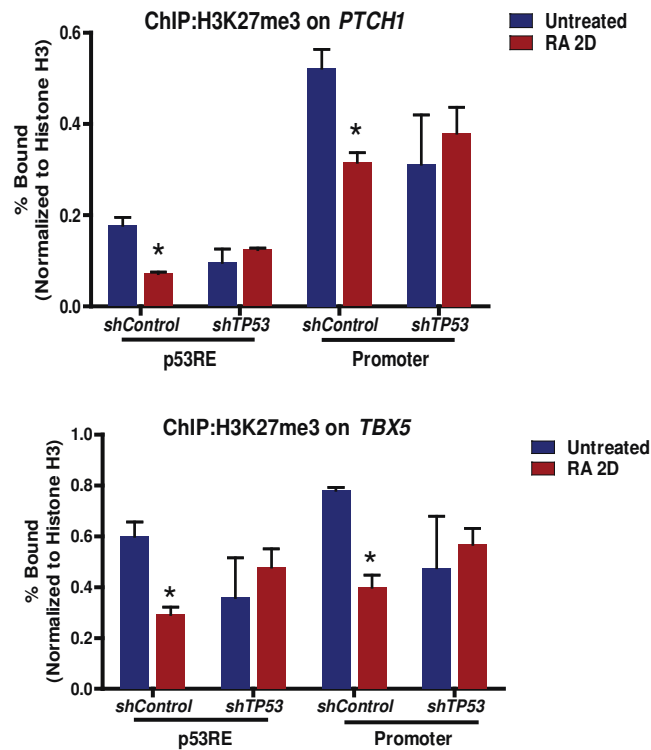


Figure 19. Enrichment of p53 at developmental genes results in activation.

C) p53 enrichment on OCT4 bound regions after sequential ChIPs. Quantitative PCR of chromatin fragments enriched by p53, OCT4 and sequential ChIP of hESCs, treated with RA for 2 days. DNA enrichments at indicated target genes were determined as fold change in % input, compared to untreated hESCs.

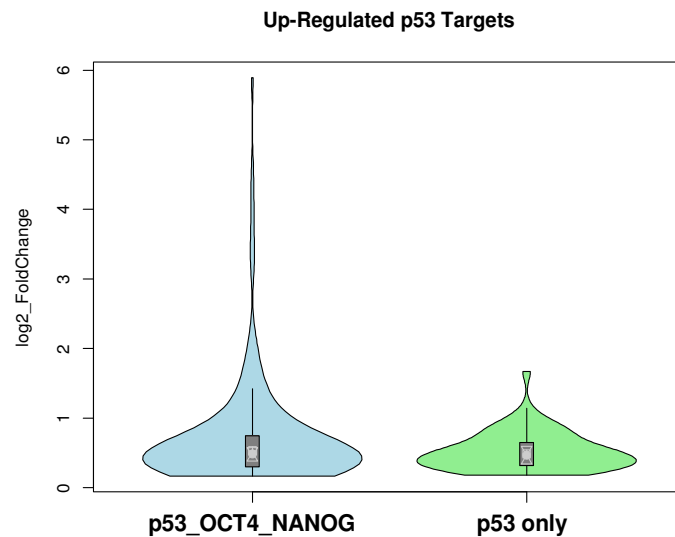
D) Histone H3K27me3 status on gene promoter or p53RE of *PTCH1* and *TBX5* in hESCs treated with RA for 2 days. Error bars represent standard deviation from three replicates (* < 0.05, ** < 0.01). [data in Figs 19C-D contributed by Kendra Alton]

3.6 p53 targets lose repressive histone marks during differentiation

We next sought to determine if changes in bivalent chromatin structure occur globally around the p53-target genes during differentiation, by analyzing genome wide histone status utilizing ChIP-Seq method for active (H3K4me3) or repressive (H3K27me3) histone marks in hESCs undergoing differentiation. To define histone tail modifications at the promoters of p53 targets, we first categorized the p53's differentially expressed targets as those that have overlapping OCT4 and/or NANOG binding sites, and the ones that are targeted by p53 only (Figures 21-22). Gene expression profiling revealed that while the average expression of the two sets are comparable, p53 gene targets that are shared with those bound by OCT4 and/or NANOG prior to differentiation are the most significantly changed (up- or down-regulated) genes (Figures 21A and 22A). Consistent with the biological functions of all differentiation-specific p53-targets (Table 2), GO-term analysis for up-regulated p53 targets with overlapping OCT4 and/or NANOG sites revealed genes responsible for pattern specification, embryonic morphogenesis and development (Figure 20B).

On the other hand, down-regulated p53 targets with overlapping OCT4 and/or NANOG sites are involved in mesodermal differentiation, metabolism and cell motion (Figure 21B).

A



B

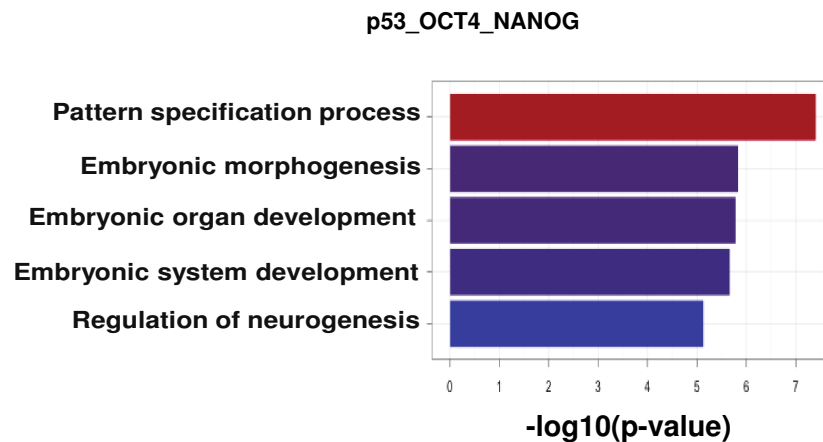
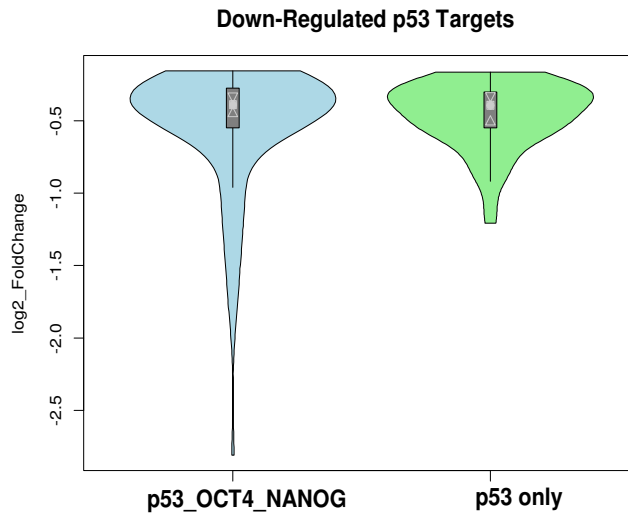


Figure 20. p53's overlapping targets with OCT4 and NANOG are more robustly expressed during differentiation

A) Violin plots representing fold changes in expression of p53 targets up-regulated during differentiation. Genes that have p53 binding sites overlapping with OCT4 and/or NANOG (p53_OCT4_NANOG) (blue); or only p53 binding sites (green).

B) The GO-Term analysis of overlapping targets of p53_OCT4_NANOG is shown.

A



B

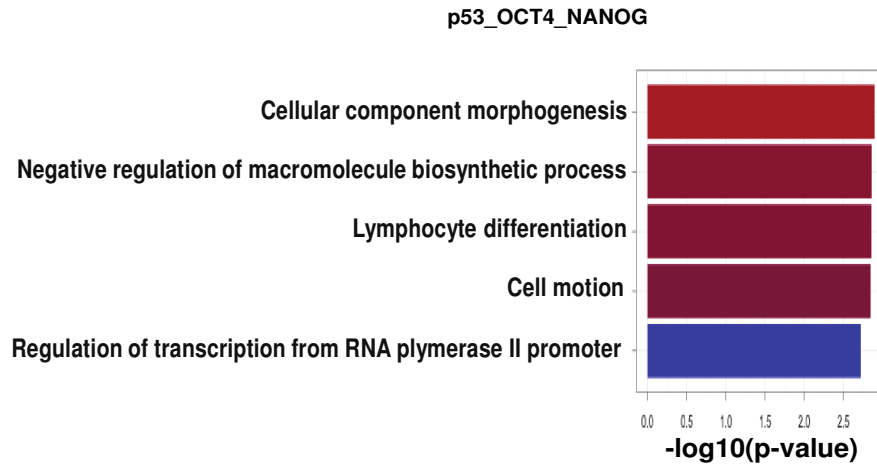


Figure 21. GO functional classification results down-regulated p53 targets with overlapping OCT4 and/or NANOG sites

A) Violin plots representing fold changes in expression of p53 targets down-regulated during differentiation. Genes that have p53 binding sites overlapping with OCT4 and/or NANOG (p53_OCT4_NANOG) (blue); or only p53 binding sites (green).

B) The GO-Term analysis of overlapping targets of p53_OCT4_NANOG is shown.

Genome wide profiling of average histone modifications confirmed that up-regulated p53-targets, overlapping with OCT4 and/or NANOG sites, are associated with bivalent histone marks (H3K4me3 and H3K27me3), which are significantly altered during differentiation (high H3K4me3, low H3K27me3), as compared to down-regulated targets (Figures 22A and 22C). However, genes targeted by p53 only gain H3K4me3 marks without a significant change in H3K27me3 status (Figures 22B and 22D).

Taken together, these results suggest that p53 plays an active role, possibly cooperating with core pluripotency factors, during differentiation of hESCs by recruitment of chromatin modifying complexes, which decrease repressive histone marks of specific developmental genes held poised in pluripotent stem cells.

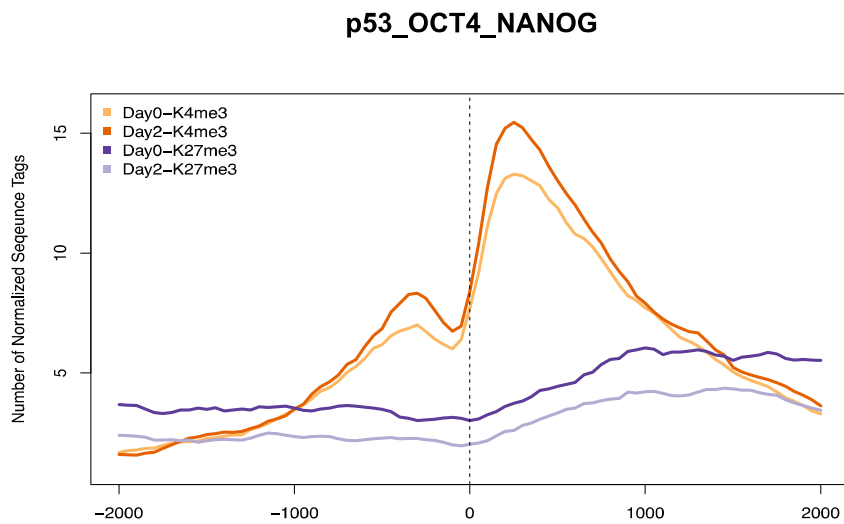
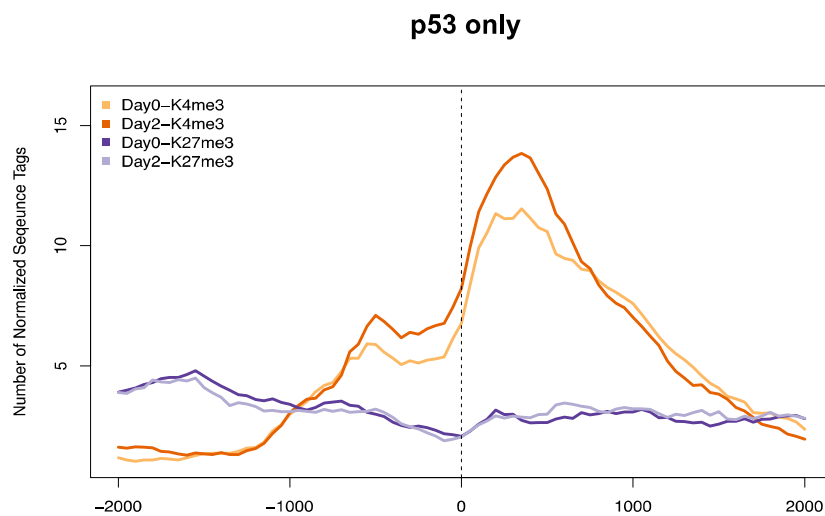
A**B**

Figure 22. Bivalent chromatin marks around promoter regions of p53 target genes in pluripotent and differentiating hESCs

Aggregate plots showing profiles of histone modifications around +/- 2KB from transcription start site (TSS) of *up-regulated* p53_OCT4_NANOG overlapping gene targets (A) and only p53 targets (B).

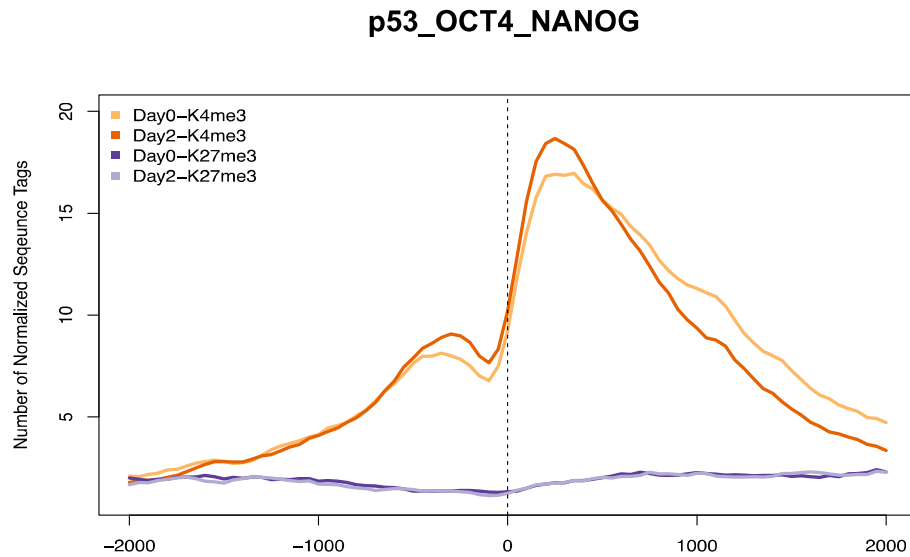
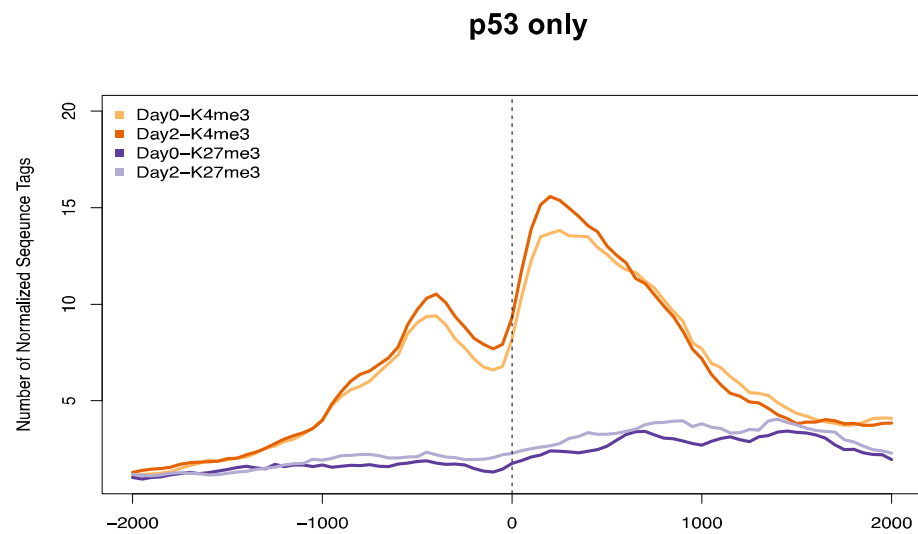
C**D**

Figure 22. Bivalent chromatin marks around promoter regions of p53 target genes in pluripotent and differentiating hESCs.

Aggregate plots showing profiles of histone modifications around +/- 2KB from transcription start site (TSS) of *down-regulated* p53_OCT4_NANOG overlapping gene targets (A) and only p53 targets (B).

CHAPTER 4 DISCUSSION AND FUTURE DIRECTIONS

4.1 Discussion

Studies of p53 are extensive; specifically its functions in cell cycle regulation and apoptosis have been scrutinized for several decades in transformed somatic cells [115,177,178]. The broader potential in regulatory roles of numerous cellular processes was only recently appreciated. For an example, p53 has been implicated in regulating cellular metabolism, deregulation of p53 leads to compromised oxidative phosphorylation chain, which is also known as Warburg effect, one of the hallmarks of cancer cells [179,180].

On the other hand, a limited knowledge of p53's function in non-transformed cells; especially in highly proliferative undifferentiated cells, such as embryonic stem cells, therefore its role in development and control of cell-fate is largely unknown [116]. In order to dissect p53's functions during transcription in human ESCs cultured under different culture conditions (Adriamycin for DNA damage and RA for differentiation), we performed genome-wide p53-chromatin binding assays along with gene expression microarrays. Integration of the data output from these comprehensive methods revealed that the RA-mediated p53-response during differentiation is highly distinct from the stress-responsive events occurring downstream of DNA damage in hESCs. During early differentiation, p53 activates the expression of several developmental transcription factor families, many of which possess homeobox protein domains. This activated cascade of transcription factors

amplifies the functional effects of p53 induction beyond the transient time period when p53 protein is elevated [140].

Differentiation-specific p53-activated genes include members of HOX, FOX, SOX, T-box (TBX) and Chromobox (CBX) gene families that are involved in differentiation and development. HOX genes are known to be involved in patterning during embryogenesis as major developmental factors [181], for example HOXA1 is essential for RA-mediated neural differentiation [98]. FOX family members have been implicated in formation of different organs during development [170], such as liver. Mutations in SOX family genes impair proper differentiation and have been related to several developmental disorders [171]. Members of the CBX family, particularly CBX2 and CBX4 are part of the Polycomb complex [182] and are vital for cell-fate determination [172]; whereas the TBX gene family regulates a diverse range of developmental processes from early body planning to late organogenesis [183].

One facet of p53 gene regulation involves repression of some transcription factors and epigenetic modifiers while activating another set of developmental genes required for RA-mediated neuro-ectodermal lineage specification. Some of the down-regulated p53 targets include regulators required for mesodermal lineage specification such as, transcription factors FOXO3 [173], HEY1 [184] and KLF6 [174]; histone deacetylases HDAC5, HDAC6 [175] and chromatin remodeler CHD7 [185]. Several proteins that are involved in transcriptional repression are also targeted by p53 for down-regulation including telomere repeat factor TERF1 [176], PcG complex

competent RNF2 [186] and Chromobox family member CBX5 [187]. Taken together, p53 might play a significant role in lineage determination by RA-induced p53-mediated repression and activation of specific genes in hESCs.

Remarkably, our motif finding analysis revealed that the differentiation-specific p53-bound sites are also enriched in OCT4:SOX2 motif. Moreover, comparison of binding sites showed that more than half of the strongest p53-bound sites are coincident with binding sites of core pluripotency factors OCT4 or NANOG, or both, in pluripotent hESCs. This suggests that there could be interplay between p53 and the core pluripotency factors, specifically during early hESC differentiation since this phenomenon is not observed for p53's binding sites during DNA-damage. Our experimental validations showed that three developmental genes HOXA1 [98], PTCH1 [188] and TBX5 [189] are up-regulated during hESC differentiation in a p53-dependent manner. Chromatin immunoprecipitation (ChIP) studies revealed that OCT4 and NANOG are bound at or in the proximity of p53-binding sites at these developmental genes during differentiation. Sequential-ChIP assay confirmed that during differentiation p53 indeed co-localizes to these regions, which are bound by OCT4. However, our current findings cannot conclude whether p53 recruitment ultimately results in displacing the bound OCT4 and/or NANOG proteins at the regulatory sites or these factors synergistically bring other chromatin modifiers to those loci, thereby activating down-stream targets expression. Elucidation of the exact mechanism requires further experiments.

Given the importance of bivalent domains in pluripotency maintenance and establishment of cell fate [35,36,37,38], we profiled the bivalent histone modifications (H3K4me3 and H3K27me3) in pluripotent and differentiating hESCs by CHIP-Seq. Our analyses revealed that up-regulated p53 targets, which are also bound by OCT4 and/or NANOG, are kept poised in ESCs by bivalent modifications and during differentiation promoter regions of these genes acquire more H3K4me3 mark while losing their H3K27me3 modifications. Furthermore, we tested if p53 has any roles regulating the chromatin modification switch near its target genes during differentiation. Notably, PTCH1 and TBX5 gene promoters could not lose their promoter-associated H3K27me3 marks during differentiation in p53-depleted hESCs. These results suggest that p53 might play a significant role in modifying chromatin structure at its poised target genes by coupling with an unknown H3K27 demethylase complexes during hESC differentiation.

The shared target genes of p53 during differentiation and DNA damage response are enriched in cell cycle regulation. p53-regulated cell-cycle control pathways play significant roles in both during hESC differentiation, by impeding cell cycle and leading differentiation [140], and DNA-damage repair by blocking the self-renewal pathway in order to prevent accumulation of chromosomal damage. Metabolism, another common GO term for conserved p53 target genes, suggests the link between p53 and metabolism could be as crucial as cell cycle pathways during both development and tumor suppression [179,190].

The most interesting GO terms that are identified specifically in p53 targets during DNA-damage, cell motion and cell migration, are the signature characteristics of metastatic carcinomas. For example, damage specific p53 gene targets listed under GO category of cell motion, FGF2 and LRP8, have been grouped into the stem-like gene expression sets that are only observed in p53 loss-of function cancers [126]. Moreover, two other cell motion-associated p53 targets, MMP14 and TNFRSF12A, are classified in epithelial to mesenchymal transition (EMT), which is a required step for metastasis [191], genes in prostate cancers [126]. Further examination of DNA-damage specific targets provides an opportunity to dissect profiles of aggressive metastatic tumors by monitoring changes in activities of these genes as an indication of deregulated p53-pathway.

Our study unveiled p53's important regulatory functions in the human embryonic differentiation, which does not align with the previously reported findings about p53's role in mESCs. Previous reports have shown that p53 binds to the promoter of Nanog in mESCs and suppresses its transcription, which leads to differentiation of mESCs [131]. Instead, we did not detect any p53 binding sites nearby NANOG regulatory regions in our p53 ChIP-Seq results in hESCs. Secondly, Li et al. recently reported that in response to DNA-damage p53 both activates differentiation-associated genes and represses ES-specific genes in mESCs [133]. However, our results in hESCs indicate that p53 targets a different set of genes during differentiation versus DNA-damage and only differentiation-specific p53 target genes are related with development and

specification. These findings implicate that unlike mouse ESCs, p53 does not repress pluripotency factors in human ESCs, yet only mediate expression of developmental genes. Moreover, p53's pro-differentiation role takes place under different environmental conditions (DNA-damage in mouse and differentiation-initiation in human ESCs) in different species (Figure 23, p53 targets several Hox genes upon DNA-damage in mESC but binds to only a single intergenic region in human HOX cluster loci after exposed to the same stress in hESCs). Observed species-specific differences in p53's functions in two organisms may be attributed to the different embryonic development stages of mouse and human ESCs [192]. In parallel, mounting evidences demonstrate a rapid evolutionary turnover for transcription factor binding sites on a genome-wide scale between species which results in regulation of a diverse set of genomic elements in different species by the same transcription factor [134,193,194,195,196].

Given the p53's significant role in promoting hESCs differentiation, viability of p53-null mice and formation of teratomas in SCID mice from p53-null hESCs raises some interesting questions [197]. In this case, we believe compensation of p53 functions in development would likely to be executed by the structurally related protein family members, p63 and p73 [198]. Notably, several developmental abnormalities such as neural tube malformations or defects in spermatogenesis and embryo implantation have been reported despite the fact that p53-null mice are not embryonic lethal [116]. This suggests that p53's functions are imperfectly compensated by other factors, but whether

p63 or p73 isoforms target any or all p53 downstream targets in hESCs
differentiation remains to be investigated.

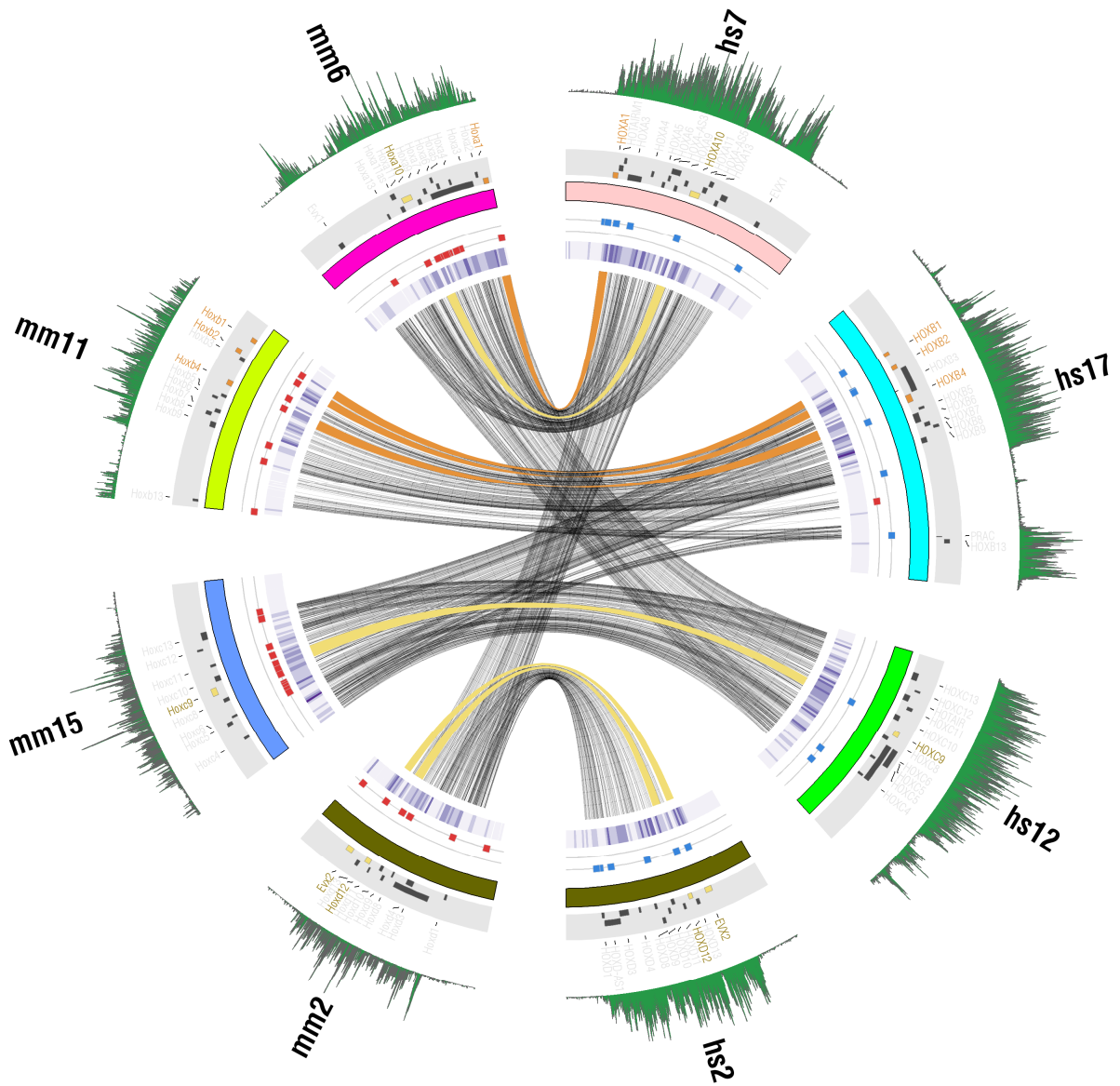


Figure 23. Species-specific binding of p53 in different environmental stimuli

Human (hs) and mouse (mm) HOX gene clusters loci are shown in this circular plot. Green track represents repressive H3K27me3 mark around the displayed regions in mESCs and hESCs. Tiles (black, orange or yellow bars) show underlying structures of HOX genes. Red (DNA-damage) and blue (differentiation) rectangles represent enriched p53 binding sites in these two conditions. Purple heatmap shows the PhastCons scores around the displayed regions. Ribbons show syntenic genomic locations between mouse and human (orange ribbons presents homologous p53 binding sites between DNA-damaged mESCs and differentiating hESCs, whereas yellow ribbons are for shifted sites for same gene targets in mESCs and hESCs).

4.2 Future Directions

Our mapping results revealed that for both DNA damage and differentiation of hESCs, p53-binding sites are enriched mostly in intergenic regions of the genome where non-coding RNA expression initiates (more than 50% of total binding sites in DNA damage and differentiation are located in gene desert regions). Binding sites of p53 around these intergenic sites gain significant value when the recent reports about ncRNAs (lncRNAs and miRNAs) and their effects in pluripotency and differentiation are taken into consideration [62]. Additional studies are required to confirm p53's regulatory significance in regulation of ncRNAs expression and possible down-stream roles of those p53-regulated RNAs in hESCs differentiation.

Members of p53 family, p63 and p73, can also regulate the gene-expression program that is mainly directed by p53, in which p73 had been shown to serve as a back-up protein for maintaining genomic integrity when p53 functions are compromised [199]. These proteins are also implicated in important developmental processes [200] such as p63 in epithelial ESC self-renewal [201] and p73 during neural cell differentiation [202]. Notably, significant portion of amino acids in DNA-binding domains, ~85%, are conserved among p53 family members, further reports revealed that p63 and p73 co-occupy target sites with a shared consensus motifs similar to those of p53 [203]. Therefore, obtaining genome-wide binding maps of p63 and p73 in differentiating or DNA-damaged hESCs would eventually lead a more

comprehensive understanding of the roles of this tumor-suppressor protein family role in human development.

Understanding the differences in regulatory networks for balancing pluripotency and differentiation between mouse and human ESCs, it would be important to establish genome-wide p53 binding sites in differentiating mouse ES and epiblast stem cells. Mouse epiblasts are considered to be more developmentally close to human ESCs [137,192] and thus determination of p53's binding sites will help to understand the regulatory functions of p53 in development of these two organisms.

REFERENCES

- [1] J.A. Thomson, J. Itskovitz-Eldor, S.S. Shapiro, M.A. Waknitz, J.J. Swiergiel, V.S. Marshall, J.M. Jones, Embryonic stem cell lines derived from human blastocysts, *Science* 282 (1998) 1145-1147.
- [2] R.A. Young, Control of the embryonic stem cell state, *Cell* 144 (2011) 940-954.
- [3] Y.H. Loh, L. Yang, J.C. Yang, H. Li, J.J. Collins, G.Q. Daley, Genomic approaches to deconstruct pluripotency, *Annu Rev Genomics Hum Genet* 12 (2011) 165-185.
- [4] M.G. Guenther, Transcriptional control of embryonic and induced pluripotent stem cells, *Epigenomics* 3 (2011) 323-343.
- [5] F. Spitz, E.E. Furlong, Transcription factors: from enhancer binding to developmental control, *Nat Rev Genet* 13 (2012) 613-626.
- [6] J.M. Vaquerizas, S.K. Kummerfeld, S.A. Teichmann, N.M. Luscombe, A census of human transcription factors: function, expression and evolution, *Nat Rev Genet* 10 (2009) 252-263.
- [7] J.C. Heng, Y.L. Orlov, H.H. Ng, Transcription factors for the modulation of pluripotency and reprogramming, *Cold Spring Harb Symp Quant Biol* 75 (2010) 237-244.
- [8] J. Nichols, B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Scholer, A. Smith, Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4, *Cell* 95 (1998) 379-391.

- [9] K. Mitsui, Y. Tokuzawa, H. Itoh, K. Segawa, M. Murakami, K. Takahashi, M. Maruyama, M. Maeda, S. Yamanaka, The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells, *Cell* 113 (2003) 631-642.
- [10] J. Silva, J. Nichols, T.W. Theunissen, G. Guo, A.L. van Oosten, O. Barrandon, J. Wray, S. Yamanaka, I. Chambers, A. Smith, Nanog is the gateway to the pluripotent ground state, *Cell* 138 (2009) 722-737.
- [11] H. Niwa, J. Miyazaki, A.G. Smith, Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells, *Nat Genet* 24 (2000) 372-376.
- [12] S. Masui, Y. Nakatake, Y. Toyooka, D. Shimosato, R. Yagi, K. Takahashi, H. Okochi, A. Okuda, R. Matoba, A.A. Sharov, M.S. Ko, H. Niwa, Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells, *Nat Cell Biol* 9 (2007) 625-635.
- [13] A.A. Avilion, S.K. Nicolis, L.H. Pevny, L. Perez, N. Vivian, R. Lovell-Badge, Multipotent cell lineages in early mouse development depend on SOX2 function, *Genes Dev* 17 (2003) 126-140.
- [14] K. Arnold, A. Sarkar, M.A. Yram, J.M. Polo, R. Bronson, S. Sengupta, M. Seandel, N. Geijsen, K. Hochedlinger, Sox2(+) adult stem and progenitor cells are important for tissue regeneration and survival of mice, *Cell Stem Cell* 9 (2011) 317-329.
- [15] L.A. Boyer, T.I. Lee, M.F. Cole, S.E. Johnstone, S.S. Levine, J.P. Zucker, M.G. Guenther, R.M. Kumar, H.L. Murray, R.G. Jenner, D.K. Gifford, D.A.

Melton, R. Jaenisch, R.A. Young, Core transcriptional regulatory circuitry in human embryonic stem cells, *Cell* 122 (2005) 947-956.

[16] Y.H. Loh, Q. Wu, J.L. Chew, V.B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K.Y. Wong, K.W. Sung, C.W. Lee, X.D. Zhao, K.P. Chiu, L. Lipovich, V.A. Kuznetsov, P. Robson, L.W. Stanton, C.L. Wei, Y. Ruan, B. Lim, H.H. Ng, The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells, *Nat Genet* 38 (2006) 431-440.

[17] T.I. Lee, R.G. Jenner, L.A. Boyer, M.G. Guenther, S.S. Levine, R.M. Kumar, B. Chevalier, S.E. Johnstone, M.F. Cole, K. Isono, H. Koseki, T. Fuchikami, K. Abe, H.L. Murray, J.P. Zucker, B. Yuan, G.W. Bell, E. Herbolsheimer, N.M. Hannett, K. Sun, D.T. Odom, A.P. Otte, T.L. Volkert, D.P. Bartel, D.A. Melton, D.K. Gifford, R. Jaenisch, R.A. Young, Control of developmental regulators by Polycomb in human embryonic stem cells, *Cell* 125 (2006) 301-313.

[18] J. Zhang, W.L. Tam, G.Q. Tong, Q. Wu, H.Y. Chan, B.S. Soh, Y. Lou, J. Yang, Y. Ma, L. Chai, H.H. Ng, T. Lufkin, P. Robson, B. Lim, Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1, *Nat Cell Biol* 8 (2006) 1114-1123.

[19] M.F. Cole, S.E. Johnstone, J.J. Newman, M.H. Kagey, R.A. Young, Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells, *Genes Dev* 22 (2008) 746-755.

[20] N. Ivanova, R. Dobrin, R. Lu, I. Kotenko, J. Levorse, C. DeCoste, X. Schafer, Y. Lun, I.R. Lemischka, Dissecting self-renewal in stem cells with RNA interference, *Nature* 442 (2006) 533-538.

- [21] J. Kim, J. Chu, X. Shen, J. Wang, S.H. Orkin, An extended transcriptional network for pluripotency of embryonic stem cells, *Cell* 132 (2008) 1049-1061.
- [22] X. Chen, H. Xu, P. Yuan, F. Fang, M. Huss, V.B. Vega, E. Wong, Y.L. Orlov, W. Zhang, J. Jiang, Y.H. Loh, H.C. Yeo, Z.X. Yeo, V. Narang, K.R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W.K. Sung, N.D. Clarke, C.L. Wei, H.H. Ng, Integration of external signaling pathways with the core transcriptional network in embryonic stem cells, *Cell* 133 (2008) 1106-1117.
- [23] R.D. Kornberg, J.O. Thomas, Chromatin structure; oligomers of the histones, *Science* 184 (1974) 865-868.
- [24] T. Kouzarides, Chromatin modifications and their function, *Cell* 128 (2007) 693-705.
- [25] M.J. Barrero, S. Boue, J.C. Izpisua Belmonte, Epigenetic mechanisms that regulate cell identity, *Cell Stem Cell* 7 (2010) 565-570.
- [26] E.M. Tomazou, A. Meissner, Epigenetic regulation of pluripotency, *Adv Exp Med Biol* 695 (2010) 26-40.
- [27] A. Rada-Iglesias, J. Wysocka, Epigenomics of human embryonic stem cells and induced pluripotent stem cells: insights into pluripotency and implications for disease, *Genome Med* 3 (2011) 36.
- [28] M. Li, G.H. Liu, J.C. Izpisua Belmonte, Navigating the epigenetic landscape of pluripotent stem cells, *Nat Rev Mol Cell Biol* 13 (2012) 524-535.
- [29] S.R. Bhaumik, E. Smith, A. Shilatifard, Covalent modifications of histones during development and disease pathogenesis, *Nat Struct Mol Biol* 14 (2007) 1008-1016.

- [30] B.D. Strahl, C.D. Allis, The language of covalent histone modifications, *Nature* 403 (2000) 41-45.
- [31] E.M. Mendenhall, B.E. Bernstein, Chromatin state maps: new technologies, new insights, *Curr Opin Genet Dev* 18 (2008) 109-115.
- [32] A. Meissner, Epigenetic modifications in pluripotent and differentiated cells, *Nat Biotechnol* 28 (2010) 1079-1088.
- [33] E. Smith, A. Shilatifard, The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes, *Mol Cell* 40 (2010) 689-701.
- [34] B.E. Bernstein, T.S. Mikkelsen, X. Xie, M. Kamal, D.J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S.L. Schreiber, E.S. Lander, A bivalent chromatin structure marks key developmental genes in embryonic stem cells, *Cell* 125 (2006) 315-326.
- [35] M.G. Guenther, S.S. Levine, L.A. Boyer, R. Jaenisch, R.A. Young, A chromatin landmark and transcription initiation at most promoters in human cells, *Cell* 130 (2007) 77-88.
- [36] T.S. Mikkelsen, M. Ku, D.B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T.K. Kim, R.P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E.S. Lander, B.E. Bernstein, Genome-wide maps of chromatin state in pluripotent and lineage-committed cells, *Nature* 448 (2007) 553-560.
- [37] X.D. Zhao, X. Han, J.L. Chew, J. Liu, K.P. Chiu, A. Choo, Y.L. Orlov, W.K. Sung, A. Shahab, V.A. Kuznetsov, G. Bourque, S. Oh, Y. Ruan, H.H. Ng, C.L. Wei, Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals

distinct genomic compartments in human embryonic stem cells, *Cell Stem Cell* 1 (2007) 286-298.

[38] N.L. Vastenhouw, A.F. Schier, Bivalent histone modifications in early embryogenesis, *Curr Opin Cell Biol* 24 (2012) 374-386.

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

[40] J.A. Simon, R.E. Kingston, Mechanisms of polycomb gene silencing: knowns and unknowns, *Nat Rev Mol Cell Biol* 10 (2009) 697-708.

[41] B. Schuettengruber, D. Chourrout, M. Vervoort, B. Leblanc, G. Cavalli, Genome regulation by polycomb and trithorax proteins, *Cell* 128 (2007) 735-745.

[42] L.A. Boyer, K. Plath, J. Zeitlinger, T. Brambrink, L.A. Medeiros, T.I. Lee, S.S. Levine, M. Wernig, A. Tajonar, M.K. Ray, G.W. Bell, A.P. Otte, M. Vidal, D.K. Gifford, R.A. Young, R. Jaenisch, Polycomb complexes repress developmental regulators in murine embryonic stem cells, *Nature* 441 (2006) 349-353.

[43] D. O'Carroll, S. Erhardt, M. Pagani, S.C. Barton, M.A. Surani, T. Jenuwein, The polycomb-group gene *Ezh2* is required for early mouse development, *Mol Cell Biol* 21 (2001) 4330-4336.

[44] J. Wang, J. Mager, E. Schnedier, T. Magnuson, The mouse PcG gene *eed* is required for Hox gene repression and extraembryonic development, *Mamm Genome* 13 (2002) 493-503.

- [45] N.S. Christophersen, K. Helin, Epigenetic control of embryonic stem cell fate, *J Exp Med* 207 (2010) 2287-2295.
- [46] J. Ding, H. Xu, F. Faiola, A. Ma'ayan, J. Wang, Oct4 links multiple epigenetic pathways to the pluripotency network, *Cell Res* 22 (2012) 155-167.
- [47] S.H. Hong, S. Rampalli, J.B. Lee, J. McNicol, T. Collins, J.S. Draper, M. Bhatia, Cell fate potential of human pluripotent stem cells is encoded by histone modifications, *Cell Stem Cell* 9 (2011) 24-36.
- [48] B. Wen, H. Wu, Y. Shinkai, R.A. Irizarry, A.P. Feinberg, Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells, *Nat Genet* 41 (2009) 246-250.
- [49] S. Bilodeau, M.H. Kagey, G.M. Frampton, P.B. Rahl, R.A. Young, SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state, *Genes Dev* 23 (2009) 2484-2489.
- [50] M. Tachibana, K. Sugimoto, M. Nozaki, J. Ueda, T. Ohta, M. Ohki, M. Fukuda, N. Takeda, H. Niida, H. Kato, Y. Shinkai, G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis, *Genes Dev* 16 (2002) 1779-1791.
- [51] O. Alder, F. Laval, A. Helness, E. Brookes, S. Pinho, A. Chandrashekan, P. Arnaud, A. Pombo, L. O'Neill, V. Azuara, Ring1B and Suv39h1 delineate distinct chromatin states at bivalent genes during early mouse lineage commitment, *Development* 137 (2010) 2483-2492.

- [52] T.G. Fazio, J.T. Huff, B. Panning, An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity, *Cell* 134 (2008) 162-174.
- [53] A. Visel, M.J. Blow, Z. Li, T. Zhang, J.A. Akiyama, A. Holt, I. Plajzer-Frick, M. Shoukry, C. Wright, F. Chen, V. Afzal, B. Ren, E.M. Rubin, L.A. Pennacchio, ChIP-seq accurately predicts tissue-specific activity of enhancers, *Nature* 457 (2009) 854-858.
- [54] A. Rada-Iglesias, R. Bajpai, T. Swigut, S.A. Brugmann, R.A. Flynn, J. Wysocka, A unique chromatin signature uncovers early developmental enhancers in humans, *Nature* 470 (2011) 279-283.
- [55] M.P. Creighton, A.W. Cheng, G.G. Welstead, T. Kooistra, B.W. Carey, E.J. Steine, J. Hanna, M.A. Lodato, G.M. Frampton, P.A. Sharp, L.A. Boyer, R.A. Young, R. Jaenisch, Histone H3K27ac separates active from poised enhancers and predicts developmental state, *Proc Natl Acad Sci U S A* 107 (2010) 21931-21936.
- [56] M. Bulger, Maps for changing landscapes: viewing epigenomic signatures through differentiation, *Cell Stem Cell* 11 (2012) 581-582.
- [57] C. Buecker, J. Wysocka, Enhancers as information integration hubs in development: lessons from genomics, *Trends Genet* 28 (2012) 276-284.
- [58] C.T. Ong, V.G. Corces, Enhancers: emerging roles in cell fate specification, *EMBO Rep* 13 (2012) 423-430.
- [59] A.T. Willingham, T.R. Gingeras, TUF love for "junk" DNA, *Cell* 125 (2006) 1215-1220.

- [60] A. Jacquier, The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs, *Nat Rev Genet* 10 (2009) 833-844.
- [61] C.A. Brosnan, O. Voinnet, The long and the short of noncoding RNAs, *Curr Opin Cell Biol* 21 (2009) 416-425.
- [62] A. Pauli, J.L. Rinn, A.F. Schier, Non-coding RNAs as regulators of embryogenesis, *Nat Rev Genet* 12 (2011) 136-149.
- [63] D.P. Bartel, MicroRNAs: target recognition and regulatory functions, *Cell* 136 (2009) 215-233.
- [64] M.S. Ebert, P.A. Sharp, Roles for microRNAs in conferring robustness to biological processes, *Cell* 149 (2012) 515-524.
- [65] Y. Wang, R. Medvid, C. Melton, R. Jaenisch, R. Blelloch, DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal, *Nat Genet* 39 (2007) 380-385.
- [66] E. Bernstein, S.Y. Kim, M.A. Carmell, E.P. Murchison, H. Alcorn, M.Z. Li, A.A. Mills, S.J. Elledge, K.V. Anderson, G.J. Hannon, Dicer is essential for mouse development, *Nat Genet* 35 (2003) 215-217.
- [67] A. Marson, S.S. Levine, M.F. Cole, G.M. Frampton, T. Brambrink, S. Johnstone, M.G. Guenther, W.K. Johnston, M. Wernig, J. Newman, J.M. Calabrese, L.M. Dennis, T.L. Volkert, S. Gupta, J. Love, N. Hannett, P.A. Sharp, D.P. Bartel, R. Jaenisch, R.A. Young, Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells, *Cell* 134 (2008) 521-533.

- [68] Y. Wang, S. Baskerville, A. Shenoy, J.E. Babiarz, L. Baehner, R. Blelloch, Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation, *Nat Genet* 40 (2008) 1478-1483.
- [69] A.J. Giraldez, Y. Mishima, J. Rihel, R.J. Grocock, S. Van Dongen, K. Inoue, A.J. Enright, A.F. Schier, Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs, *Science* 312 (2006) 75-79.
- [70] Z. Lichner, E. Pall, A. Kerekes, E. Pallinger, P. Maraghechi, Z. Bosze, E. Gocza, The miR-290-295 cluster promotes pluripotency maintenance by regulating cell cycle phase distribution in mouse embryonic stem cells, *Differentiation* 81 (2011) 11-24.
- [71] Y. Tay, J. Zhang, A.M. Thomson, B. Lim, I. Rigoutsos, MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation, *Nature* 455 (2008) 1124-1128.
- [72] N. Xu, T. Papagiannakopoulos, G. Pan, J.A. Thomson, K.S. Kosik, MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells, *Cell* 137 (2009) 647-658.
- [73] M. Guttman, I. Amit, M. Garber, C. French, M.F. Lin, D. Feldser, M. Huarte, O. Zuk, B.W. Carey, J.P. Cassady, M.N. Cabili, R. Jaenisch, T.S. Mikkelsen, T. Jacks, N. Hacohen, B.E. Bernstein, M. Kellis, A. Regev, J.L. Rinn, E.S. Lander, Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals, *Nature* 458 (2009) 223-227.
- [74] M.N. Cabili, C. Trapnell, L. Goff, M. Koziol, B. Tazon-Vega, A. Regev, J.L. Rinn, Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses, *Genes Dev* 25 (2011) 1915-1927.

- [75] J.L. Rinn, H.Y. Chang, Genome regulation by long noncoding RNAs, *Annu Rev Biochem* 81 (2012) 145-166.
- [76] S.Y. Ng, L.W. Stanton, Long non-coding RNAs in stem cell pluripotency, *Wiley Interdiscip Rev RNA* 4 (2013) 121-128.
- [77] M. Guttman, J. Donaghey, B.W. Carey, M. Garber, J.K. Grenier, G. Munson, G. Young, A.B. Lucas, R. Ach, L. Bruhn, X. Yang, I. Amit, A. Meissner, A. Regev, J.L. Rinn, D.E. Root, E.S. Lander, lincRNAs act in the circuitry controlling pluripotency and differentiation, *Nature* 477 (2011) 295-300.
- [78] A.M. Khalil, M. Guttman, M. Huarte, M. Garber, A. Raj, D. Rivea Morales, K. Thomas, A. Presser, B.E. Bernstein, A. van Oudenaarden, A. Regev, E.S. Lander, J.L. Rinn, Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression, *Proc Natl Acad Sci U S A* 106 (2009) 11667-11672.
- [79] M. Guttman, J.L. Rinn, Modular regulatory principles of large non-coding RNAs, *Nature* 482 (2012) 339-346.
- [80] S.Y. Ng, R. Johnson, L.W. Stanton, Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors, *EMBO J* 31 (2012) 522-533.
- [81] R. Bianco, D. Melisi, F. Ciardiello, G. Tortora, Key cancer cell signal transduction pathways as therapeutic targets, *Eur J Cancer* 42 (2006) 290-294.
- [82] M.F. Pera, P.P. Tam, Extrinsic regulation of pluripotent stem cells, *Nature* 465 (2010) 713-720.
- [83] H.H. Ng, M.A. Surani, The transcriptional and signalling networks of pluripotency, *Nat Cell Biol* 13 (2011) 490-496.

- [84] L. Daheron, S.L. Opitz, H. Zaehres, M.W. Lensch, P.W. Andrews, J. Itskovitz-Eldor, G.Q. Daley, LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells, *Stem Cells* 22 (2004) 770-778.
- [85] R.H. Xu, X. Chen, D.S. Li, R. Li, G.C. Addicks, C. Glennon, T.P. Zwaka, J.A. Thomson, BMP4 initiates human embryonic stem cell differentiation to trophoblast, *Nat Biotechnol* 20 (2002) 1261-1264.
- [86] D. James, A.J. Levine, D. Besser, A. Hemmati-Brivanlou, TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells, *Development* 132 (2005) 1273-1282.
- [87] L. Vallier, D. Reynolds, R.A. Pedersen, Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway, *Dev Biol* 275 (2004) 403-421.
- [88] L. Vallier, M. Alexander, R.A. Pedersen, Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells, *J Cell Sci* 118 (2005) 4495-4509.
- [89] R.H. Xu, T.L. Sampsel-Barron, F. Gu, S. Root, R.M. Peck, G. Pan, J. Yu, J. Antosiewicz-Bourget, S. Tian, R. Stewart, J.A. Thomson, NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs, *Cell Stem Cell* 3 (2008) 196-206.
- [90] A.C. Mullen, D.A. Orlando, J.J. Newman, J. Loven, R.M. Kumar, S. Bilodeau, J. Reddy, M.G. Guenther, R.P. DeKoter, R.A. Young, Master transcription factors determine cell-type-specific responses to TGF-beta signaling, *Cell* 147 (2011) 565-576.

- [91] G. Dravid, Z. Ye, H. Hammond, G. Chen, A. Pyle, P. Donovan, X. Yu, L. Cheng, Defining the role of Wnt/beta-catenin signaling in the survival, proliferation, and self-renewal of human embryonic stem cells, *Stem Cells* 23 (2005) 1489-1501.
- [92] L.J. Gudas, Retinoids and vertebrate development, *J Biol Chem* 269 (1994) 15399-15402.
- [93] M. Rhinn, P. Dolle, Retinoic acid signalling during development, *Development* 139 (2012) 843-858.
- [94] K. Niederreither, V. Subbarayan, P. Dolle, P. Chambon, Embryonic retinoic acid synthesis is essential for early mouse post-implantation development, *Nat Genet* 21 (1999) 444-448.
- [95] T. Pennimpede, D.A. Cameron, G.A. MacLean, H. Li, S. Abu-Abed, M. Petkovich, The role of CYP26 enzymes in defining appropriate retinoic acid exposure during embryogenesis, *Birth Defects Res A Clin Mol Teratol* 88 (2010) 883-894.
- [96] L.J. Gudas, J.A. Wagner, Retinoids regulate stem cell differentiation, *J Cell Physiol* 226 (2011) 322-330.
- [97] M. Bibel, J. Richter, K. Schrenk, K.L. Tucker, V. Staiger, M. Korte, M. Goetz, Y.A. Barde, Differentiation of mouse embryonic stem cells into a defined neuronal lineage, *Nat Neurosci* 7 (2004) 1003-1009.
- [98] E. Martinez-Ceballos, L.J. Gudas, *Hoxa1* is required for the retinoic acid-induced differentiation of embryonic stem cells into neurons, *J Neurosci Res* 86 (2008) 2809-2819.

- [99] K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors, *Cell* 126 (2006) 663-676.
- [100] J. Yu, M.A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J.L. Frane, S. Tian, J. Nie, G.A. Jonsdottir, V. Ruotti, R. Stewart, Slukvin, II, J.A. Thomson, Induced pluripotent stem cell lines derived from human somatic cells, *Science* 318 (2007) 1917-1920.
- [101] M. Wernig, A. Meissner, J.P. Cassady, R. Jaenisch, c-Myc is dispensable for direct reprogramming of mouse fibroblasts, *Cell Stem Cell* 2 (2008) 10-12.
- [102] M. Nakagawa, M. Koyanagi, K. Tanabe, K. Takahashi, T. Ichisaka, T. Aoi, K. Okita, Y. Mochiduki, N. Takizawa, S. Yamanaka, Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts, *Nat Biotechnol* 26 (2008) 101-106.
- [103] S. Loewer, M.N. Cabili, M. Guttman, Y.H. Loh, K. Thomas, I.H. Park, M. Garber, M. Curran, T. Onder, S. Agarwal, P.D. Manos, S. Datta, E.S. Lander, T.M. Schlaeger, G.Q. Daley, J.L. Rinn, Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells, *Nat Genet* 42 (2010) 1113-1117.
- [104] R.L. Judson, J.E. Babiarz, M. Venere, R. Blelloch, Embryonic stem cell-specific microRNAs promote induced pluripotency, *Nat Biotechnol* 27 (2009) 459-461.
- [105] R. Sridharan, J. Tchieu, M.J. Mason, R. Yachechko, E. Kuoy, S. Horvath, Q. Zhou, K. Plath, Role of the murine reprogramming factors in the induction of pluripotency, *Cell* 136 (2009) 364-377.

- [106] A. Soufi, G. Donahue, K.S. Zaret, Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome, *Cell* 151 (2012) 994-1004.
- [107] M.G. Guenther, G.M. Frampton, F. Soldner, D. Hockemeyer, M. Mitalipova, R. Jaenisch, R.A. Young, Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells, *Cell Stem Cell* 7 (2010) 249-257.
- [108] H. Zhu, M.W. Lensch, P. Cahan, G.Q. Daley, Investigating monogenic and complex diseases with pluripotent stem cells, *Nat Rev Genet* 12 (2011) 266-275.
- [109] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, *Cell* 144 (2011) 646-674.
- [110] J. Kim, S.H. Orkin, Embryonic stem cell-specific signatures in cancer: insights into genomic regulatory networks and implications for medicine, *Genome Med* 3 (2011) 75.
- [111] D.J. Wong, H. Liu, T.W. Ridky, D. Cassarino, E. Segal, H.Y. Chang, Module map of stem cell genes guides creation of epithelial cancer stem cells, *Cell Stem Cell* 2 (2008) 333-344.
- [112] I. Ben-Porath, M.W. Thomson, V.J. Carey, R. Ge, G.W. Bell, A. Regev, R.A. Weinberg, An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors, *Nat Genet* 40 (2008) 499-507.
- [113] J. Kim, A.J. Woo, J. Chu, J.W. Snow, Y. Fujiwara, C.G. Kim, A.B. Cantor, S.H. Orkin, A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs, *Cell* 143 (2010) 313-324.

[114] R. Beroukhim, C.H. Mermel, D. Porter, G. Wei, S. Raychaudhuri, J. Donovan, J. Barretina, J.S. Boehm, J. Dobson, M. Urashima, K.T. Mc Henry, R.M. Pinchback, A.H. Ligon, Y.J. Cho, L. Haery, H. Greulich, M. Reich, W. Winckler, M.S. Lawrence, B.A. Weir, K.E. Tanaka, D.Y. Chiang, A.J. Bass, A. Loo, C. Hoffman, J. Prensner, T. Liefeld, Q. Gao, D. Yecies, S. Signoretti, E. Maher, F.J. Kaye, H. Sasaki, J.E. Tepper, J.A. Fletcher, J. Taberero, J. Baselga, M.S. Tsao, F. Demichelis, M.A. Rubin, P.A. Janne, M.J. Daly, C. Nucera, R.L. Levine, B.L. Ebert, S. Gabriel, A.K. Rustgi, C.R. Antonescu, M. Ladanyi, A. Letai, L.A. Garraway, M. Loda, D.G. Beer, L.D. True, A. Okamoto, S.L. Pomeroy, S. Singer, T.R. Golub, E.S. Lander, G. Getz, W.R. Sellers, M. Meyerson, The landscape of somatic copy-number alteration across human cancers, *Nature* 463 (2010) 899-905.

[115] K.H. Vousden, C. Prives, Blinded by the Light: The Growing Complexity of p53, *Cell* 137 (2009) 413-431.

[116] B.T. Spike, G.M. Wahl, p53, Stem Cells, and Reprogramming: Tumor Suppression beyond Guarding the Genome, *Genes Cancer* 2 (2011) 404-419.

[117] H. Hong, K. Takahashi, T. Ichisaka, T. Aoi, O. Kanagawa, M. Nakagawa, K. Okita, S. Yamanaka, Suppression of induced pluripotent stem cell generation by the p53-p21 pathway, *Nature* 460 (2009) 1132-1135.

[118] T. Kawamura, J. Suzuki, Y.V. Wang, S. Menendez, L.B. Morera, A. Raya, G.M. Wahl, J.C. Belmonte, Linking the p53 tumour suppressor pathway to somatic cell reprogramming, *Nature* 460 (2009) 1140-1144.

- [119] H. Li, M. Collado, A. Villasante, K. Strati, S. Ortega, M. Canamero, M.A. Blasco, M. Serrano, The Ink4/Arf locus is a barrier for iPS cell reprogramming, *Nature* 460 (2009) 1136-1139.
- [120] R.M. Marion, K. Strati, H. Li, M. Murga, R. Blanco, S. Ortega, O. Fernandez-Capetillo, M. Serrano, M.A. Blasco, A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity, *Nature* 460 (2009) 1149-1153.
- [121] J. Utikal, J.M. Polo, M. Stadtfeld, N. Maherali, W. Kulalert, R.M. Walsh, A. Khalil, J.G. Rheinwald, K. Hochedlinger, Immortalization eliminates a roadblock during cellular reprogramming into iPS cells, *Nature* 460 (2009) 1145-1148.
- [122] V. Krizhanovsky, S.W. Lowe, Stem cells: The promises and perils of p53, *Nature* 460 (2009) 1085-1086.
- [123] A.M. Puzio-Kuter, A.J. Levine, Stem cell biology meets p53, *Nat Biotechnol* 27 (2009) 914-915.
- [124] S. Menendez, S. Camus, J.C. Izpisua Belmonte, p53: guardian of reprogramming, *Cell Cycle* 9 (2010) 3887-3891.
- [125] H. Mizuno, B.T. Spike, G.M. Wahl, A.J. Levine, Inactivation of p53 in breast cancers correlates with stem cell transcriptional signatures, *Proc Natl Acad Sci U S A* 107 (2010) 22745-22750.
- [126] E.K. Markert, H. Mizuno, A. Vazquez, A.J. Levine, Molecular classification of prostate cancer using curated expression signatures, *Proc Natl Acad Sci U S A* 108 (2011) 21276-21281.

- [127] J.A. Fagin, K. Matsuo, A. Karmakar, D.L. Chen, S.H. Tang, H.P. Koeffler, High prevalence of mutations of the p53 gene in poorly differentiated human thyroid carcinomas, *J Clin Invest* 91 (1993) 179-184.
- [128] M.R. Junttila, A.N. Karnezis, D. Garcia, F. Madriles, R.M. Kortlever, F. Rostker, L. Brown Swigart, D.M. Pham, Y. Seo, G.I. Evan, C.P. Martins, Selective activation of p53-mediated tumour suppression in high-grade tumours, *Nature* 468 (2010) 567-571.
- [129] Z. Zhao, J. Zuber, E. Diaz-Flores, L. Lintault, S.C. Kogan, K. Shannon, S.W. Lowe, p53 loss promotes acute myeloid leukemia by enabling aberrant self-renewal, *Genes Dev* 24 (2010) 1389-1402.
- [130] E. Feinstein, R.P. Gale, J. Reed, E. Canaani, Expression of the normal p53 gene induces differentiation of K562 cells, *Oncogene* 7 (1992) 1853-1857.
- [131] T. Lin, C. Chao, S. Saito, S.J. Mazur, M.E. Murphy, E. Appella, Y. Xu, p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression, *Nat Cell Biol* 7 (2005) 165-171.
- [132] K.H. Lee, M. Li, A.M. Michalowski, X. Zhang, H. Liao, L. Chen, Y. Xu, X. Wu, J. Huang, A genomewide study identifies the Wnt signaling pathway as a major target of p53 in murine embryonic stem cells, *Proc Natl Acad Sci U S A* 107 (2010) 69-74.
- [133] M. Li, Y. He, W. Dubois, X. Wu, J. Shi, J. Huang, Distinct regulatory mechanisms and functions for p53-activated and p53-repressed DNA damage response genes in embryonic stem cells, *Mol Cell* 46 (2012) 30-42.

- [134] G. Kunarso, N.Y. Chia, J. Jeyakani, C. Hwang, X. Lu, Y.S. Chan, H.H. Ng, G. Bourque, Transposable elements have rewired the core regulatory network of human embryonic stem cells, *Nat Genet* 42 (2010) 631-634.
- [135] D.T. Odom, R.D. Dowell, E.S. Jacobsen, W. Gordon, T.W. Danford, K.D. MacIsaac, P.A. Rolfe, C.M. Conboy, D.K. Gifford, E. Fraenkel, Tissue-specific transcriptional regulation has diverged significantly between human and mouse, *Nat Genet* 39 (2007) 730-732.
- [136] A. De Los Angeles, Y.H. Loh, P.J. Tesar, G.Q. Daley, Accessing naive human pluripotency, *Curr Opin Genet Dev* 22 (2012) 272-282.
- [137] P.J. Tesar, J.G. Chenoweth, F.A. Brook, T.J. Davies, E.P. Evans, D.L. Mack, R.L. Gardner, R.D. McKay, New cell lines from mouse epiblast share defining features with human embryonic stem cells, *Nature* 448 (2007) 196-199.
- [138] T. Kunath, Primed for pluripotency, *Cell Stem Cell* 8 (2011) 241-242.
- [139] K.A. Becker, J.L. Stein, J.B. Lian, A.J. van Wijnen, G.S. Stein, Establishment of histone gene regulation and cell cycle checkpoint control in human embryonic stem cells, *J Cell Physiol* 210 (2007) 517-526.
- [140] A.K. Jain, K. Allton, M. Iacovino, E. Mahen, R.J. Milczarek, T.P. Zwaka, M. Kyba, M.C. Barton, p53 regulates cell cycle and microRNAs to promote differentiation of human embryonic stem cells, *PLoS biology* 10 (2012) e1001268.
- [141] C. Hindley, A. Philpott, The cell cycle and pluripotency, *Biochem J* 451 (2013) 135-143.
- [142] R.C. Hardison, J. Taylor, Genomic approaches towards finding cis-regulatory modules in animals, *Nat Rev Genet* 13 (2012) 469-483.

- [143] B. Ren, F. Robert, J.J. Wyrick, O. Aparicio, E.G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T.L. Volkert, C.J. Wilson, S.P. Bell, R.A. Young, Genome-wide location and function of DNA binding proteins, *Science* 290 (2000) 2306-2309.
- [144] T.C. Mockler, S. Chan, A. Sundaresan, H. Chen, S.E. Jacobsen, J.R. Ecker, Applications of DNA tiling arrays for whole-genome analysis, *Genomics* 85 (2005) 1-15.
- [145] D.S. Johnson, A. Mortazavi, R.M. Myers, B. Wold, Genome-wide mapping of in vivo protein-DNA interactions, *Science* 316 (2007) 1497-1502.
- [146] P.J. Park, ChIP-seq: advantages and challenges of a maturing technology, *Nat Rev Genet* 10 (2009) 669-680.
- [147] T.S. Furey, ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions, *Nat Rev Genet* 13 (2012) 840-852.
- [148] Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, X.S. Liu, Model-based analysis of ChIP-Seq (MACS), *Genome Biol* 9 (2008) R137.
- [149] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics* 26 (2010) 841-842.
- [150] A. Siepel, G. Bejerano, J.S. Pedersen, A.S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L.W. Hillier, S. Richards, G.M. Weinstock, R.K. Wilson, R.A. Gibbs, W.J. Kent, W. Miller, D. Haussler, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes, *Genome Res* 15 (2005) 1034-1050.

- [151] H. Shin, T. Liu, A.K. Manrai, X.S. Liu, CEAS: cis-regulatory element annotation system, *Bioinformatics* 25 (2009) 2605-2606.
- [152] P. Machanick, T.L. Bailey, MEME-ChIP: motif analysis of large DNA datasets, *Bioinformatics* 27 (2011) 1696-1697.
- [153] T.L. Bailey, C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in biopolymers, *Proc Int Conf Intell Syst Mol Biol* 2 (1994) 28-36.
- [154] T.L. Bailey, DREME: motif discovery in transcription factor ChIP-seq data, *Bioinformatics* 27 (2011) 1653-1659.
- [155] R.C. McLeay, T.L. Bailey, Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data, *BMC Bioinformatics* 11 (2010) 165.
- [156] H.H. He, C.A. Meyer, H. Shin, S.T. Bailey, G. Wei, Q. Wang, Y. Zhang, K. Xu, M. Ni, M. Lupien, P. Mieczkowski, J.D. Lieb, K. Zhao, M. Brown, X.S. Liu, Nucleosome dynamics define transcriptional enhancers, *Nat Genet* 42 (2010) 343-347.
- [157] D. Karolchik, A.S. Hinrichs, T.S. Furey, K.M. Roskin, C.W. Sugnet, D. Haussler, W.J. Kent, The UCSC Table Browser data retrieval tool, *Nucleic Acids Res* 32 (2004) D493-496.
- [158] T.J. Klisch, Y. Xi, A. Flora, L. Wang, W. Li, H.Y. Zoghbi, In vivo Atoh1 targetome reveals how a proneural transcription factor regulates cerebellar development, *Proc Natl Acad Sci U S A* 108 (2011) 3288-3293.
- [159] W. Huang da, B.T. Sherman, Q. Tan, J.R. Collins, W.G. Alvord, J. Roayaei, R. Stephens, M.W. Baseler, H.C. Lane, R.A. Lempicki, The DAVID

Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists, *Genome Biol* 8 (2007) R183.

[160] S. Povey, R. Lovering, E. Bruford, M. Wright, M. Lush, H. Wain, The HUGO Gene Nomenclature Committee (HGNC), *Hum Genet* 109 (2001) 678-680.

[161] C.Y. McLean, D. Bristor, M. Hiller, S.L. Clarke, B.T. Schaar, C.B. Lowe, A.M. Wenger, G. Bejerano, GREAT improves functional interpretation of cis-regulatory regions, *Nat Biotechnol* 28 (2010) 495-501.

[162] R. Edgar, M. Domrachev, A.E. Lash, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository, *Nucleic Acids Res* 30 (2002) 207-210.

[163] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, M.A. Marra, Circos: an information aesthetic for comparative genomics, *Genome Res* 19 (2009) 1639-1645.

[164] R.M. Myers, J. Stamatoyannopoulos, M. Snyder, I. Dunham, R.C. Hardison, B.E. Bernstein, T.R. Gingeras, W.J. Kent, E. Birney, B. Wold, G.E. Crawford, A user's guide to the encyclopedia of DNA elements (ENCODE), *PLoS Biol* 9 (2011) e1001046.

[165] B. Qin, M. Zhou, Y. Ge, L. Taing, T. Liu, Q. Wang, S. Wang, J. Chen, L. Shen, X. Duan, S. Hu, W. Li, H. Long, Y. Zhang, X.S. Liu, CistromeMap: a knowledgebase and web server for ChIP-Seq and DNase-Seq studies in mouse and human, *Bioinformatics* 28 (2012) 1411-1412.

- [166] R.A. Irizarry, B. Hobbs, F. Collin, Y.D. Beazer-Barclay, K.J. Antonellis, U. Scherf, T.P. Speed, Exploration, normalization, and summaries of high density oligonucleotide array probe level data, *Biostatistics* 4 (2003) 249-264.
- [167] G.K. Smyth, Linear models and empirical bayes methods for assessing differential expression in microarray experiments, *Stat Appl Genet Mol Biol* 3 (2004) Article3.
- [168] R. Beckerman, C. Prives, Transcriptional regulation by p53, *Cold Spring Harb Perspect Biol* 2 (2010) a000935.
- [169] O. Hobert, H. Westphal, Functions of LIM-homeobox genes, *Trends Genet* 16 (2000) 75-83.
- [170] O.J. Lehmann, J.C. Sowden, P. Carlsson, T. Jordan, S.S. Bhattacharya, Fox's in development and disease, *Trends Genet* 19 (2003) 339-344.
- [171] G.E. Schepers, R.D. Teasdale, P. Koopman, Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families, *Dev Cell* 3 (2002) 167-170.
- [172] C.S. Merzdorf, Emerging roles for zic genes in early development, *Dev Dyn* 236 (2007) 922-940.
- [173] X. Zhang, S. Yalcin, D.F. Lee, T.Y. Yeh, S.M. Lee, J. Su, S.K. Mungamuri, P. Rimmel, M. Kennedy, R. Sellers, M. Landthaler, T. Tuschl, N.W. Chi, I. Lemischka, G. Keller, S. Ghaffari, FOXO1 is an essential regulator of pluripotency in human embryonic stem cells, *Nat Cell Biol* 13 (2011) 1092-1099.
- [174] N. Matsumoto, A. Kubo, H. Liu, K. Akita, F. Laub, F. Ramirez, G. Keller, S.L. Friedman, Developmental regulation of yolk sac hematopoiesis by Kruppel-like factor 6, *Blood* 107 (2006) 1357-1365.

- [175] S. Chang, T.A. McKinsey, C.L. Zhang, J.A. Richardson, J.A. Hill, E.N. Olson, Histone deacetylases 5 and 9 govern responsiveness of the heart to a subset of stress signals and play redundant roles in heart development, *Mol Cell Biol* 24 (2004) 8467-8476.
- [176] J. Karlseder, L. Kachatrian, H. Takai, K. Mercer, S. Hingorani, T. Jacks, T. de Lange, Targeted deletion reveals an essential function for the telomere length regulator Trf1, *Mol Cell Biol* 23 (2003) 6533-6541.
- [177] A.J. Levine, M. Oren, The first 30 years of p53: growing ever more complex, *Nat Rev Cancer* 9 (2009) 749-758.
- [178] D. Menendez, A. Inga, M.A. Resnick, The expanding universe of p53 targets, *Nat Rev Cancer* 9 (2009) 724-737.
- [179] K.H. Vousden, K.M. Ryan, p53 and metabolism, *Nat Rev Cancer* 9 (2009) 691-700.
- [180] A.J. Levine, A.M. Puzio-Kuter, The control of the metabolic switch in cancers by oncogenes and tumor suppressor genes, *Science* 330 (2010) 1340-1344.
- [181] J.C. Pearson, D. Lemons, W. McGinnis, Modulating Hox gene functions during animal body patterning, *Nat Rev Genet* 6 (2005) 893-904.
- [182] L. Morey, G. Pascual, L. Cozzuto, G. Roma, A. Wutz, S.A. Benitah, L. Di Croce, Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells, *Cell Stem Cell* 10 (2012) 47-62.
- [183] C. Showell, O. Binder, F.L. Conlon, T-box genes in early embryogenesis, *Dev Dyn* 229 (2004) 201-218.

- [184] A. Fischer, J. Klattig, B. Kneitz, H. Diez, M. Maier, B. Holtmann, C. Englert, M. Gessler, Hey basic helix-loop-helix transcription factors are repressors of GATA4 and GATA6 and restrict expression of the GATA target gene ANF in fetal hearts, *Mol Cell Biol* 25 (2005) 8960-8970.
- [185] D.M. Martin, Chromatin remodeling in development and disease: focus on CHD7, *PLoS Genet* 6 (2010) e1001010.
- [186] P. van der Stoop, E.A. Boutsma, D. Hulsman, S. Noback, M. Heimerikx, R.M. Kerkhoven, J.W. Voncken, L.F. Wessels, M. van Lohuizen, Ubiquitin E3 ligase Ring1b/Rnf2 of polycomb repressive complex 1 contributes to stable maintenance of mouse embryonic stem cells, *PLoS One* 3 (2008) e2235.
- [187] Y.H. Yu, G.Y. Chiou, P.I. Huang, W.L. Lo, C.Y. Wang, K.H. Lu, C.C. Yu, G. Alterovitz, W.C. Huang, J.F. Lo, H.S. Hsu, S.H. Chiou, Network biology of tumor stem-like cells identified a regulatory role of CBX5 in lung cancer, *Sci Rep* 2 (2012) 584.
- [188] P. Maye, S. Becker, H. Siemen, J. Thorne, N. Byrd, J. Carpentino, L. Grabel, Hedgehog signaling is required for the differentiation of ES cells into neurectoderm, *Dev Biol* 265 (2004) 276-290.
- [189] J.E. Dixon, E. Dick, D. Rajamohan, K.M. Shakesheff, C. Denning, Directed differentiation of human embryonic stem cells to interrogate the cardiac gene regulatory network, *Mol Ther* 19 (2011) 1695-1703.
- [190] C.D. Folmes, P.P. Dzeja, T.J. Nelson, A. Terzic, Metabolic plasticity in stem cell homeostasis and differentiation, *Cell Stem Cell* 11 (2012) 596-606.
- [191] J. Yang, R.A. Weinberg, Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis, *Dev Cell* 14 (2008) 818-829.

- [192] F.J. Najm, J.G. Chenoweth, P.D. Anderson, J.H. Nadeau, R.W. Redline, R.D. McKay, P.J. Tesar, Isolation of epiblast stem cells from preimplantation mouse embryos, *Cell Stem Cell* 8 (2011) 318-325.
- [193] I.G. Romero, I. Ruvinsky, Y. Gilad, Comparative studies of gene expression and the evolution of gene regulation, *Nat Rev Genet* 13 (2012) 505-516.
- [194] R.D. Dowell, Transcription factor binding variation in the evolution of gene regulation, *Trends Genet* 26 (2010) 468-475.
- [195] M.D. Wilson, D.T. Odom, Evolution of transcriptional control in mammals, *Curr Opin Genet Dev* 19 (2009) 579-585.
- [196] M. O'Bleness, V.B. Searles, A. Varki, P. Gagneux, J.M. Sikela, Evolution of genetic and genomic features unique to the human lineage, *Nat Rev Genet* 13 (2012) 853-866.
- [197] H. Song, S.K. Chung, Y. Xu, Modeling disease in human ESCs using an efficient BAC-based homologous recombination system, *Cell Stem Cell* 6 (2010) 80-89.
- [198] V. Dotsch, F. Bernassola, D. Coutandin, E. Candi, G. Melino, p63 and p73, the ancestors of p53, *Cold Spring Harb Perspect Biol* 2 (2010) a004887.
- [199] F. Talos, A. Nemajerova, E.R. Flores, O. Petrenko, U.M. Moll, p73 suppresses polyploidy and aneuploidy in the absence of functional p53, *Mol Cell* 27 (2007) 647-659.
- [200] U.M. Moll, N. Slade, p63 and p73: roles in development and tumor formation, *Mol Cancer Res* 2 (2004) 371-386.

- [201] M. Senoo, F. Pinto, C.P. Crum, F. McKeon, p63 Is essential for the proliferative potential of stem cells in stratified epithelia, *Cell* 129 (2007) 523-536.
- [202] M. Agostini, P. Tucci, R. Killick, E. Candi, B.S. Sayan, P. Rivetti di Val Cervo, P. Nicotera, F. McKeon, R.A. Knight, T.W. Mak, G. Melino, Neuronal differentiation by TAp73 is mediated by microRNA-34a regulation of synaptic protein targets, *Proc Natl Acad Sci U S A* 108 (2011) 21093-21098.
- [203] A. Yang, Z. Zhu, A. Kettenbach, P. Kapranov, F. McKeon, T.R. Gingeras, K. Struhl, Genome-wide mapping indicates that p73 and p63 co-occupy target sites and have similar dna-binding profiles in vivo, *PLoS One* 5 (2010) e11572.

VITA

Kadir Caner Akdemir was born on April 5th 1984, in Siirt Turkey. He is an alumnus of Yeditepe University in Istanbul, Turkey where he obtained Bachelor of Science degree in Computer Science (2007) and in Genetics (2008). Upon completion, he was admitted to Graduate School of Biomedical Sciences at the University of Texas Health Science Center at Houston in August 2008.

Permanent Address:

Fatih Cad. No:80/1

Yalova/TURKEY