

**MODULATION OF TRANSCRIPTION FACTOR CHROMATIN ASSOCIATION
AND GENE TRANSCRIPTION PROGRAM IN EMBRYONIC STEM CELL
AND TRIPLE NEGATIVE BREAST CANCER BY
POLY (ADP-RIBOSE) POLYMERASE 1**

APPROVED BY SUPERVISORY COMMITTEE

NOTE: The top line is for the Supervising Professor's name. There should be as many lines as there are members of the committee. All signatures must be original and in ink. Adjust "Approved by Supervisory Committee" line upward if the committee list is very large.

W. Lee Kraus, Ph.D

Chun-li Zhang, Ph.D

Sean Morrison, Ph.D

Bing Li, Ph.D

Taekyung Kim, Ph.D

DEDICATION

To my family, for their unconditional support and love

**MODULATION OF TRANSCRIPTION FACTOR CHROMATIN ASSOCIATION
AND GENE TRANSCRIPTION PROGRAM IN EMBRYONIC STEM CELL
AND TRIPLE NEGATIVE BREAST CANCER BY
POLY (ADP-RIBOSE) POLYMERASE 1**

By

Ziying Liu

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

May, 2016

ACKNOWLEDGEMENT

First, I would like to express my sincere gratitude to my mentor, Dr. W. Lee Kraus, for the support and guidance throughout my graduate program. He is a great mentor, who provided me a tremendous amount of scientific insights. He is always encouraging me to try new ideas, exploring the unknown territories of life sciences. I learnt from him the critical things for becoming a good scientist: work hard, be patient, be creative, and be brave.

I would also like to thank my committee members: Dr. Chun-li Zhang for following my research with interest and for his expert advice throughout my graduate program, Dr. Sean Morrison for great advice and support on my projects, Dr. Bing Li and Dr. Taekyung Kim for providing scientific insights and encouraging me to be more confident and a better scientist.

I would like to thank all the past and current members of Kraus lab for their great support, stimulating scientific interactions, and for making the Kraus lab environment fun and enjoyable. Special thanks to Dr. Shrikanth Gadad, Dr. Bryan Gibson, Dr. Rebecca Gupte, Dr. Hector Franco, Dr. Daeseok Kim, Dr. Ken Lin, Dr. Xin Luo, Dr. Miao Sun, Dr. Balaji Parameswaran, Keun Ryu, Rui Li, Shino Murakami, Rachel Ramirez, for being both wonderful colleagues and great friends of me.

A special thanks to all my dear friends: Hui Xu, Luying Jia, Rui Zhong, Yuan Lin, Jueqi Chen, Pei Wang, Zhao Jin, Yi Liu, Xiuli Liu, Muqing Cao, Xin Luo, Miao Sun, Rui Li, Ting Zhou, Keun Ryu. I am lucky to have you all as my friends and grateful for your support and friendship.

Lastly, I would like to express my greatest gratitude and love to my family: my parents and grandparents, who stand by me, love me, support me, and believe in me unconditionally in every possible way; my husband Lei Wang, who is my cheerleader,

always accompany me, encourage me and love me. I would not have been here without you all!

BIOGRAPHICAL SKETCH

Ziying Liu was born in Shenyang, China. In 2005, she started the undergraduate program of Biological Science and Technology in Tsinghua University, Beijing, China, where she developed her strong interest in life sciences. In 2009, she received her Bachelor's degree in Biological Science from Tsinghua University. In the same year, she joined the Graduate Field of Biochemistry, Molecular and Cell Biology at Cornell University. In 2010, she joined the laboratory of Dr. W. Lee Kraus. In the August of 2010, she moved with the laboratory of Dr. W. Lee Kraus from Cornell University to the University of Texas Southwestern Medical Center. In her graduate studies, Ziying initiated and developed a project studying the transcriptional regulation functions of PARP-1 in embryonic stem cells using a combination of cell biology, molecular biology, as well as computational biology methods. She also expanded her studies in triple negative breast cancer system.

MODULATION OF TRANSCRIPTION FACTOR CHROMATIN ASSOCIATION
AND GENE TRANSCRIPTION PROGRAM IN EMBRYONIC STEM CELL
AND TRIPLE NEGATIVE BREAST CANCER BY
POLY (ADP-RIBOSE) POLYMERASE 1

Ziying Liu, Ph.D

The University of Texas Southwestern Medical Center at Dallas

2016

Supervising Professor: W. Lee Kraus, Ph.D

Abstract

Poly(ADP-ribose) polymerase-1 (PARP-1), also referred to as ADP-ribosyltransferase Diphtheria toxin-like 1 (ARTD1), is an abundant nuclear protein that plays key roles in a variety of nuclear processes, including the regulation of transcription. PARP-1 possesses an intrinsic enzymatic activity that catalyzes the transfer of ADP-ribose (ADPR) units from nicotinamide adenine dinucleotide (NAD⁺) onto target gene regulatory proteins, thereby modulating their activities. Although great strides have been made in the past decade in deciphering the seemingly opposing and varied roles of PARP-1 in gene regulation, many puzzles remain in this field. Using a combination of cell biology, molecular biology, genomics and biochemistry methods, I investigated the functions of PARP-1 in regulating gene transcription program in mouse embryonic stem cells and human triple negative breast cancer cells. I found that in mouse embryonic stem cells, PARP-1 functions as a pre-pioneering factor, stabilizing transcription factor Sox2 interaction with nucleosomes. This function is required for maintaining gene transcription program in embryonic stem cells. Depletion of PARP-1 causes disrupted embryonic stem cell gene expression profile, including decreased expression of Nanog, as well as increased expression of differentiation genes. Furthermore, using human triple negative breast cancer cells, I showed that this gene transcriptional regulation mechanism through PARP-1-Sox2 interplay is conserved in different physiological models. Interestingly,

inhibiting PARylation activity causes gain of Sox2 binding to a set of genomic locations in TNBC cells, indicating that PARylation activity plays an antagonizing role in PARP-1-regulated Sox2 chromatin interaction. In summary, our results illustrate how PARP-1 can act at the level of the nucleosome to produce global effects on transcription factor binding and biologically important gene expression outcomes.

TABLE OF CONTENTS

CHAPTER 1. Regulation of Chromatin Structure and Function by PARP-1 and ADP-ribosylation	
1.1. Summary	2
1.2. Introduction	2
1.3. PARP-1, PARylation, and the PARP Family	3
1.3.1. PARP-1 structure and function	3
1.3.2. PARP-1 targets.....	5
1.3.3. PAR binding modules in histone variants and chromatin-regulating proteins	7
1.3.4. The PARP family	8
1.4. Modulation of Chromatin Structure and Gene Expression by PARP-1	10
1.4.1. PARP-1, open chromatin, and enhanced gene expression.....	10
1.4.2. PARP-1 and the linker histone H1	11
1.4.3. PARP-1 and core histones	13
1.4.4. PARP-1 and histone modifications	14
1.4.5. PARP-1 and chromatin remodelers	15
1.4.6. PARP-1 and DNA methylation.....	16
1.4.7. PARP-1, CTCF, and insulator function	17
1.5. PARP-1 functions in heterochromatin	19
1.5.1. A link between PARP-1 function and heterochromatin	19
1.5.2. PARP-1, nucleolar function, and rDNA silencing.....	20
1.5.3. PARP-1, X chromosome inactivation, and macroH2A	22
1.6. Regulation of PARP-1 Localization and Activity in Chromatin	25
1.6.1. Regulation of PARP-1 through interactions with nucleosomes	25
1.6.2. Regulation of PARP-1 by non-histone protein binding partners and post-translational modifications.....	26
1.6.3. Regulation of PARP-1 by cellular signaling pathways.....	27
1.6.4. Modulation of PAR activity by nuclear NAD ⁺ metabolism	28
1.7. Conclusions	29
REFERENCES	30

CHAPTER 2. PARP-1 Facilitates Sox2 Binding to Nucleosomes to Maintain a Pluripotency Gene Expression Program in Embryonic Stem Cells

2.1. Summary	39
2.2. Introduction	40
2.3. Results	41
2.3.1. PARP-1 safeguards the “stemness” of embryonic stem cells by transcriptional regulation.	41
2.3.2. PARP-1 globally co-localizes and interacts with Sox2.	46
2.3.3. PARP-1 is required to stabilize the chromatin association of Sox2 to a subset of genomic locations.	50
2.3.4. PARP-1-dependent Sox2 binding sites are associated with closed chromatin conformation.	56
2.3.5. PARP-1 stabilizes Sox2 binding to nucleosome in vitro.	64
2.3.6. Formation of PARP-1-Sox2 complex on the nucleosome stabilizes the interaction between Sox2 and nucleosome.	67
2.3.7. Regulation of Sox2 interaction with nucleosome by the position of Sox motif.....	71
2.4. Discussion	72
2.4.1. Modulation of Oct4 and Sox2 DNA binding by nucleosome occupancy.....	72
2.4.2. Regulation of Sox2 binding to nucleosome by position of Sox motif in the nucleosome	72
2.4.3. Role of PARP-1 PARylation in regulating chromatin association of Sox2.....	73
2.4.4. Potential mechanisms of PARP-1 regulating somatic reprogramming	74
2.5. Materials and Methods	75
REFERENCES	92

CHAPTER 3. Regulation of Triple Negative Breast Cancer by Sox2 and PARP-1

3.1. Summary	98
3.2. Introduction	99
3.3. Results	101
3.3.1. Sox2-expressing TNBC cells proliferate slower and are more invasive.	101
3.3.2. Ectopic Sox2 expression alters gene transcription through chromatin association in MDA-MB-231 cells.	103
3.3.3. Sox2 regulates gene transcription by direct association with distal regulatory elements	105

3.3.4. Sox2 binds to a distinct cohort of genomic locations in triple negative breast cancer compared to embryonic stem cells	108
3.3.5. Regulation of Sox2 chromatin association by PARP-1	109
3.3.6. Blocking PARylation activity reshapes Sox2 genomic localization profile	111
3.3.7. Modulation of Sox2-mediated gene transcription regulation through PAR inhibition.	113
3.4. Discussion	116
3.4.1. Potential regulators mediating cell type-specific function of Sox2 in triple negative breast cancer.	117
3.4.2. Interplay between PAR inhibition and Sox2 in gene transcription regulation	117
3.4.3. Potential beneficial role of PAR inhibitor in treating Sox2-positive TNBC	119
3.5. Materials and Methods.....	120
REFERENCES.....	128

PRIOR PUBLICATIONS

Doiguchi, M., Nakagawa, T., Imamura, Y., Yoneda, M., Higashi, M., Kubota, K., Yamashita, S., Asahara, H., Iida, M., Fujii, S., Ikura, T., **Liu, Z.**, Nandu T., Kraus, W.L., Hitoshi, U., Ito, T., SMARCAD1 is an ATP-dependent stimulator of nucleosomal H2A acetylation via CBP, resulting in transcriptional regulation. *Scientific Reports*. 6:20179 | DOI: 10.1038/srep20179

Liu, Z., Kraus, W.L., Regulation of Chromatin Structure and Function by PARP-1 and ADP-Ribosylation. *Fundamentals of Chromatin*. (Book Chapter)., Springer. Edited by Jerry Workman.

TABLE OF FIGURES

CHAPTER 1. Regulation of Chromatin Structure and Function by PARP-1 and ADP-ribosylation

Figure 1.1. Structural and functional domains of PARP-1, an ADP-ribosyl transferase that links nuclear NAD ⁺ metabolism to protein modification.	4
Figure 1.2. Binding partners of PARP-1.....	6
Figure 1.3. Effects of PARylation target protein.	7
Figure 1.4. Molecular actions of PARP-1.....	9
Figure 1.5. PARP-1 functions in heterochromatin.....	19
Figure 1.6. Regulatory inputs and outputs for PARP-1	24

CHAPTER 2. PARP-1 Facilitates Sox2 Binding to Nucleosomes to Maintain a Pluripotency Gene Expression Program in Embryonic Stem Cells

Figure 2.1. Knockout of Parp1 alters the gene expression program in mESCs.....	43
Figure 2.2. The promoters of genes negatively regulated by PARP-1 are enriched for both H3K4me3 and H3K27me3	44
Figure 2.3. Gene expression changes in mESC upon Parp1 depletion.....	45
Figure 2.4. The catalytic activity of PARP-1 is not required for the regulation of gene expression in undifferentiated mESCs.....	46
Figure 2.5. PARP-1 is enriched at the promoters of highly expressed genes.....	47
Figure 2.6. PARP-1 colocalizes with Sox2 genome-wide and helps to maintain a pluripotency gene expression program in mouse embryonic stem cells.....	49
Figure 2.7. Decreased gene expression in mESCs upon Parp1 knockout correlate with decreased Sox2 binding	52
Figure 2.8. Sox2 binding sites without Oct4, but not Oct4 binding sites without Sox2, exhibit PARP-1-dependent binding	53
Figure 2.9. Increase of gene expression upon Parp-1 depletion is partially explained by decreased level of Nanog.....	54
Figure 2.10. PARP-1 is required for the binding of Sox2 to a subset of its genomic sites in mESCs	55
Figure 2.11. PARP-1-dependent Sox2 binding sites exhibit higher nucleosome occupancy and lower co-occupancy by other transcription factors than PARP-1-independent Sox2 binding sites	57

Figure 2.12. PARP-1-dependent Sox2 binding sites exhibit lower co-occupancy by other transcription factors than PARP-1-independent Sox2 binding sites.....	58
Figure 2.13. Parp-1 dependent Sox2 binding sites have less cooperation with other transcription factors	59
Figure 2.14. Summary of genomic features for PARP-1-independent and PARP-1-dependent Sox2 binding sites.....	60
Figure 2.15. Recombinant proteins and nucleosomes used in biochemical assays exploring functional interactions between PARP-1 and Sox2.....	61
Figure 2.16. PARP-1 stabilizes Sox2 binding to nucleosomes.....	63
Figure 2.17. The PARP-1 BRCT domain mediates interactions between PARP-1 and Sox2	66
Figure 2.18. Sox2 redirects PARP-1 binding on the nucleosome.....	67
Figure 2.19. Assembly of mononucleosomes containing a natural Pou/Sox motif from the Nanog promoter and binding by Oct4 to the mononucleosomes.....	69
Figure 2.20. PARP-1 facilitates Sox2 binding to nucleosomes at Sox sites located in wide, outward-facing minor grooves	70
Figure 2.21. Model for PARP-1-dependent binding of Sox2 to nucleosomes	75

CHAPTER 3. Regulation of Triple Negative Breast Cancer by Sox2 and PARP-1

Figure 3.1. Sox2 promotes metastasis in triple negative breast cancer	102
Figure 3.2. Ectopic Sox2 expression inhibits cell growth in HCC1143 cells.....	103
Figure 3.3. Ectopic Sox2 expression alters global gene expression in MDA-MB-231 cells	104
Figure 3.4. Sox2 directly binds to distal regions and regulates gene expression.....	106
Figure 3.5. Characterization of chromatin features of Sox2 peaks in MDA-MB-231 cells	107
Figure 3.6. Sox2 binds to a distinct cohort of genomic locations in triple negative breast cancer compared to embryonic stem cells	109
Figure 3.7. Sox2 association with chromatin requires PARP-1.....	110
Figure 3.8. Inhibition of PARP-1 activity using PJ34.....	110
Figure 3.9. Blocking PARP-1 activity promotes chromatin association of Sox2.....	112
Figure 3.10. Newly gained Sox2 binding upon PAR inhibition are associated with closed chromatin	113
Figure 3.11. Modulation of Sox2-mediated gene transcription regulation through PAR inhibition	114
Figure 3.12. Modulation of Sox2-mediated gene transcription regulation through PAR inhibition.....	115

Figure 3.13. Modulation of Sox2-mediated gene transcription regulation through PAR inhibition	116
Figure 3.14. Model for regulation of Sox2 chromatin association by modulating PARylation activity.....	119

LIST OF DEFINITIONS

PARP-1 – Poly (ADP-Ribose) Polymerase 1

PAR - poly(ADP-ribose)

NAD+ - Nicotinamide adenine dinucleotide

DBD - DNA binding domain

BRCT - BRCA1 C terminus

CAT - Catalytic

PARylate – Poly ADP-Ribosylate

ESC – Embryonic stem cell

iPSC - Induced pluripotent stem cell

ChIP - Chromatin immunoprecipitation

TSS – Transcription start site

TTS - Transcription termination site

GO - Gene Ontology

GSEA - Gene set enrichment analysis

DHS – DNase I hypersensitivity sites

NRP – Nuclosome rotational positioning

MGW – Minor groove width

TNBC – Triple negative breast cancer

CSC – Cancer stem cells

CHAPTER 1

Regulation of Chromatin Structure and Function by PARP-1 and ADP-ribosylation

1.1. Summary

Poly(ADP-ribose) polymerase-1 (PARP-1), also referred to as ADP-ribosyltransferase Diphtheria toxin-like 1 (ARTD1), is an abundant nuclear protein that plays key roles in a variety of nuclear processes, including the regulation of transcription. PARP-1 possesses an intrinsic enzymatic activity that catalyzes the transfer of ADP-ribose (ADPR) units from nicotinamide adenine dinucleotide (NAD⁺) onto target gene regulatory proteins, thereby modulating their activities. Although great strides have been made in the past decade in deciphering the seemingly opposing and varied roles of PARP-1 in gene regulation, many puzzles remain. In this review, we discuss the current state of understanding in this area, especially how PARP-1 interfaces with various components of gene regulatory pathways (e.g., the basal transcription machinery, DNA-binding transcription factors, coregulators, chromatin remodeling, histone modifications, and DNA methylation). In addition, we discuss some gene-specific, cell type-specific, cell state-specific effects of PARP-1 on gene regulation, which might contribute to its biological functions. Finally, we review some of the recent progress targeting PARPs using chemical inhibitors, some of which may alter PARP-1-dependent gene regulatory programs to promote therapeutic outcomes.

1.2. Introduction

Chromatin, a repeating array of nucleosomes (i.e., 146 bp of DNA wrapped around an octamer of core histone proteins) and its associated linker histones and non-histone proteins, plays key roles in a variety of nuclear functions related to genomic DNA, including transcription, replication, repair, and recombination¹. A wide array of chromatin-modulating proteins have evolved to exploit the regulatory potential of chromatin and to ensure that the fidelity of these processes remains intact. Poly(ADP-ribose) polymerase-1 (PARP-1) is a ubiquitous and abundant nuclear protein, whose functions are dependent on its interactions with chromatin, as well as chromatin-modulating proteins². Many studies over the past few decades have identified and characterized the physical and functional interaction of PARP-1 with chromatin and, more recently, have begun to elucidate the effect they play in physiology and pathology. In this review, we summarize the current knowledge in this area and highlight key examples that provide insights into the functional interplay between PARP-1 and chromatin. In addition, where appropriate, we will describe the related activities of PARP-2 and other PARP family members.

1.3. PARP-1, PARylation, and the PARP Family

1.3.1. PARP-1 structure and function

PARP-1 is a 1014 amino acid protein (~116 kDa) (Fig. 1.1A), which possesses an intrinsic ADP-ribosyltransferase activity that transfers ADP-ribose moieties from donor NAD⁺ molecules (Fig. 1.1B) to glutamate, aspartate, and lysine residues of its target proteins and joins them in chains of poly(ADP-ribose) (PAR) (Fig. 1.1C)³. PARP-1 contains six independently folded domains, comprising three major functional units (Fig. 1.1A): (1) an amino-terminal DNA binding domain (DBD), which contains two zinc finger domains followed by a newly identified zinc-binding domain⁴; (2) an automodification domain (AMD) containing a BRCT fold thought to play an important role in mediating protein-protein interactions⁵; and (3) a well-conserved carboxyl-terminal catalytic domain (CD) containing a WGR motif, which may function in nucleic acid binding⁴, an α -helical PARP regulatory domain (PRD), which interacts with the substrate-

binding site and may modulate branching of the PAR chain, and an NAD⁺-binding “PARP signature” motif, which is conserved across PARP family members and is critical for PARP-1 enzymatic activity (Fig. 1.1A)⁶. These functional units interact in the overall structure of PARP-1 to confer the chromatin- and gene-regulating properties of PARP-1.

PARP-1 functions in a wide spectrum of physiological processes and pathological processes by controlling key molecular events in the nucleus^{7,8}. Although PARP-1 was originally characterized as a DNA-damage response protein, studies over the past decade have identified key roles for PARP in the modulation of chromatin structure and transcription. And, in spite of the unremarkable phenotype reported in the original report of PARP-1 null (*PARP-1*^{-/-}) mice, recent studies in which *PARP-1*^{-/-} mice have been subjected various stressors have begun to expand our understanding of PARP-1 biology⁸. For example, *PARP-1*^{-/-} mice subjected to immune challenges, high fat diet⁹, or altered light/dark cycles have revealed key roles for PARP-1 in innate immune response, inflammation, cell and organismal metabolism¹⁰⁻¹², and circadian rhythms¹³. In addition, PARP-1 has also been shown to play roles in hormone-dependent cellular outcomes, cellular differentiation, and neuronal function. Nonetheless, key questions about PARP-1 function remain.

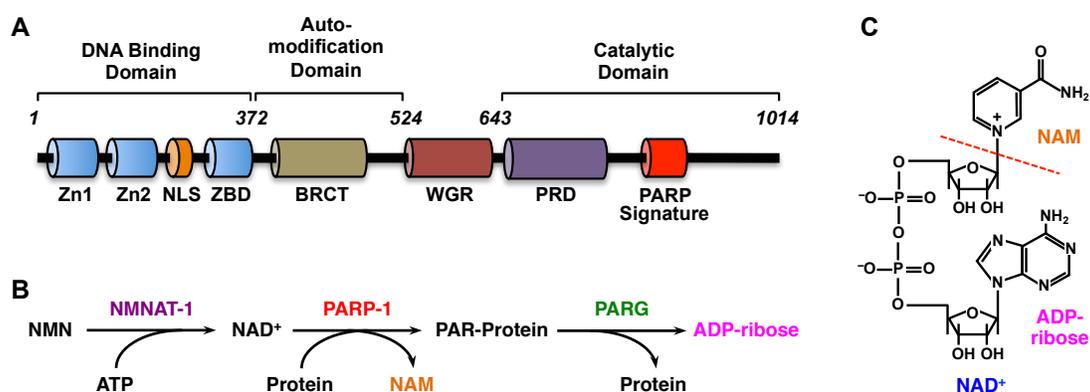


Figure 1.1. Structural and functional domains of PARP-1, an ADP-ribosyl transferase that links nuclear NAD⁺ metabolism to protein modification.

A) PARP-1 contains six independently folded domains, comprising three major functional units (1) an amino-terminal DNA-binding domain, which contains two zinc finger domains (Zn1 and Zn2) followed by a newly identified zinc-binding domain (ZBD); (2) an

automodification domain containing a BRCA1 C terminus (BRCT) fold thought to play an important role in mediating protein–protein interactions; and (3) a well-conserved carboxyl-terminal catalytic domain containing a WGR (Trp–Gly–Arg) motif, which may function in nucleic acid binding, an α -helical PARP regulatory domain (PRD), which interacts with the substrate-binding site and may modulate branching of the PAR chain, and an NAD⁺-binding “PARP signature” motif, which is conserved across PARP family members and is critical for PARP-1 enzymatic activity.

B) Enzymes, substrates, and products in the nuclear NAD⁺ metabolic pathway. Nicotinamide mononucleotide adenylyltransferase-1 (NMNAT-1) catalyzes the synthesis of NAD⁺ from nicotinamide mononucleotide (NMN) and ATP. PARP-1 uses ADP-ribose units donated by NAD⁺ to catalyze the addition of poly(ADP-ribose) (PAR) polymers on target proteins, release nicotinamide (NAM). Poly(ADP-ribose) glycohydrolase (PARG) hydrolyzes the PAR chains, releasing ADP-ribose.

C) Chemical structure of NAD⁺ showing the component ADP-ribose and NAM moieties.

1.3.2. PARP-1 targets

PARP-1’s enzymatic activity is required for many of PARP-1’s functions and a large number of proteins have been identified as targets of PARP-1-mediated PARylation, including PARP-1 itself (through an automodification reaction) and a wide variety of nuclear proteins (e.g., core histones, linker histone H1, chromatin-modulating enzymes^{3,14} etc.; discussed in more detail below). The addition of tens, or even hundreds, of ADP-ribose units significantly changes the biochemical properties of the target proteins¹⁵. As such, PARylation functions as an important post-translational modification in a variety of fundamental cellular functions. PARylation can alter target protein activity by (1) altering the affinity of protein-protein or protein-nucleic acid interactions (Fig. 1.3A-B), (2) modulating protein stability through crosstalk with ubiquitylation pathways (Fig. 1.3C), (3) creating a protein interaction scaffold that can promote the recruitment of PAR-binding proteins to specific sites of action in the nucleus³ (discussed in more detail below) (Fig. 1.3D) or (4) regulation of target protein enzymatic activity (Fig. 1.3E). Although many functions of PARP-1 are dependent on its catalytic activity, PARP-1 may function through catalytic-independent mechanisms as well¹⁵. For example, as we discuss below, PARP-1 can directly modulate chromatin structure through its nucleosome-binding activity, independent of its enzymatic activity¹⁶. PARP-1 can also function as a transcriptional coregulator independent of its enzymatic activity^{1,2}. The interplay between PARP-1’s catalytic-dependent and –independent activities controls PARP-1’s functions in the nucleus.

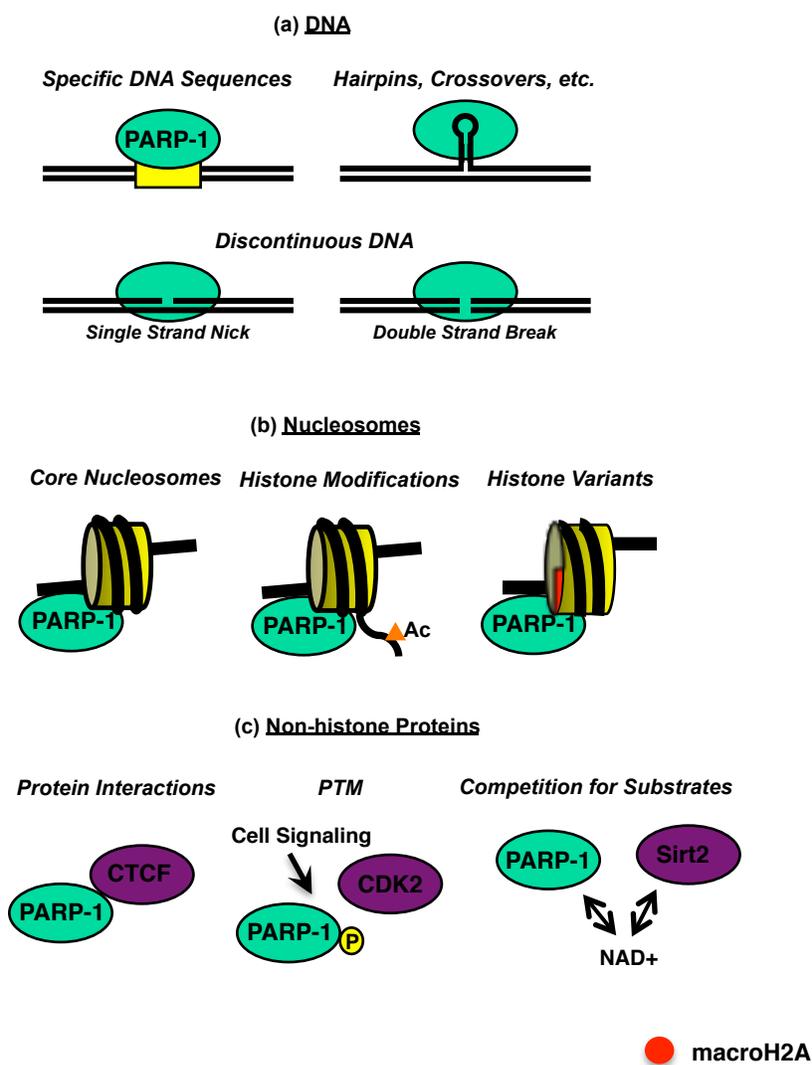


Figure 1.2. Binding partners of PARP-1.

PARP-1 interacts with DNA, proteins, and protein–DNA complexes. (a) PARP-1 can bind to specific DNA sequences, hairpins, or sites of DNA damage. (b) PARP-1 binds to protein–DNA complexes called nucleosomes (two copies each of the four core histones H2A, H2B, H3, and H4 plus 146 bp of DNA) at the dyad axis (*red triangle*) and to the linker DNA where it exits the nucleosome. (c) PARP-1 binds to the proteins that it targets for PARylation.

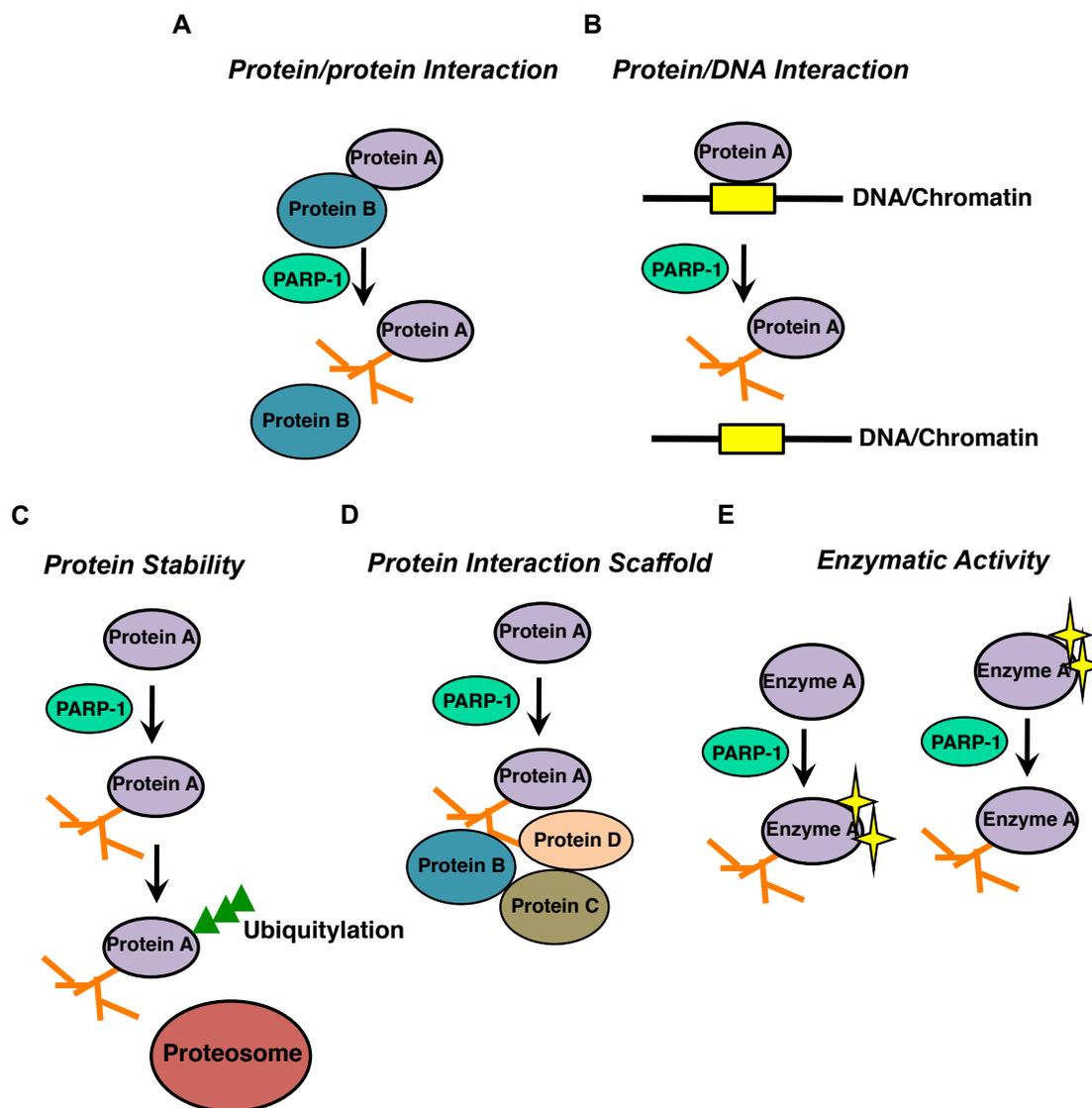


Figure 1.3. Effects of PARylation target protein.

PARylation by PARP-1 has a variety of effects on target proteins, including the following: (A) disruption of protein–protein interactions. (B) Disruption of protein–DNA interactions. (C) Promotion of ubiquitylation and proteasome-mediated degradation. (D) Formation of protein scaffolds. (E) inhibition (*left*) or enhancement/activation (*right*) of enzyme activity.

1.3.3. PAR binding modules in histone variants and chromatin regulating proteins

As noted above, the binding of proteins to PAR can be a key regulatory mechanism by PARP-1 and PARylation. Four different types of PAR-binding modules in proteins have been identified to date^{3,17}: (1) PAR-binding motifs (PBM), a short sequence of amino acids comprising Lys-Arg clusters interspersed with hydrophobic amino acids, as

exemplified by *Drosophila* protein MI2, a homologue of mammalian nucleosome remodeling enzyme CHD4¹⁸; (2) PAR-binding zinc finger (the PBZ domain), as found in the DNA damage response proteins APLF and CHFR¹⁸; (3) macrodomain folds, which is contained in histone variant macroH2A, macroPARP family members as well as some chromatin-remodelling enzymes^{19,20}; and (4) WWE domain that is present in various ubiquitin ligases including RNF146 and ULF^{21,22}. As will be discussed below, proteins containing PAR binding modules can recognize and bind specifically to PARylation signals, which further has important regulatory roles by modulating protein localization or enzymatic activities.

1.3.4. The PARP family

Proteins with ADP-ribosyltransferase activity have been characterized across a wide array of species in all kingdoms of life, including eukaryotes (except for yeast), eubacteria, archaeobacteria, and even some DNA viruses²³. PARP-1, which was the first protein identified with poly(ADP-ribosyl) transferase, is the founding member of the PARP family of proteins, which is defined based on homology to the PARP signature motif. PARP family can be further classified into four subfamilies based on their structures, associated functional domains, and enzymatic activities, including (1) DNA damage-dependent PARPs (PARP-1, PARP-2, PARP-3), which are activated by damaged DNA and other DNA structures through their N-terminal DNA-binding domains; (2) tankyrases (tankyrase 1 and tankyrase 2), which contain large ankyrin domain repeats that facilitate target selection and activation; (3) CCCH PARPs (PARP-7, PARP-12, PARP-13.1, and PARP-13.2), which contain RNA-binding Cys-Cys-Cys-His zinc fingers and PAR-binding WWE domains; and (4) macrodomain PARPs (BAL1/PARP-9, BAL2/PARP14, and BAL3/PARP-15), which contain ADP-ribose- and PAR-binding macrodomain folds; as well as a few additional members that do not belong to these subfamilies^{3,6}. PARP family members vary in their enzymatic activities, and some are even catalytically inactive: PARPs 1 and 2, vPARP, and tankyrase 1 and 2 catalyse poly(ADP-ribosyl)ation, PARPs 3, 10, 14 and 15 catalyse mono(ADP-ribosyl)ation, and the remaining PARP family members are thought to be inactive. Recently, a new nomenclature for the PARP family has been proposed – the ADP-ribosyltransferase Diphtheria toxin-like (ARTD) family, which is based on a more accurate description of

the mode of catalysis²⁴. In this new nomenclature, PARP-1 is referred to as ARTD1, recognizing it as the prototypical PARP family member.

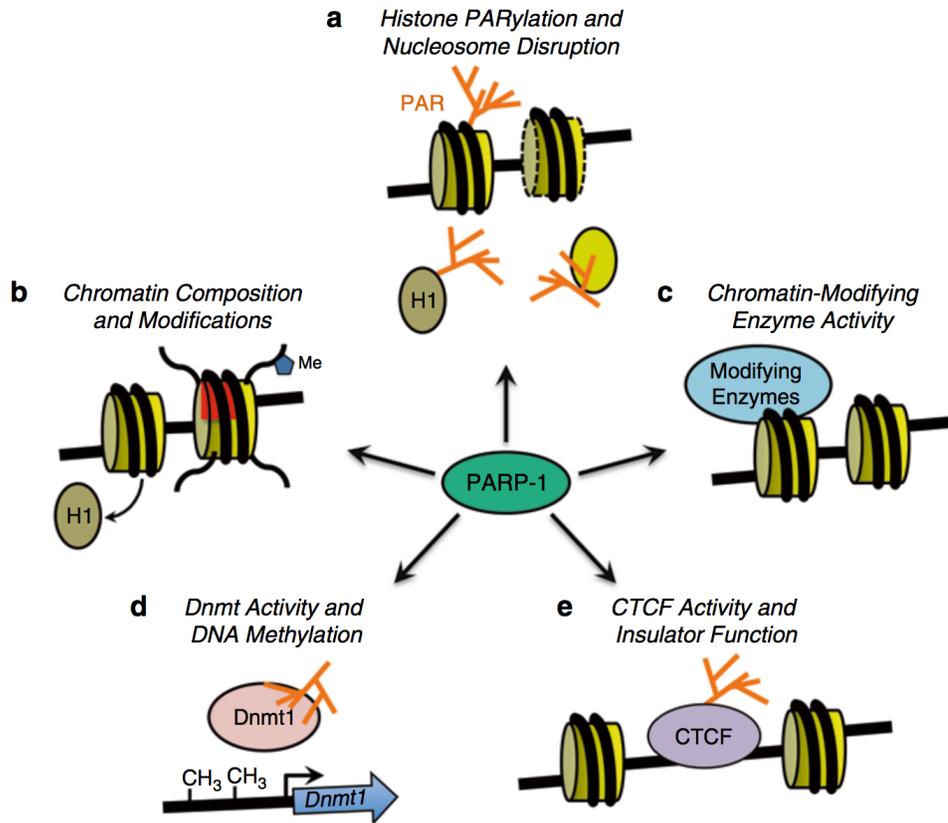


Figure 1.4. Molecular actions of PARP-1.

PARP-1 plays a variety of roles in the nucleus, many of which are targeted toward the modulation of chromatin structure, including the following (a) PARylation of histones, which may disrupt nucleosome structure, or serve as a negatively charged matrix that can bind histones. (b) Modulating the composition of chromatin (e.g., inhibiting the binding of the linker histone H1, promoting the incorporation of histone variants) or the posttranslational modification state of core histones (i.e., by altering the activity of histone-modifying enzymes). (c) Modulating the activity of chromatin-modifying enzymes, such as the ATP-dependent nucleosome remodeling enzymes ISWI and ALC1 and the histone demethylase KDM5B. (d) Modulating the activity of the DNA methyltransferase Dnmt1, which affects the extent of DNA methylation. (e) Modulating the activity of CTCF, which affects its insulator function.

1.4. Modulation of Chromatin Structure and Gene Expression by PARP-1

In spite of the initial, seemingly singular, focus on the role of PARP-1 in DNA damage detection and repair responses in the historical literature, a considerable amount of evidence from the past decade supports a key role for PARP-1 in the modulation of chromatin structure and gene expression – a role that may be its most important cellular function in normal physiological states^{1,2,16}. In this section, we discuss the regulation of gene expression by PARP-1 via chromatin-dependent mechanisms. Note that PARP-1 also controls gene expression by acting as a classical transcriptional coregulator with a number of different sequence-specific DNA binding transcription factors (e.g., NF- κ B¹⁷, nuclear receptors^{1,25}, and many others). The coregulator activity of PARP-1 is beyond the scope of this review on chromatin-dependent mechanisms, but this topic has been reviewed elsewhere and the reader is directed there^{1,7}.

1.4.1. PARP-1, open chromatin, and enhanced gene expression

Early biochemical studies suggested that PARP-1 preferentially associates with open, transcriptionally active regions of chromatin. Studies by Huletsky et al. provided the first evidence that PARP-1 and PARylation promote the formation of an open chromatin structure in biochemical studies using polynucleosomes isolated from calf thymus nuclei^{26 15}. They showed that PARylation of chromatin promotes decondensation and maintains polynucleosomes in a more accessible and open state. This observation is supported by in vivo studies from Tulin et al., who showed using isolated salivary glands from *Drosophila* larvae that the *Drosophila* PARP-1 homolog dPARP and its activity are

required for the formation of ecdysone- or heat shock-induced “puffs” (regions of transcriptionally active decondensed chromatin) on polytene chromosomes²⁷. Although dPARP is broadly localized across euchromatin, PAR activity is localized preferentially to the “puff” regions. Importantly, inhibition of dPARP enzymatic activity blocks both puff formation and heat shock-induced gene expression²⁷.

In mammalian cells, PARP-1 also maintains an open, transcriptionally active chromatin architecture at the promoters of subsets of its target genes. In chromatin immunoprecipitation (ChIP)-based genome-wide PARP-1 localization analyses in MCF-7 human breast cancer cells, PARP-1 is enriched at the transcription start sites (TSSs) of actively transcribed genes^{23,28}, correlating strongly with the enrichment of histone H3 lysine 4 trimethylation (H3K4me3)²⁸, a mark of active promoters²³. RNAi-mediated knockdown of PARP-1 in MCF-7 cells decreases chromatin accessibility (as determined by digestion with MNase) at the promoters of genes positively regulated by PARP-1. This is accompanied by reduced loading of the RNA polymerase II transcription machinery, H3K4me3, and target gene expression²⁸.

Together, these studies in insect and mammalian cells support a role for PARP-1 in modulating chromatin structure in transcriptionally active regions of the genome. PARP-1, however, does not act exclusively to decondense chromatin and destabilizing nucleosomes in euchromatin regions. Rather, PARP-1 modulates chromatin structure in a context-dependent manner. For example, in *Drosophila*, depletion of dPARP causes early embryonic lethality with defects in heterochromatin^{15,29}, suggesting that dPARP may also be important for the maintenance of the appropriate state of compaction in heterochromatin. In the remaining text in this section, we focus on the detailed mechanisms of PARP-1 function in regions of open chromatin. The specific roles of PARP-1 in heterochromatin are discussed in the following section.

1.4.2. PARP-1 and the linker histone H1

Linker histones, such as H1, bind to nucleosomes³⁰ at the dyad axis and promote the formation of higher-order chromatin structures by compacting nucleosomes^{15,16}. One mechanism by which PARP-1 causes the condensation of chromatin is by promoting the exclusion of H1 from regions of chromatin (Fig 1.4A). In both *Drosophila* and mammals, PARP-1 binding to chromatin is negatively correlated with the binding of H1^{16,30}. dPARP

occupies different regions of chromatin than H1 on *Drosophila* polytene chromosomes¹⁶. Likewise, PARP-1 occupies different regions of chromatin than H1 across the mammalian genome³⁰, showing a reciprocal binding pattern with peaks of PARP-1 localizing to troughs of H1. Importantly, knockdown of PARP-1 leads to increased levels of H1 at the promoters of PARP-1 target genes^{28,30}, suggesting a direct functional interplay between PARP-1 and H1 at these promoters.

The displacement of H1 by PARP-1 can occur through direct competition for binding sites on nucleosomes, independent of PARP-1 catalytic activity. Using a set of biochemical experiments with reconstituted chromatin, Kim et al. showed that PARP-1 binds at or near the dyad axis of nucleosomes with a stoichiometry of one, in a manner largely resembling the binding of H1 to nucleosomes¹⁶. Increasing the concentration of PARP-1 reduced the binding of H1 to chromatin, and vice versa, supporting a model of competition for a common binding site on nucleosomes¹⁶. Alternatively, direct PARylation of H1 may promote its dissociation from chromatin. H1 is PARylated by PARP-1 and may be preferentially PARylated over other chromatin proteins, as demonstrated in studies with isolated native chromatin containing both linker histone and core histones³¹⁻³³. PARylation of histone H1 decreases its affinity for nucleosomes, thus facilitating chromatin decondensation²⁶. These models are not mutually exclusive, but further studies are needed to determine when and where these distinct mechanisms are used in vivo.

PARP-1-mediated displacement of H1 may be regulated in a signal-dependent manner. This has been demonstrated previously for genes activated by estrogen signaling, where the displacement of H1 from the promoter occurs in a topoisomerase II β -dependent manner^{34,35}, and progestin signaling, where the displacement of H1 occurs in cdk2-dependent manner³⁶. In these examples, topoisomerase II β and cdk2 are required to activate PARP-1 enzymatic activity, which is required for the H1 displacement. In contrast, for genes inhibited by the phorbol ester 12-O-tetradecanoylphorbol-13-acetate (TPA), PARP-1 is displaced from the promoter as H1 is loaded²⁸. These examples, highlight the reciprocal functional interactions that occur between PARP-1 and H1 during gene regulation.

1.4.3. PARP-1 and core histones

In addition to its effects on the linker histone H1, PARP-1 may also modulate chromatin structure by acting on core histones in nucleosomes (Fig 1.4A). One possible mechanism for PARP-1-dependent destabilization of nucleosomes is through the PARylation of core histones, although the importance of this as a regulatory mechanism in vivo is unclear. Core histones are mono(ADP-ribosyl)ated in vivo^{14,16}, and, although they can be poly(ADP-ribosyl)ated in vitro^{26,33,37,38} the extent to which core histone PARylation occurs in vivo under normal physiological conditions has not been definitively determined. Although the results of some studies support the existence of histone PARylation in vivo^{15,39}, preferentially on H2A and H2B, this remains an open question and active area of investigation. Adding anionic PAR polymers to core histones may decrease their affinity for negatively charged DNA³³. Using a crude nucleosomal “core particle” containing PARylation activity prepared from isolated rat liver nuclei followed by micrococcal nuclease digestion, Mathis and Althaus showed that PARylation of “core particle proteins”, presumably core histones, by addition of NAD⁺ significantly decreased the affinity of “core particle proteins” for the 146bp DNA³⁷. In contrast, Kim et al¹⁶. did not detect PARylation of nucleosomal core histones using physiological levels of PARP-1 and purified reconstituted polynucleosomal arrays containing only core histones and un-nicked, circular DNA, even though autoPARylation of PARP-1 and structural changes in chromatin were robustly detected in the presence of NAD⁺. The differences between these two studies may reflect differences that occur in vivo between damaged genomic DNA (i.e., exposure to free ends) and undamaged genomic DNA¹. In the former, PARP-1-dependent PARylation of core histones may be dramatically potentiated.

A role for DNA damage in promoting histone PARylation is supported by studies using DNA damaging agents⁴⁰. For example, under conditions of alkylation-induced DNA damage, histone H2B is a major PARylation acceptor in SV40 minichromosomes³⁹. Moreover, during free radical-induced DNA damage, about 2-3% of histones H1, H3, H2B, and H4 in cells are PARylated. Again, it remains unclear if this represents a general mechanism applicable to physiological conditions, when DNA damage or other potent stimuli of PARylation activity are absent¹⁶. Also, given the preference of PARP-1 for histone H1 as a substrate, the extent to which core histone PARylation occurs or is

physiologically relevant in vivo is uncertain. In the end, mono(ADP-ribosyl)ation of core histones may occur more frequently than PARylation in vivo and may be more physiologically important. Clearly, further studies are needed to address these questions.

Another possible mechanism for PARP-1-dependent destabilization of nucleosomes is through noncovalent interactions between core histones and PARylated PARP-1 or free PAR¹⁵. PARP-1 is the major target of PARylation activity in cells, accepting approximately 90 percent of PAR through an automodification reaction^{14,26,41}. Free PAR can be generated by catabolism of protein-linked PAR by the enzymatic activity of PAR glycohydrolase (PARG)³. Core histones bind in a polymer-length dependent manner to either free or covalently-linked PAR^{33,37,42}, raising the possibility that PAR can function as core histone “sink”. In such a scenario, PAR may bind core histones transiently dissociated from nucleosomes by chromatin-modulating enzymes or actively promote the removal of core histones by competing with genomic DNA.

A recent study by Petesch and Lis provide support for PARP-1-dependent destabilization of nucleosomes. They observed a dramatic, transcription-independent decrease of nucleosome occupancy in the *hsp70* gene body immediately following heat shock in *Drosophila* S2 cells⁴³, which was dependent on dPARP and its catalytic activity; RNAi-mediated knock down of dPARP or chemical inhibition of its enzymatic activity inhibits heat shock-induced nucleosome loss and *hsp70* transcription⁴³⁻⁴⁵. dPARP initially localizes at the first nucleosome downstream of the TSS, but becomes rapidly redistributed through the gene body right after heat shock, tracking with elongating RNA polymerase II, dependent on dPARP activation. The target of the dPARP enzymatic activity in these studies is unclear, but the authors propose that PARylation of dPARP dissociates it from nucleosomes, causing it to relocate throughout gene body, where accumulated PAR may strip gene body core histones from nucleosomes. Further studies of this and other robustly inducible gene expression systems should help to clarify the specific roles and mechanisms of PARP-1-dependent nucleosome destabilization.

1.4.4. PARP-1 and histone modifications

Covalent, post-translational modification of histone can dramatically affect chromatin architecture function, and represents another target point for regulation by PARP-1. PARP-1's effects on histone modifications can occur through its effects on the

localization and activity of histone-modifying enzymes (Fig 1.4B). For example, Krishnakumar et al. (2010) showed that PARP-1 interacts with and PARylates histone demethylase KDM5B, thus preventing it from binding to the promoters of genes positively regulated by PARP-1^{23,28}. RNAi-mediated knockdown of PARP-1 permits increased binding of KDM5B and concomitant reductions in H3K4me3 levels on these genes. Not surprisingly given these results, the genomic localization of PARP-1 significantly correlates with H3K4me3 levels. In another recent study focused on reprogramming of somatic cells, PARP-1, as well as its activity, have been shown to play critical role in enhancing the efficiency of induced pluripotency in fibroblasts. PARP-1 deficiency causes a decrease in H3K4me2 levels and an increase in H3K27me3 levels at the *NANOG* and *ESRRB* loci, which encode pluripotency transcription factors. The altered chromatin state correlates with decreased transcription factor accessibility⁴⁶. However, it is unclear whether these changes are a direct effect of PARP-1 on the enzymes that control H3K4me2 and H3K27me3 levels, or is it simply a consequence of changes in gene expression. Identifying and characterizing more direct targets of PARP-1-dependent regulation, such as KDM5B, will provide mechanistic insights on the regulation of histone modification by PARP-1.

1.4.5. PARP-1 and chromatin remodelers

PARP-1 can also affect chromatin structure by regulating the activity of ATP-dependent nucleosome remodeling enzymes (Fig 1.4C). For example, in *Drosophila*, functions of ISWI, a nucleosome remodeling enzyme having chromosome-compacting effects⁴⁷, can be counteracted by dPARP⁴⁸. ISWI catalyzes nucleosome spacing and sliding reactions using energy released from ATP hydrolysis⁴⁹, and plays important role in mediating chromosome compaction in *Drosophila*, at least in part, by facilitating linker histone association^{48,50}. ISWI occupies distinct regions from dPARP across the genome and loss of dPARP causes a global increase in chromatin association by ISWI. Furthermore, ISWI is a target of PARylation by dPARP both in vitro and in vivo; PARylation of ISWI inhibits its ATPase activity and reduces its nucleosome binding capacity⁴⁸. These observations suggest an antagonistic effect of dPARP on the formation of higher-order chromatin structures mediated by ISWI, yet it remains to be determined

whether the same functional interplay between PARP-1 and ISWI exists in mammalian cells.

PARP-1 also affects the activity of ATP-dependent nucleosome remodeling enzyme, ALC1 (Amplified in Liver Cancer 1; a.k.a. CHD1L), through a unique mechanism that involves the binding of PAR by ALC1^{8,18,51}. ALC1 lacks chromatin targeting domains that are found in other nucleosome remodeling enzymes (e.g., bromo- or chromodomains), but it does contain a macrodomain at its carboxyl-terminal end, which serves as a PAR-binding domain^{18,51}. In response to DNA damage, ALC1 is recruited to of DNA damage-induced PAR foci in the genome via its macrodomain. Association with PAR dramatically increases ALC1 ATPase and nucleosome remodeling activities, thus promoting the formation of a relaxed chromatin conformation that facilitates DNA repair. Although these PAR-induced effects of ALC1 have been studied in response to DNA damage, these mechanisms may also apply to signal-dependent gene regulation where nucleosome remodeling events are required.

1.4.6. PARP-1 and DNA methylation

DNA methylation, predominantly 5-methyl-cytosine, plays an important role in modulating chromatin conformation and functions, with high level of DNA methylation associated with the formation of compact chromatin structures⁵². The first direct evidence of a role for PARP-1 and PAR in the regulation of DNA methylation (Fig 1.4D) came from Zardo et al. (1998)⁵³, who showed using methylation-sensitive restriction enzymes and bisulfate conversion of methyl-DNA followed by restriction enzyme digestion that PARylation decreases DNA methylation levels within the CpG island of *HTF8* gene promoter in L929 mouse fibroblasts. Furthermore, blocking PARylation causes chromatin compaction and induces a global DNA hypermethylation, as shown by imaging in the L929 cells⁵⁴. One potential mechanism for PARP-1-mediated regulation of DNA methylation is through Dnmt1, the major mammalian “maintenance” DNA methyltransferase, which maintains DNA methylation patterns through DNA replication^{55,56}. PARP-1 can regulate Dnmt1 through affecting its transcription level^{55,57}, with PARP-1 protein binding to Dnmt1 gene promoter⁵⁸. PARylation is likely to be involved in PARP-1 regulation of Dnmt1 gene transcription, since overexpressing PARG

which decreases PARylation level causes an abnormality in DNA methylation pattern on CpG island in Dnmt1 promoter and inhibits Dnmt1 gene transcription⁵⁸. In addition, Dnmt1 protein contains a consensus PAR binding motif and is able to interact noncovalently with PAR polymers⁵⁹. Blocking PARylation in L929 as well as NIH/3T3 mouse fibroblast cells increases Dnmt1 methyltransferase activity, whereas binding to PAR polymers inhibits its activity⁵⁹. Note, however, that the studies described here are mostly based on inhibitor treatment or in vitro biochemical assays. To fully understand the regulatory effects of PARP-1 on Dnmt1, more thorough in vivo studies characterizing the functional interplay between PARP-1 and Dnmt1 are needed.

Recent progress in understanding mechanisms of active DNA demethylation has added another potential layer to the regulation of DNA methylation by PARP-1⁶⁰⁻⁶². In this regard, genome-wide DNA demethylation in primordial germ cells between embryonic days 10.5 and 12.5 correlates with the appearance of single-strand DNA breaks, an activation of the base excision repair (BER) pathway, and a high level of PARylation⁶³. Inhibition of PAR formation and BER activity inhibits DNA demethylation in the paternal pronucleus in zygotes shortly after fertilization. These results suggest that PARP-1 and PARylation are required for active DNA demethylation during PGC development, probably functioning through the BER pathway. Another potential mechanism for the regulation of DNA methylation by PARP-1 involving active DNA demethylation involves the TET family proteins, which can convert 5-methyl-cytosine to 5-hydroxymethyl-cytidine^{46,62}. The latter can be further removed and replaced by unmethylated cytidine. Although there is no direct evidence functionally linking PARP-1 to TET enzymes yet, recent results may hint at some functional interactions. For example, both PARP-1 and TET2 are required for somatic cell reprogramming, and deficiency in both factors causes a change in patterns of histone modification as well as transcription level at the *NANOG* and *ESSRB* genes in a similar manner⁴⁶. Further studies are required to determine the underlying mechanisms and potential functional interplay between PARP-1 and TET2.

1.4.7. PARP-1, CTCF, and insulator function

Insulators play important roles in establishing boundaries between heterochromatin and euchromatin, as well as protecting genes from the influence of regulatory elements in

their chromatin environment⁶⁴. A potential link between PARP-1 and insulator function was made with the identification of CCCTC-binding factor (CTCF) (Fig 1.4E), a chromatin insulator protein, as a target of PARylation and PAR-dependent regulation⁶⁵. Yu et al. (2004) detected PARs at the imprinting control region (ICR) of the *Igf2-H19* locus, which contains a CTCF-binding element and functions as an insulator in a parent-of-origin-specific manner⁶⁶. Interestingly, the PAR signal is only detected on the maternal-origin ICR, which is hypomethylated, associated with CTCF, and functions as an insulator, but not on the paternal-origin ICR, which is hypermethylated, not associated with CTCF, and non-functional as an insulator. Yu et al. also showed that CTCF is PARylated, with a PARylation signal corresponding to most CTCF-binding sites, as demonstrated by ChIP-chip assays. Although PARylation does not affect CTCF DNA binding activity, it is required for CTCF-dependent insulator function, as shown by insulator-trap assays. Another study identified PARP-1 as a CTCF-associated protein using immunoprecipitation followed by proteomics and mass spectrometry, providing additional evidence for a functional link between these two proteins. Together, these studies provide initial evidence suggesting a role for PARP-1 in insulator function, but more studies are needed to confirm these findings, elucidate the molecular mechanisms, and determine the biological contexts in which they might apply.

Interestingly, CTCF can activate PARP-1 enzymatic activity in the absence of DNA *in vitro*⁶⁷. If this holds true *in vivo*, it may serve as a potential mechanism for regulating PARylation at CTCF binding sites. Moreover, Dnmt1 has been identified to associate with CTCF protein. It colocalizes with CTCF and PARylated PARP-1 on specific CTCF target locus⁶⁷. As discussed above, Dnmt1 activity is inhibited by noncovalent interaction with PAR polymers. Therefore, presence of PARP-1 protects the CTCF-bound CpG DNA sequences from being methylated by Dnmt1. Further studies will shed light on understanding the interplay between PARP-1, CTCF, and DNA methylation, as well as its effect on gene transcriptional regulation.

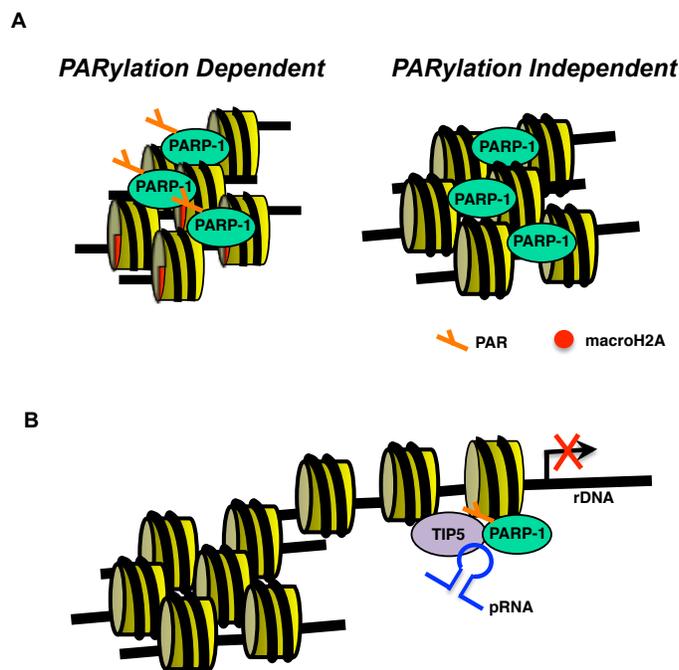


Figure 1.5. PARP-1 functions in heterochromatin

A) PARylation -dependent and -independent functions of PARP-1 in regulating heterochromatin.

B) PARP-1 regulation of rDNA transcription mediated by TIP5 and pRNA.

1.5. PARP-1 functions in heterochromatin

The previous examples of PARP-1 function in gene regulation were related predominantly to euchromatic regions of the genome, where PARP-1 is generally associated with transcriptionally active genes. A considerable body of evidence, however, also supports a role for PARP-1 in regulating chromatin in heterochromatic regions of the genome, which tend to be more transcriptionally repressed. In this section, we discuss these aspects of PARP-1 function.

1.5.1. A link between PARP-1 function and heterochromatin

The earliest studies of PARP-1 function in heterochromatin came from *Drosophila*, which has only two PARP-related genes, homologs of PARP-1 (dPARP) and tankyrase^{15,29}. The *dPARP* gene localizes to a centromeric heterochromatin region spanning more than 150 kb, which is enriched in transposons. Disruption of the *dPARP*

gene by an insertion mutation near its upstream promoter, which inhibits expression of dPARP, causes abnormalities in heterochromatin formation during *Drosophila* development. In *dPARP* mutant strains, sensitivity to micrococcal nuclease in heterochromatic regions is dramatically increased, and is accompanied by a more homogenous nuclear morphology with a loss of the distinct chromocenter and nucleolus normally observed in wild-type strains²⁹. These results suggest a role for dPARP in the proper formation or maintenance of heterochromatin (Fig 1.5A).

These in vivo studies are supported by biochemical studies showing that PARP-1 can promote the formation of compact, repressive chromatin structures, in spite of PARP-1's clear role in supporting open chromatin in euchromatic regions. For example, in studies with purified, reconstituted chromatin, PARP-1 promotes the compaction of chromatin in the absence of NAD⁺, as demonstrated by both MNase digestion and atomic force microscopy⁶⁸. Thus, hypo- or un-PARylated PARP-1 protein, controlled ultimately by NAD⁺ availability⁶⁹, can act to compact chromatin structure. This effect of PARP-1 in promoting the formation of higher-order chromatin structures depends on both its DBD domain, which mediates its binding to nucleosomes, and the catalytic domain, which cooperates with DBD in compacting nucleosomes⁶⁸. This observation may seem to be at odds with the positive effects of PARP-1 on chromatin structure and gene regulation reported in other systems, but may be explained by different NAD⁺ concentrations available in the different experiments. Physiological concentrations of NAD⁺ may be as high as 200 to 300 μM ¹⁶ (although nuclear NAD⁺ concentrations have yet to be measured directly), but PARP-1 activity is tightly regulated, may be directly linked to nuclear NAD⁺ synthesis (see below), and may vary significantly under different physiological conditions. In addition, depending on the chromatin environment, different pools of PARP-1 may have varied catalytic activity and autoPARylation status, providing a potential explanation for the apparently opposite functions of PARP-1 in different chromatin regions.

1.5.2. PARP-1, nucleolar function, and rDNA silencing

The nucleolus is a nuclear compartment where regions of the genome containing hundreds of ribosomal RNA genes (referred to as rDNA) are localized and transcribed by RNA polymerase I⁷⁰. The rDNA is either transcriptionally active (i.e., being transcribed into rRNA, which is further assembled into ribosomes) or transcriptionally silent, forming

heterochromatin structures, which are maintained during cell propagation. A nucleolar pool of PARP-1 has long been observed by immunostaining, with about 40% of PARP-1 protein localizing in the nucleolus⁷¹. Proteomic analyses of the nucleolus has further confirmed a nucleolar localization of PARP-1^{72,73}. In addition, PARP-1 has been shown to interact with nucleolar proteins⁷⁴, including nucleophosmin/B23^{75,76}.

In *Drosophila*, disrupting dPARP enzymatic activity causes abnormalities in nucleolar structure as well as a mislocalization of nucleolar-associated proteins. This indicates a role played by PARP-1 and its activity in the nucleolus⁷⁷. A recent study by Guetg et al. has confirmed the functional link between PARP-1 and rDNA silencing in the nucleolus of mammalian cells (Fig 1.5B)⁷¹. They showed that PARP-1 binds to TIP5, a component of the nucleolar remodeling complex (NoRC), which has previously been shown to play important role in maintaining rDNA silencing⁷⁸. The interaction between PARP-1 and TIP5 is mediated by pRNA, a noncoding RNA synthesized from active rDNA⁷⁹, which is required for NoRC function. PARP-1 protein binds to silenced rDNA and is required for the maintenance of rDNA silencing. RNAi-mediated knockdown of PARP-1 significantly increases the levels of 45S pre-rRNA. Furthermore, the localization of PARP-1 with rDNA occurs in mid-late S phase, after silent rDNA are replicated, indicating a role of PARP-1 in re-establishing heterochromatin in the rDNA locus after DNA replication. PARylation is required for PARP-1-mediated rDNA silencing and silent rDNA chromatin is PARylated, providing an additional link between PARP-1 and the regulation of rDNA. More studies are needed, however, to fully understand the detailed mechanisms of this regulation. Specifically, it is unclear whether PARP-1 and PARylation establish or maintain rDNA silencing through direct modulation of heterochromatin structure or by regulating the RNA polymerase I transcription machinery. Moreover, the functional effects of PARylation on nucleolar proteins is unclear. For example, TIP5 and other heterochromatin proteins are substrates for PARylation by PARP-1, but the role this modification plays in rDNA silencing is unclear. Identifying the key functional targets of PARylation in the nucleolus will provide additional mechanistic insights into PARP-1-mediated rDNA silencing⁸⁰. On the other hand, previous high resolution immunohistochemical studies have demonstrated that nucleolar PARP-1 is more concentrated in dense fibrillar foci, the site of nucleolar transcription. The

localization of PARP-1 to the nucleolus was found to be sensitive to an RNA polymerase I inhibitor⁸¹, indicating a connection between PARP and active rDNA transcription in the nucleolus. It remains to be determined whether these seemingly conflicting observations are due to cell type or cell state difference in various studies. Moreover, more studies are required to uncover the potential more complicated interplay between rDNA transcription and PARP-1 as suggested in the studies described above.

1.5.3. PARP-1, X chromosome inactivation, and macroH2A

Inactivation of the X chromosome in female mammals (“X inactivation”) is a process by which one of the two copies of the X chromosome is rendered transcriptionally inactive by assembly into heterochromatin⁸². PARP-1 has been linked to X inactivation, with the first evidence supporting this role coming from mouse genetic studies examining functional interplay between PARP-1 and PARP-2. Interestingly, *PARP-1^{+/-}/PARP-2^{-/-}* mutant mice exhibit female-specific embryonic lethality due to X chromosome instability^{15,83}. Additional studies have shown that PARP-1 depletion causes de-repression of a GFP reporter gene integrated in the inactive X chromosome (*Xi*) in female mouse embryonic fibroblasts⁸⁴. These studies have suggested a link between PARP-1 and X inactivation, but more studies are needed to confirm a direct role for PARP-1, as well as the underlying mechanisms.

If PARP-1 does play a role in X inactivation, what might the potential mechanisms be? One possibility might be functional interactions with macroH2A⁸⁴, a vertebrate-specific histone variant that contains a histone region homologous to canonical histone H2A^{85,86}, as well as a large carboxyl-terminal non-histone domain called a macrodomain. The macrodomain of the macroH2A1.1 isoform functions as a PAR-binding module, as noted above. MacroH2A is highly enriched in heterochromatin regions of the genome, including the *Xi* and senescence-associated heterochromatin, where it plays an important role in maintaining a compact and repressive chromatin configuration.

Recent studies have provided clues of how PARP-1 and macroH2A may functionally interact. For example, Timinszky et al. used a combination of biochemical, structural, and cell-based assays to show that macroH2A1.1 functions as a PAR sensor, by binding PAR through its macrodomain¹⁹. The outcome is a rearrangement of chromatin structure, leading to the formation of compact chromatin regions. Although these observations were

made in the context of a DNA damage response, the underlying mechanisms may be applicable to the regulatory mechanisms controlling X inactivation. Alternatively, PARP-1 may act directly to compact chromatin, independent of its catalytic activity, as discussed above. In this regard, macroH2A has been shown to be able to inhibit PARP-1 activity *in vitro*⁸⁴, which is consistent with the observation that heterochromatin-associated PARP-1 tends to be less active. Further studies are needed, however, to fully test this model.

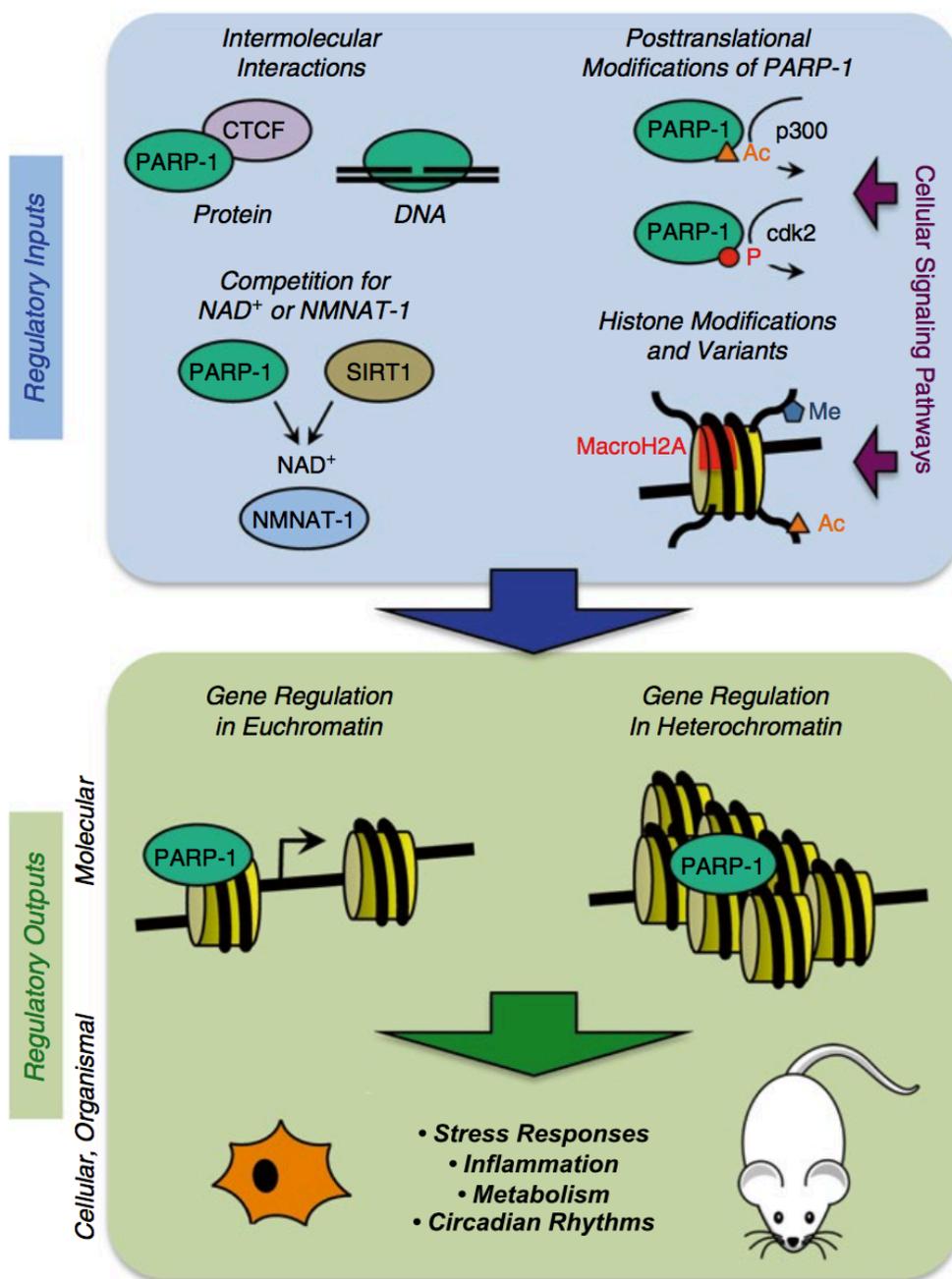


Figure 1.6. Regulatory inputs and outputs for PARP-1

A number of distinct “regulatory inputs” modulate the activity and localization of PARP-1 (*top/blue*). Likewise, PARP-1 functions to promote “regulatory outputs” that

control a wide variety of molecular, cellular, and organismal outcomes (*bottom/green*), as illustrated in the figure and described in the text.

1.6. Regulation of PARP-1 Localization and Activity in Chromatin

As discussed above, PARP-1 plays a wide variety of regulatory roles in chromatin, which in many cases are dependent on the chromatin context and cellular state. The appropriate activity of PARP-1 in different conditions is controlled through a variety of mechanisms involving DNA conformations, nucleosome conformations or composition, chromatin states (including the repertoire of histone modifications and chromatin-associated proteins), posttranslational modifications, cellular signaling pathways, and cellular metabolic status^{3,8}. Thus, PARP-1 and PARylation are integrated into a greater control network, enabling context-dependent modulation of chromatin by PARP-1. Understanding how various combinations of these different components regulate PARP-1 activity will aid in understanding the mechanisms by which PARP-1 senses and controls the chromatin environment. In this section, we discuss about the mechanisms controlling PARP-1 localization to chromatin and enzymatic activity.

1.6.1. Regulation of PARP-1 through interactions with nucleosomes

PARP-1's enzymatic activity can be stimulated by interactions with nucleosomes¹⁵, a process that can be modulated by the presence of histone modifications, histone variants, and higher-order nucleosome interactions ([Fig. 1.6, top](#)). One example of how histone modifications can modulate PARP-1 catalytic activity is observed in *Drosophila* at the *Hsp70* promoter upon heat shock. Heat shock factor binds to its response elements in the promoter and recruits the histone acetyltransferase Tip60 to acetylate histone H2A on lysine 5. This stimulates the catalytic activity of promoter-localized dPARP, which triggers it to dissociate from the promoter nucleosomes and spread across the gene body⁴⁴. Likewise, histone variants can also modulate PARP-1 catalytic activity, as discussed above for macroH2A⁸⁴. Another example is *Drosophila* H2Av, a homolog of mammalian histone variants H2Az and H2Ax⁸⁷, which has been shown to colocalize with dPARP in nuclei as shown by immunofluorescence, as well as at the *Hsp70* promoter by ChIP⁸⁸. Depletion of H2Av causes a mislocalization of dPARP at the *Hsp70* promoter, indicating proper localization of dPARP requires H2Av. Furthermore, H2Av phosphorylation is

required for dPARP activation, as well as dPARP-mediated heat shock-induced gene transcription and genotoxic stress responses. The signal that triggers H2Av phosphorylation, however, is still unknown. Interestingly, results from the same studies using biochemical assays suggest that instead of activating dPARP directly, phosphorylated H2Av modulates dPARP activity in a nucleosome-dependent manner, probably by modulating nucleosome conformation and facilitating interactions between PARP and histone H4, which ultimately activates dPARP enzymatic activity. At present, the functional links between histone H2A variants and PARP-1 have not been well characterized in mammalian cells, but PARP-1 and H2Az both localize to gene promoters, which would at least provide the opportunity for functional interactions.

The regulation of dPARP by H2Av indicates that nucleosome composition and conformation can have important effects on PARP-1 binding and activity. In this regard, it seems likely that other components of nucleosomes or modulators of nucleosome structure may have similar effects on PARP-1.

1.6.2. Regulation of PARP-1 by non-histone protein binding partners and post-translational modifications

Non-histone, chromatin-associated proteins can also regulate PARP-1 activity and function, either through direct interactions with PARP-1 or by post-translationally modifying PARP-1. For example, as discussed above, interactions with CTCF can activate PARP-1 independent of DNA^{1,15,67}. A host of other PARP-1 interactions partners can do the same, such as DNA binding transcription factor YY1^{1,89}. More broadly, post-translational modifications can regulate various aspects of PARP-1 function, including its catalytic activity, binding affinity for other proteins or chromatin, and stability⁸. PARP-1 is subjected to wide array of post-translational modifications, including PARylation, phosphorylation, acetylation, ubiquitylation, and sumoylation. These have been reviewed extensively elsewhere (Luo and Krishnakumar reviews)^{8,23}; we provide a few examples here for illustrative purposes.

PARP-1 is acetylated both *in vitro* and *in vivo* by p300^{8,17}, a chromatin-associated protein acetyltransferase, upon proinflammatory stimuli in macrophages. Acetylation of PARP-1 stabilizes its interaction with the proinflammatory transcription factor p50, and is

required for PARP-1 function as a PARylation-independent coactivator of NF- κ B, as shown by luciferase reporter assay using reporter gene driven by endogenous proinflammatory responsive gene promoters¹⁷. Moreover, using co-immunoprecipitation, it was shown that PARP-1 protein interacts with members of HDAC class I, and is potentially deacetylated by HDAC1, HDAC2 or HDAC3, since overexpressing HDAC1-3 proteins decreases acetylation level of PARP-1 protein; also, the coactivator activity of PARP-1 is negatively regulated by HDAC1-3.

PARylation of PARP-1, either through an automodification reaction or by other PARPs, can also regulate PARP-1 function. Extensive autoPARylation of PARP-1, as observed under conditions of DNA damage, decreases PARP-1 binding to chromatin, although it is unclear whether the levels of auto PARylation achieved under normal physiological conditions would be sufficient to achieve the same effect. AutoPARylation of PARP-1 is rapidly reversed by PARG⁹⁰, which acts digest PAR chains. Interestingly, PARG shuttles between the nucleus and cytoplasm^{91,92}, providing a potential mechanism for modulating its effects on PARylated PARP-1, as well as other PARylated nuclear proteins. Accumulating evidence suggests a function for PARG in transcriptional regulation^{8,90,93}, in part through catabolism of PAR in the context of chromatin.

As these examples, as well as others described elsewhere, clearly illustrate, PARP-1 activity and function can be modulated by protein binding partners and reversible post-translational modifications. The post-translational modifications of PARP-1 are, in turn, controlled in many cases as endpoints of cellular signaling pathways.

1.6.3.Regulation of PARP-1 by cellular signaling pathways

PARP-1 is also regulated by various cellular signaling pathways, which are activated or inhibited in response to external or internal cues⁸. Kinases are critical components of many cellular signaling pathways, functioning to transmit signals either by directly phosphorylating downstream targets, or by interacting with them and allosterically modulating their biochemical properties. PARP-1 is regulated by various kinases through both mechanisms.

In human breast cancer cells, PARP-1 is phosphorylated in its catalytic domain on Ser785 and Ser786 by the hormone-activated cyclin-dependent kinase CDK2 in response to treatment with the synthetic progestin R5020. Phosphorylation of PARP-1 by CDK2 is

required for the R5020-induced, rapid and transient activation of PARP-1, which in turn enhances progesterin-induced gene transcription³⁶. PARP-1 is also phosphorylated by the extracellular signal-regulated kinases, ERK1/2, on Ser372 and Thr373, which potentiates PARP-1-dependent PARylation after DNA damage⁹⁴. In addition, the stress-activated kinase JNK1 can phosphorylate PARP-1 on undetermined residues to stimulate PARP-1's enzymatic activity⁹⁵. Phosphorylation may also have an inhibitory effect on PARP-1. For example, phosphorylation of PARP-1 by protein kinase C causes decreased DNA binding capacity and catalytic activity of PARP-1.

More studies are needed to explore the detailed mechanisms by which phosphorylation of PARP-1 can alter its enzymatic activity. It may alter PARP-1's ability to bind NAD⁺ or its target proteins, or it may affect its catalytic properties, both of which may be the result of phosphorylation-induced changes in PARP-1 structure or chemistry, or allosteric changes in PARP-1 induced by the kinase. With respect to the latter, PARP-1 activity is stimulated by association with phosphorylated ERK2⁹⁶. Interestingly, activated PARP-1 in turn promotes phosphorylation of the transcription factor Elk1 by ERK2, which regulates Elk1-mediated gene transcription^{94,96}. This suggests a feed-back loop between PARP-1 and kinases. Together, interactions of PARP-1 with kinases, as well as the resulting phosphorylation of PARP-1, provides a number of avenues for the regulation of PARP-1 activity in chromatin, as well as connections to cellular signaling pathways.

1.6.4. Modulation of PAR activity by nuclear NAD⁺ metabolism

As shown in biochemical assays, the available NAD⁺ concentration dramatically affects the kinetics of PARylation by PARP-1, with higher concentrations of NAD⁺ promoting the length of PAR polymer synthesis. The presence of a separate nuclear pool of NAD⁺ may have important regulatory functions, as implied by the exclusive nuclear localization of the most predominant mammalian NAD⁺ synthase, NMNAT1. However, largely being limited by the availability of cell imaging technology allowing a high-resolution in vivo real-time visualization of NAD⁺ subcellular distribution, it remains to be determined how NAD⁺ distributes under different cell states.

In addition to serving to supply NAD⁺ substrate, NMNAT1 also acts to modulate PARP-1 directly. NMNAT1 has been shown to be able to bind to PARP-1 and stimulate

its PARylation activity, while phosphorylation of NMNAT1 by protein kinase C decreases its effect on PARP-1. Furthermore, Zhang et al. have shown that NMNAT1 is recruited to gene promoters by PARP-1, where it not only locally supplies substrate for PARP-1, but also allosterically enhances its PARylation activity in a NAD⁺ production-independent manner. This interplay between NMNAT1 with PARP-1 seems to take place in a context-dependent manner, since only a subset of genes are regulated by both factors in the same way. Thus, it remains to understand how the recruitment of NMNAT1 by PARP-1, the NAD⁺ synthase activity and its effect on PARP-1 are regulated.

Moreover, it is speculated that other NAD⁺-dependent nuclear enzymes, such as protein deacetylase SIRT1, might compete with PARP-1 for NAD⁺ supply or protein-protein interaction with NMNAT1, thus potentially adding another layer of regulation by nuclear NAD⁺ metabolism. However, the relation between PARP-1, NAD⁺ synthase, and NAD⁺-dependent nuclear enzymes as well as the mechanisms remains to be clarified.

1.7. Conclusions

As discussed above, PARP-1 plays important role in regulating chromatin structure in both euchromatin and heterochromatin regions. It regulates gene transcription through maintaining an open chromatin conformation by regulating various aspects of chromatin including linker histone, core histones, histone posttranslational modification, chromatin remodelers as well as DNA methylation. Remarkably, PARP-1 also acts as a key modulator in heterochromatin regions. Chromatin modulation functions of PARP-1 is affected by its chromatin environment as well as cell signaling. Although a significant progress has been made in understanding chromatin regulation by PARP-1, more studies are needed to fully explore the detailed mechanisms as well as their implications in physiological contexts, such as dissecting the functions played by PARP-1 protein as well as its catalytic activity in various biological processes, exploring the regulatory mechanisms of a context-dependent function of PARP-1 in chromatin regulation, as well as understanding how chromatin structure modulation is translated to effects on gene transcription.

REFERENCES

1. Kraus, W.L. Transcriptional control by PARP-1: chromatin modulation, enhancer-binding, coregulation, and insulation. *Curr Opin Cell Biol* **20**, 294-302 (2008).
2. Kraus, W.L. & Lis, J.T. PARP goes transcription. *Cell* **113**, 677-83 (2003).
3. Gibson, B.A. & Kraus, W.L. New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nat Rev Mol Cell Biol* **13**, 411-24 (2012).
4. Langelier, M.F., Planck, J.L., Roy, S. & Pascal, J.M. Structural basis for DNA damage-dependent poly(ADP-ribosylation) by human PARP-1. *Science* **336**, 728-32 (2012).
5. Loeffler, P.A. *et al.* Structural studies of the PARP-1 BRCT domain. *BMC Struct Biol* **11**, 37 (2011).
6. Schreiber, V., Dantzer, F., Ame, J.C. & de Murcia, G. Poly(ADP-ribose): novel functions for an old molecule. *Nat Rev Mol Cell Biol* **7**, 517-28 (2006).
7. Ji, Y. & Tulin, A.V. The roles of PARP1 in gene control and cell differentiation. *Curr Opin Genet Dev* **20**, 512-8 (2010).
8. Luo, X. & Kraus, W.L. On PAR with PARP: cellular stress signaling through poly(ADP-ribose) and PARP-1. *Genes Dev* **26**, 417-32 (2012).
9. Devalaraja-Narashimha, K. & Padanilam, B.J. PARP1 deficiency exacerbates diet-induced obesity in mice. *J Endocrinol* **205**, 243-52 (2010).
10. Bai, P. *et al.* PARP-1 inhibition increases mitochondrial metabolism through SIRT1 activation. *Cell Metab* **13**, 461-8 (2011).
11. Luo, X. & Kraus, W.L. A one and a two ... expanding roles for poly(ADP-ribose) polymerases in metabolism. *Cell Metab* **13**, 353-5 (2011).
12. Yang, F., Baumann, C. & De La Fuente, R. Persistence of histone H2AX phosphorylation after meiotic chromosome synapsis and abnormal centromere cohesion in poly (ADP-ribose) polymerase (Parp-1) null oocytes. *Dev Biol* **331**, 326-38 (2009).
13. Asher, G. *et al.* Poly(ADP-ribose) polymerase 1 participates in the phase entrainment of circadian clocks to feeding. *Cell* **142**, 943-53 (2010).
14. D'Amours, D., Desnoyers, S., D'Silva, I. & Poirier, G.G. Poly(ADP-ribosylation) reactions in the regulation of nuclear functions. *Biochem J* **342** (Pt 2), 249-68 (1999).
15. Kim, M.Y., Zhang, T. & Kraus, W.L. Poly(ADP-ribosylation) by PARP-1: 'PAR-laying' NAD⁺ into a nuclear signal. *Genes Dev* **19**, 1951-67 (2005).

16. Kim, M.Y., Mauro, S., Gevry, N., Lis, J.T. & Kraus, W.L. NAD⁺-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1. *Cell* **119**, 803-14 (2004).
17. Hassa, P.O. *et al.* Acetylation of poly(ADP-ribose) polymerase-1 by p300/CREB-binding protein regulates coactivation of NF-kappaB-dependent transcription. *J Biol Chem* **280**, 40450-64 (2005).
18. Ahel, I. *et al.* Poly(ADP-ribose)-binding zinc finger motifs in DNA repair/checkpoint proteins. *Nature* **451**, 81-5 (2008).
19. Timinszky, G. *et al.* A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. *Nat Struct Mol Biol* **16**, 923-9 (2009).
20. Karras, G.I. *et al.* The macro domain is an ADP-ribose binding module. *EMBO J* **24**, 1911-20 (2005).
21. Aravind, L. The WWE domain: a common interaction module in protein ubiquitination and ADP ribosylation. *Trends Biochem Sci* **26**, 273-5 (2001).
22. Wang, Z. *et al.* Recognition of the iso-ADP-ribose moiety in poly(ADP-ribose) by WWE domains suggests a general mechanism for poly(ADP-ribosyl)ation-dependent ubiquitination. *Genes Dev* **26**, 235-40 (2012).
23. Krishnakumar, R. & Kraus, W.L. The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. *Mol Cell* **39**, 8-24 (2010).
24. Hottiger, M.O., Hassa, P.O., Luscher, B., Schuler, H. & Koch-Nolte, F. Toward a unified nomenclature for mammalian ADP-ribosyltransferases. *Trends Biochem Sci* **35**, 208-19 (2010).
25. Pavri, R. *et al.* PARP-1 determines specificity in a retinoid signaling pathway via direct modulation of mediator. *Mol Cell* **18**, 83-96 (2005).
26. Huletsky, A. *et al.* The effect of poly(ADP-ribosyl)ation on native and H1-depleted chromatin. A role of poly(ADP-ribosyl)ation on core nucleosome structure. *J Biol Chem* **264**, 8878-86 (1989).
27. Tulin, A. & Spradling, A. Chromatin loosening by poly(ADP)-ribose polymerase (PARP) at *Drosophila* puff loci. *Science* **299**, 560-2 (2003).
28. Krishnakumar, R. & Kraus, W.L. PARP-1 regulates chromatin structure and transcription through a KDM5B-dependent pathway. *Mol Cell* **39**, 736-49 (2010).

29. Tulin, A., Stewart, D. & Spradling, A.C. The *Drosophila* heterochromatic gene encoding poly(ADP-ribose) polymerase (PARP) is required to modulate chromatin structure during development. *Genes Dev* **16**, 2108-19 (2002).
30. Krishnakumar, R. *et al.* Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. *Science* **319**, 819-21 (2008).
31. Poirier, G.G., de Murcia, G., Jongstra-Bilen, J., Niedergang, C. & Mandel, P. Poly(ADP-ribosylation) of polynucleosomes causes relaxation of chromatin structure. *Proc Natl Acad Sci U S A* **79**, 3423-7 (1982).
32. Huletsky, A. *et al.* Sequential ADP-ribosylation pattern of nucleosomal histones. ADP-ribosylation of nucleosomal histones. *Eur J Biochem* **146**, 277-85 (1985).
33. Wesierska-Gadek, J. & Sauermann, G. The effect of poly(ADP-ribose) on interactions of DNA with histones H1, H3 and H4. *Eur J Biochem* **173**, 675-9 (1988).
34. Ju, B.G. *et al.* A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. *Science* **312**, 1798-802 (2006).
35. Ju, B.G. & Rosenfeld, M.G. A breaking strategy for topoisomerase IIbeta/PARP-1-dependent regulated transcription. *Cell Cycle* **5**, 2557-60 (2006).
36. Wright, R.H. *et al.* CDK2-dependent activation of PARP-1 is required for hormonal gene regulation in breast cancer cells. *Genes Dev* **26**, 1972-83 (2012).
37. Mathis, G. & Althaus, F.R. Release of core DNA from nucleosomal core particles following (ADP-ribose)_n-modification in vitro. *Biochem Biophys Res Commun* **143**, 1049-54 (1987).
38. Althaus, F.R. *et al.* Histone shuttling by poly ADP-ribosylation. *Mol Cell Biochem* **138**, 53-9 (1994).
39. Adamietz, P. & Rudolph, A. ADP-ribosylation of nuclear proteins in vivo. Identification of histone H2B as a major acceptor for mono- and poly(ADP-ribose) in dimethyl sulfate-treated hepatoma AH 7974 cells. *J Biol Chem* **259**, 6841-6 (1984).
40. Kreimeyer, A., Wielckens, K., Adamietz, P. & Hilz, H. DNA repair-associated ADP-ribosylation in vivo. Modification of histone H1 differs from that of the principal acceptor proteins. *J Biol Chem* **259**, 890-6 (1984).
41. Ogata, N., Ueda, K., Kawaichi, M. & Hayaishi, O. Poly(ADP-ribose) synthetase, a main acceptor of poly(ADP-ribose) in isolated nuclei. *J Biol Chem* **256**, 4135-7 (1981).

42. Realini, C.A. & Althaus, F.R. Histone shuttling by poly(ADP-ribosylation). *J Biol Chem* **267**, 18858-65 (1992).
43. Petesch, S.J. & Lis, J.T. Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at Hsp70 loci. *Cell* **134**, 74-84 (2008).
44. Petesch, S.J. & Lis, J.T. Activator-induced spread of poly(ADP-ribose) polymerase promotes nucleosome loss at Hsp70. *Mol Cell* **45**, 64-74 (2012).
45. Petesch, S.J. & Lis, J.T. Overcoming the nucleosome barrier during transcript elongation. *Trends Genet* **28**, 285-94 (2012).
46. Doege, C.A. *et al.* Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* **488**, 652-5 (2012).
47. Langst, G. & Becker, P.B. Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. *J Cell Sci* **114**, 2561-8 (2001).
48. Sala, A. *et al.* The nucleosome-remodeling ATPase ISWI is regulated by poly-ADP-ribosylation. *PLoS Biol* **6**, e252 (2008).
49. Corona, D.F. & Tamkun, J.W. Multiple roles for ISWI in transcription, chromosome organization and DNA replication. *Biochim Biophys Acta* **1677**, 113-9 (2004).
50. Corona, D.F. *et al.* ISWI regulates higher-order chromatin structure and histone H1 assembly in vivo. *PLoS Biol* **5**, e232 (2007).
51. Gottschalk, A.J. *et al.* Poly(ADP-ribosylation) directs recruitment and activation of an ATP-dependent chromatin remodeler. *Proc Natl Acad Sci U S A* **106**, 13770-4 (2009).
52. Robertson, K.D. DNA methylation and chromatin - unraveling the tangled web. *Oncogene* **21**, 5361-79 (2002).
53. Zardo, G. & Caiafa, P. The unmethylated state of CpG islands in mouse fibroblasts depends on the poly(ADP-ribosylation) process. *J Biol Chem* **273**, 16517-20 (1998).
54. de Capoa, A. *et al.* Reduced levels of poly(ADP-ribosylation) result in chromatin compaction and hypermethylation as shown by cell-by-cell computer-assisted quantitative analysis. *FASEB J* **13**, 89-93 (1999).
55. Caiafa, P., Guastafierro, T. & Zampieri, M. Epigenetics: poly(ADP-ribosylation) of PARP-1 regulates genomic methylation patterns. *FASEB J* **23**, 672-8 (2009).

56. Svedruzic, Z.M. Mammalian cytosine DNA methyltransferase Dnmt1: enzymatic mechanism, novel mechanism-based inhibitors, and RNA-directed DNA methylation. *Curr Med Chem* **15**, 92-106 (2008).
57. Caiafa, P. & Zlatanova, J. CCCTC-binding factor meets poly(ADP-ribose) polymerase-1. *J Cell Physiol* **219**, 265-70 (2009).
58. Zampieri, M. *et al.* Parp1 Localizes within the Dnmt1 Promoter and Protects Its Unmethylated State by Its Enzymatic Activity. *Plos One* **4**(2009).
59. Reale, A., Matteis, G.D., Galleazzi, G., Zampieri, M. & Caiafa, P. Modulation of DNMT1 activity by ADP-ribose polymers. *Oncogene* **24**, 13-9 (2005).
60. Williams, K., Christensen, J. & Helin, K. DNA methylation: TET proteins-guardians of CpG islands? *EMBO Rep* **13**, 28-35 (2012).
61. Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev* **25**, 2436-52 (2011).
62. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-5 (2009).
63. Ciccarone, F. *et al.* Poly(ADP-ribosyl)ation Acts in the DNA Demethylation of Mouse Primordial Germ Cells Also with DNA Damage-Independent Roles. *PLoS One* **7**, e46927 (2012).
64. Barkess, G. & West, A.G. Chromatin insulator elements: establishing barriers to set heterochromatin boundaries. *Epigenomics* **4**, 67-80 (2012).
65. Dunn, K.L. & Davie, J.R. The many roles of the transcriptional regulator CTCF. *Biochem Cell Biol* **81**, 161-7 (2003).
66. Yu, W. *et al.* Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nat Genet* **36**, 1105-10 (2004).
67. Zampieri, M. *et al.* ADP-ribose polymers localized on Ctfp-Parp1-Dnmt1 complex prevent methylation of Ctfp target sites. *Biochem J* **441**, 645-52 (2012).
68. Wacker, D.A. *et al.* The DNA binding and catalytic domains of poly(ADP-ribose) polymerase 1 cooperate in the regulation of chromatin structure and transcription. *Mol Cell Biol* **27**, 7475-85 (2007).
69. Alvarez-Gonzalez, R., Pacheco-Rodriguez, G. & Mendoza-Alvarez, H. Enzymology of ADP-ribose polymer synthesis. *Mol Cell Biochem* **138**, 33-7 (1994).

70. Boisvert, F.M., van Koningsbruggen, S., Navascues, J. & Lamond, A.I. The multifunctional nucleolus. *Nat Rev Mol Cell Biol* **8**, 574-85 (2007).
71. Guetg, C., Scheifele, F., Rosenthal, F., Hottiger, M.O. & Santoro, R. Inheritance of silent rDNA chromatin is mediated by PARP1 via noncoding RNA. *Mol Cell* **45**, 790-800 (2012).
72. Andersen, J.S. *et al.* Directed proteomic analysis of the human nucleolus. *Curr Biol* **12**, 1-11 (2002).
73. Scherl, A. *et al.* Functional proteomic analysis of human nucleolus. *Mol Biol Cell* **13**, 4100-9 (2002).
74. Kotova, E., Jarnik, M. & Tulin, A.V. Poly (ADP-ribose) polymerase 1 is required for protein localization to Cajal body. *PLoS Genet* **5**, e1000387 (2009).
75. Chan, P.K. Characterization and cellular localization of nucleophosmin/B23 in HeLa cells treated with selected cytotoxic agents (studies of B23-translocation mechanism). *Exp Cell Res* **203**, 174-81 (1992).
76. Meder, V.S., Boeglin, M., de Murcia, G. & Schreiber, V. PARP-1 and PARP-2 interact with nucleophosmin/B23 and accumulate in transcriptionally active nucleoli. *J Cell Sci* **118**, 211-22 (2005).
77. Boamah, E.K., Kotova, E., Garabedian, M., Jarnik, M. & Tulin, A.V. Poly(ADP-Ribose) polymerase 1 (PARP-1) regulates ribosomal biogenesis in *Drosophila* nucleoli. *PLoS Genet* **8**, e1002442 (2012).
78. Mayer, C., Schmitz, K.M., Li, J., Grummt, I. & Santoro, R. Intergenic transcripts regulate the epigenetic state of rRNA genes. *Mol Cell* **22**, 351-61 (2006).
79. Mayer, C., Neubert, M. & Grummt, I. The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus. *EMBO Rep* **9**, 774-80 (2008).
80. Guetg, C. & Santoro, R. Noncoding RNAs link PARP1 to heterochromatin. *Cell Cycle* **11**, 2217-8 (2012).
81. Desnoyers, S., Kaufmann, S.H. & Poirier, G.G. Alteration of the nucleolar localization of poly(ADP-ribose) polymerase upon treatment with transcription inhibitors. *Exp Cell Res* **227**, 146-53 (1996).

82. Brockdorff, N. Chromosome silencing mechanisms in X-chromosome inactivation: unknown unknowns. *Development* **138**, 5057-65 (2011).
83. Menissier de Murcia, J. *et al.* Functional interaction between PARP-1 and PARP-2 in chromosome stability and embryonic development in mouse. *EMBO J* **22**, 2255-63 (2003).
84. Nusinow, D.A. *et al.* Poly(ADP-ribose) polymerase 1 is inhibited by a histone H2A variant, MacroH2A, and contributes to silencing of the inactive X chromosome. *J Biol Chem* **282**, 12851-9 (2007).
85. Costanzi, C., Stein, P., Worrad, D.M., Schultz, R.M. & Pehrson, J.R. Histone macroH2A1 is concentrated in the inactive X chromosome of female preimplantation mouse embryos. *Development* **127**, 2283-9 (2000).
86. Changolkar, L.N. & Pehrson, J.R. macroH2A1 histone variants are depleted on active genes but concentrated on the inactive X chromosome. *Mol Cell Biol* **26**, 4410-20 (2006).
87. Redon, C. *et al.* Histone H2A variants H2AX and H2AZ. *Curr Opin Genet Dev* **12**, 162-9 (2002).
88. Kotova, E. *et al.* Drosophila histone H2A variant (H2Av) controls poly(ADP-ribose) polymerase 1 (PARP1) activation in chromatin. *Proc Natl Acad Sci U S A* **108**, 6205-10 (2011).
89. Griesenbeck, J., Ziegler, M., Tomilin, N., Schweiger, M. & Oei, S.L. Stimulation of the catalytic activity of poly(ADP-ribosyl) transferase by transcription factor Yin Yang 1. *FEBS Lett* **443**, 20-4 (1999).
90. Frizzell, K.M. *et al.* Global analysis of transcriptional regulation by poly(ADP-ribose) polymerase-1 and poly(ADP-ribose) glycohydrolase in MCF-7 human breast cancer cells. *J Biol Chem* **284**, 33926-38 (2009).
91. Ohashi, S. *et al.* Subcellular localization of poly(ADP-ribose) glycohydrolase in mammalian cells. *Biochem Biophys Res Commun* **307**, 915-21 (2003).
92. Bonicalzi, M.E., Vodenicharov, M., Coulombe, M., Gagne, J.P. & Poirier, G.G. Alteration of poly(ADP-ribose) glycohydrolase nucleocytoplasmic shuttling characteristics upon cleavage by apoptotic proteases. *Biol Cell* **95**, 635-44 (2003).
93. Rapizzi, E., Fossati, S., Moroni, F. & Chiarugi, A. Inhibition of poly(ADP-ribose) glycohydrolase by gallotannin selectively up-regulates expression of proinflammatory genes. *Mol Pharmacol* **66**, 890-8 (2004).

94. Cohen-Armon, M. *et al.* DNA-independent PARP-1 activation by phosphorylated ERK2 increases Elk1 activity: a link to histone acetylation. *Mol Cell* **25**, 297-308 (2007).
95. Zhang, S. *et al.* c-Jun N-terminal kinase mediates hydrogen peroxide-induced cell death via sustained poly(ADP-ribose) polymerase-1 activation. *Cell Death Differ* **14**, 1001-10 (2007).
96. Kauppinen, T.M. *et al.* Direct phosphorylation and regulation of poly(ADP-ribose) polymerase-1 by extracellular signal-regulated kinases 1/2. *Proc Natl Acad Sci U S A* **103**, 7136-41 (2006).

CHAPTER 2

PARP-1 Facilitates Sox2 Binding to Nucleosomes to Maintain a Pluripotency Gene Expression Program in Embryonic Stem Cells

2.1. Summary

Sox2 is a transcription factor that acts to promote a pluripotency gene expression gene program in embryonic stem cells. Here we show using a complementary set of cell-based genomic and biochemical assays that PARP-1, a chromatin-modulating protein, facilitates the binding of Sox2 to a cohort of its genomic binding sites in mouse embryonic stem cells, thus helping to maintain the Sox2 gene expression program. We find that PARP-1-dependent Sox2 binding sites reside in closed regions of chromatin with high nucleosome occupancy and low co-occupancy by partner transcription factors (e.g., Oct4, Nanog). Furthermore, PARP-1 promotes Sox2 binding to nucleosomes at suboptimal sites located in wide minor grooves facing away from the histone octamer by forming a complex with Sox2 on the nucleosome. Our results illustrate how PARP-1 can act at the level of the nucleosome to produce global effects on transcription factor binding and biologically important gene expression outcomes.

2.2. Introduction

PARP-1, or Poly(ADP-Ribose) polymerase 1, is an abundant nucleus protein that is involved in multiple biological pathways ¹. Recently, multiple studies have suggested a link between PARP-1 and pluripotency. PARP-1 is required for reprogramming somatic cells into induced pluripotent stem cells (iPSCs) ². In addition, PARP-1 depletion in embryonic stem cells causes decreased expression of certain pluripotency marks and increased expression of some differentiation genes ³. Despite the strong evidence of PARP-1 playing a role in regulating pluripotency, the mechanism for this remains poorly understood.

As a nucleosome-binding protein, PARP-1 is able to regulate gene transcription through modulating chromatin structure ⁴. PARP-1 blocks histone H1 binding to actively transcribed gene promoters and maintains an open chromatin conformation ^{5 6}. PARP-1 also has an inhibitory effect on histone demethylase KDM5B through PARylation ⁷. In addition to functioning as a chromatin structure modulator, PARP-1 is able to function as a coregulator as well, directly regulating sequence-specific DNA-binding transcription factors such as NF- κ B ⁸ and E2F-1 ⁹. It is thus important to identify which chromatin-associating protein or transcription factor is the regulatory target of PARP-1 in embryonic stem cells in order to better understand the mechanism of PARP-1 regulating pluripotency.

The pluripotency of embryonic stem cell is controlled by a set of “master regulators” including Sox2, Oct4, Nanog, Klf4, etc ¹⁰. The HMG box containing transcription factor Sox2 plays fundamental role in regulating embryonic stem cell identity and differentiation. It binds to consensus DNA sequence and regulates gene transcription ¹¹. Regulation of Sox2 chromatin association is thus critical for controlling the fate of ES cells. Sox2 frequently co-binds with other transcription factors such as Oct4 and BRN2 in a cell type-specific manner ¹². Association with distinct transcription factors is believed to be the mechanism for Sox2 to achieve cell type-specific DNA binding. Despite this, one question remains: in genomic regions where other sequence-specific DNA binding transcription factor binding is lacking, how does Sox2 get access to its binding site, especially to those sites with a nucleosome barrier? It was shown by previous studies that both Sox2 and Oct4

are able to bind to nucleosome^{13 14}. During early stages of somatic cell reprogramming, Oct4, Sox2, and Klf4 are able to access genomic regions with a closed chromatin conformation¹³. In addition, Soufi et al. also showed *in vitro* that both Oct4 and Sox2 are able to bind to nucleosomes, with an affinity comparable to naked DNA¹⁴. However, unlike Oct4, Sox2 interaction with nucleosomes is a mixture of sequence specific and non-specific contacts. In addition, Sox2 interaction with nucleosomes can be efficiently competed by non-specific DNA¹⁴. Therefore, it is still not clear how Sox2 binds specifically and stably to its target sequence in a nucleosome context.

In this study, we identified an unexpected, catalytic activity-independent function of PARP-1 as a “pre-pioneer factor” regulating the chromatin association of Sox2. PARP-1 is specifically required for stabilizing Sox2 binding to genomic regions with higher nucleosome occupancy as well as lower co-localization of other transcription factors. In addition, PARP-1 protein is able to stabilize Sox2 interaction with nucleosomes *in vitro*. PARP-1 therefore safeguards the “stemness” of embryonic stem cells by controlling Sox2 chromatin association. Our study has provided a better understanding of the mechanism of regulating ES cell pluripotency by PARP-1, as well as how the chromatin association of Sox2 is regulated.

2.3. Results

2.3.1. PARP-1 safeguards the “stemness” of embryonic stem cells by transcriptional regulation.

To explore the PARP-1-dependent regulation of ESC biology, we used mouse ESCs (mESCs) with genetic deletion of the *Parp1* gene¹⁵. Alterations in the morphology of the undifferentiated *Parp1*^{-/-} mESCs were not observed compared to wild type cells (data not shown). In addition, the expression level of *Rex1*, which is a sensitive mark of embryonic stem cell pluripotency state, remains unchanged (Fig. 2.1A), indicating the stem cell identity remains essentially the same upon the loss of PARP-1. Despite this, analysis of gene expression by RT-qPCR revealed that PARP-1 depletion results in a significant down-regulation of some pluripotency genes including *Nanog*, while not affecting the expression of other pluripotency factors such as *Oct4* and *Sox2* (Fig. 2.1A). Moreover, many differentiation-related genes were found to be significantly up-regulated

upon PARP-1 depletion (Fig. 2.1A). Importantly, when we induced ES cells to differentiate by forming embryoid bodies, differentiation occurred significantly faster in *Parp1* knockout cells as compared to wild type ES cells, as demonstrated by performing RT-qPCR for 19 differentiation markers (Fig. 2.6C).

To investigate the effect of PARP-1 depletion on ES cell gene transcription on a global scale, we performed RNA-seq using *Parp1* knockout as well as wild type ES cells. Agreeing on the gene specific RT-qPCR results, *Nanog* was significantly down-regulated, with *Oct4* and *Sox2* expression unchanged (Fig. 2.1B, Fig 2.2A). Furthermore, differentiation genes like *Pax6* and *Gata6* were also up-regulated. H3K4me3, which is a marker of active transcription, correlated with gene expression level changes (Fig. 2.1B). We identified 801 genes whose expression levels were significantly increased (defined as PARP-1 positively regulated, $p < 0.01$, fold > 2) in *Parp1* knockout ES cells, and 421 genes with a significant down-regulation (defined as PARP-1 negatively regulated, $p < 0.01$, fold > 2) upon PARP-1 depletion (Fig. 2.1C). Among the genes positively regulated by PARP-1 are the pluripotency genes *Klf4* and *Tbx3* (Fig. 2.2C). Gene ontology analysis showed a significant enrichment of negative regulators of differentiation as well as those that maintain stem cell identity (Fig. 2.3). On the other hand, genes negatively regulated by PARP-1 were enriched with those that are associated with cell differentiation (Fig. 2.3). Consistent with this, PARP-1 negatively regulated genes were enriched with the bivalent histone marks, H3K4me3 and H3K27me3, which is a chromatin signature of differentiation genes in embryonic stem cells (Fig. 2.2A). Interestingly, undifferentiated mESCs have nearly undetectable basal levels of PARP-1-mediated poly(ADP-ribosylation) (Fig. 2.4A). The effects of PARP-1 on gene expression likely did not require its catalytic activity, as the PARP inhibitor PJ34 did not promote ESC differentiation (Fig. 2.4B), suggesting non-catalytic mechanisms for PARP-1-dependent gene regulation.

Together, our results showed that PARP-1 is required for maintaining the transcriptional program of mouse embryonic stem in a PARylation activity-independent manner.

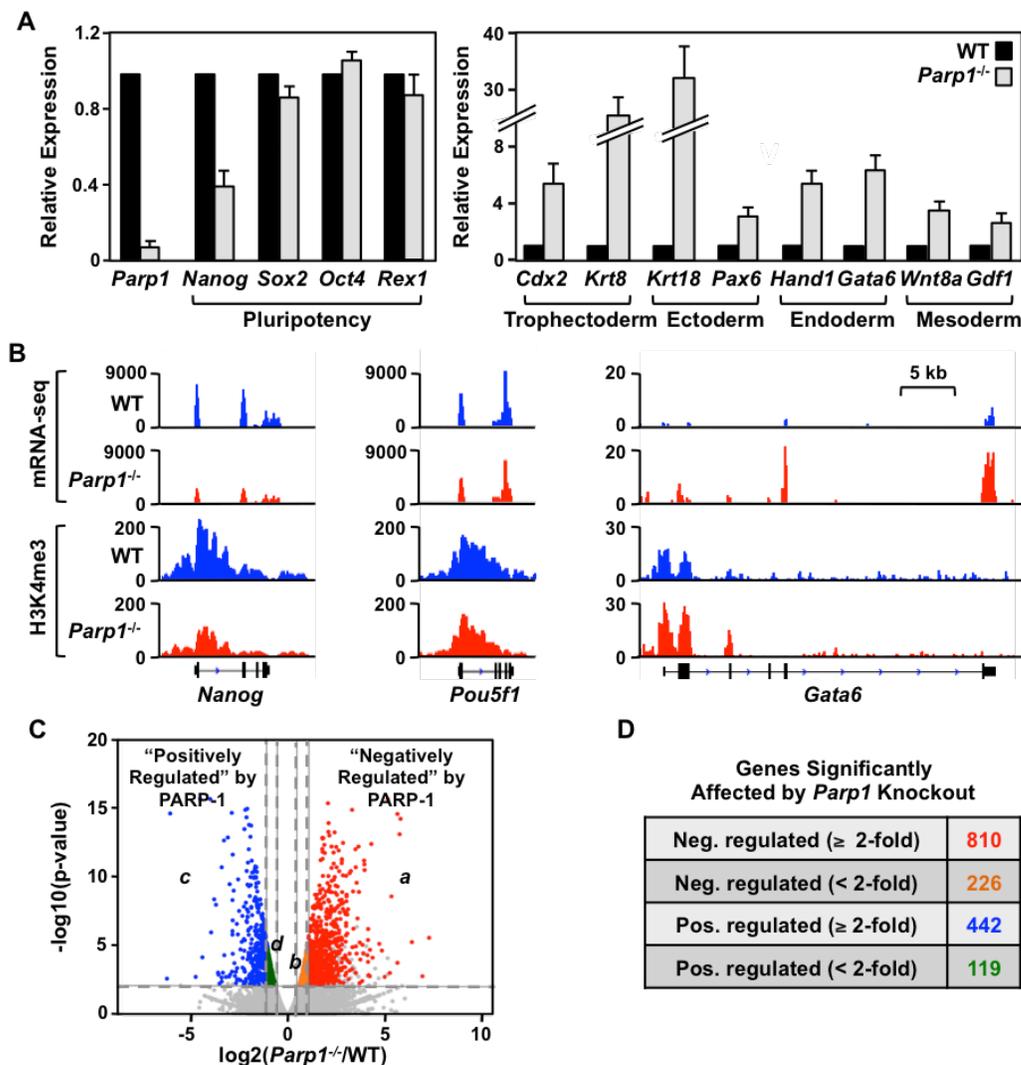


Figure 2.1. Knockout of *Parp1* alters the gene expression program in mESCs.

A) Effect of *Parp1* knockout on the expression of pluripotency-associated genes (*Left*) and differentiation-associated genes (*Right*) in mESCs, as determined by RT-qPCR. The data for *Parp1*^{-/-} mESCs are expressed relative to WT ESCs. Each bar represents the mean plus the SEM, $n \geq 3$. The differences observed for *Nanog*, *Parp1*, and all of the differentiation genes are significant (Student’s t test, p -value < 0.05).

B) Genome browser tracks of mRNA-seq data (*top*) and H3K4me3 ChIP-seq data (*bottom*) around the *Nanog*, *Oct4*, and *Gata6* genes in WT and *Parp1*^{-/-} mESCs.

C) Volcano plot of RNA-seq data from WT and *Parp1*^{-/-} mESCs showing genes whose expression significantly increases upon *Parp1* knockout (“Negatively Regulated” by PARP-1) or significantly decreases upon *Parp1* knockout (“Positively Regulated” by PARP-1). The patterns of significant regulation are color-coded as follows: (a) *Red*, negatively regulated genes with ≥ 2 -fold increase upon *Parp1* knockout; (b) *Orange*, negatively regulated genes with < 2 -fold increase upon *Parp1* knockout; (c) *Blue*,

positively regulated genes with ≥ 2 -fold decrease upon *Parp1* knockout; (d) *Green*, positively regulated genes with < 2 -fold decrease upon *Parp1* knockout.

D) Numbers of genes positively or negatively regulated by PARP-1, from panel C, with the same color coding and cutoffs.

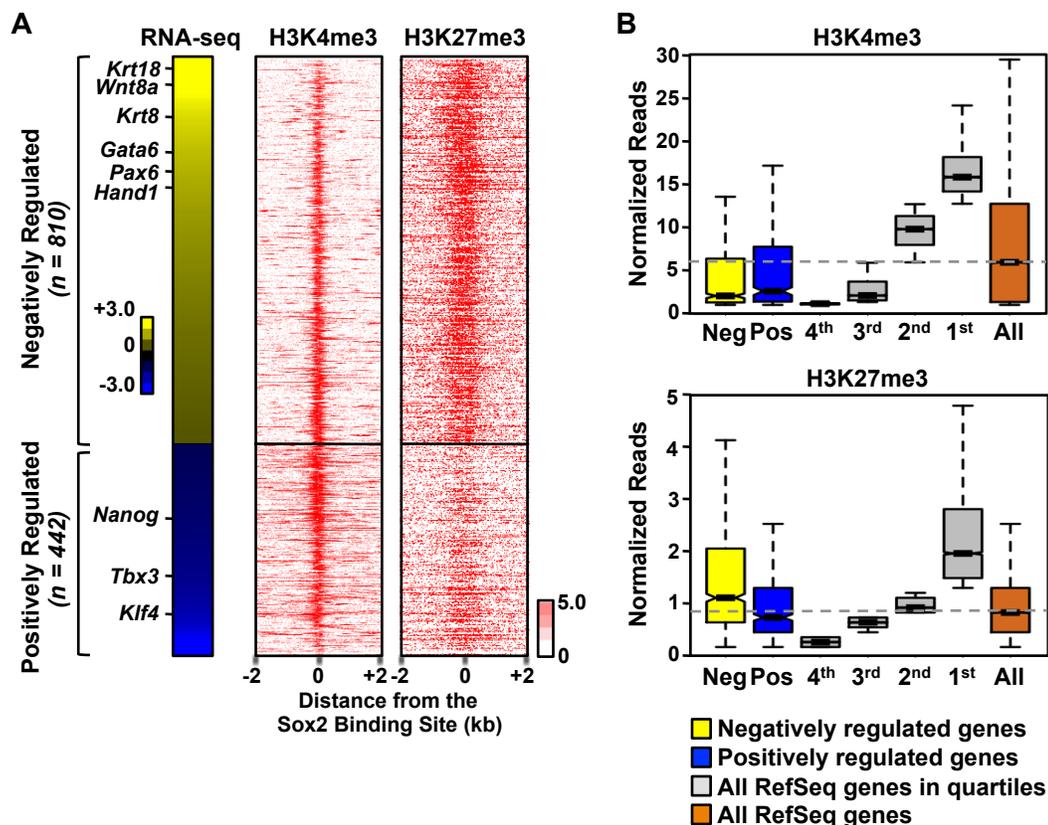


Figure 2.2. The promoters of genes negatively regulated by PARP-1 are enriched for both H3K4me3 and H3K27me3.

A) Status of H3K4me3 and H3K27me3 at the promoters of PARP-1-regulated genes in mESCs. *Left*, Heatmap of the relative expression levels of genes whose expression significantly ($FDR < 5\%$) increases upon *Parp1* knockout (“Negatively Regulated” by PARP-1) or decreases upon *Parp1* knockout (“Positively Regulated” by PARP-1) from RNA-seq, as in Fig. 1A. *Right*, Heatmaps of the relative levels of H3K4me3 and H3K27me3 from ChIP-seq, as indicated. The ChIP-seq data are centered on the transcription start sites of the genes (TSSs; ± 2 kb). The genes are listed in the same order top to bottom for the RNA-seq and ChIP-seq data.

B) Box plots of normalized H3K4me3 (*top*) and H3K27me3 (*bottom*) ChIP-seq read counts in a 1 kb window at the TSSs of genes negatively regulated by PARP-1 ($n = 810$) or positively regulated by PARP-1 ($n = 442$), as well as all RefSeq genes binned in quartiles based on expression levels from RNA-seq data.

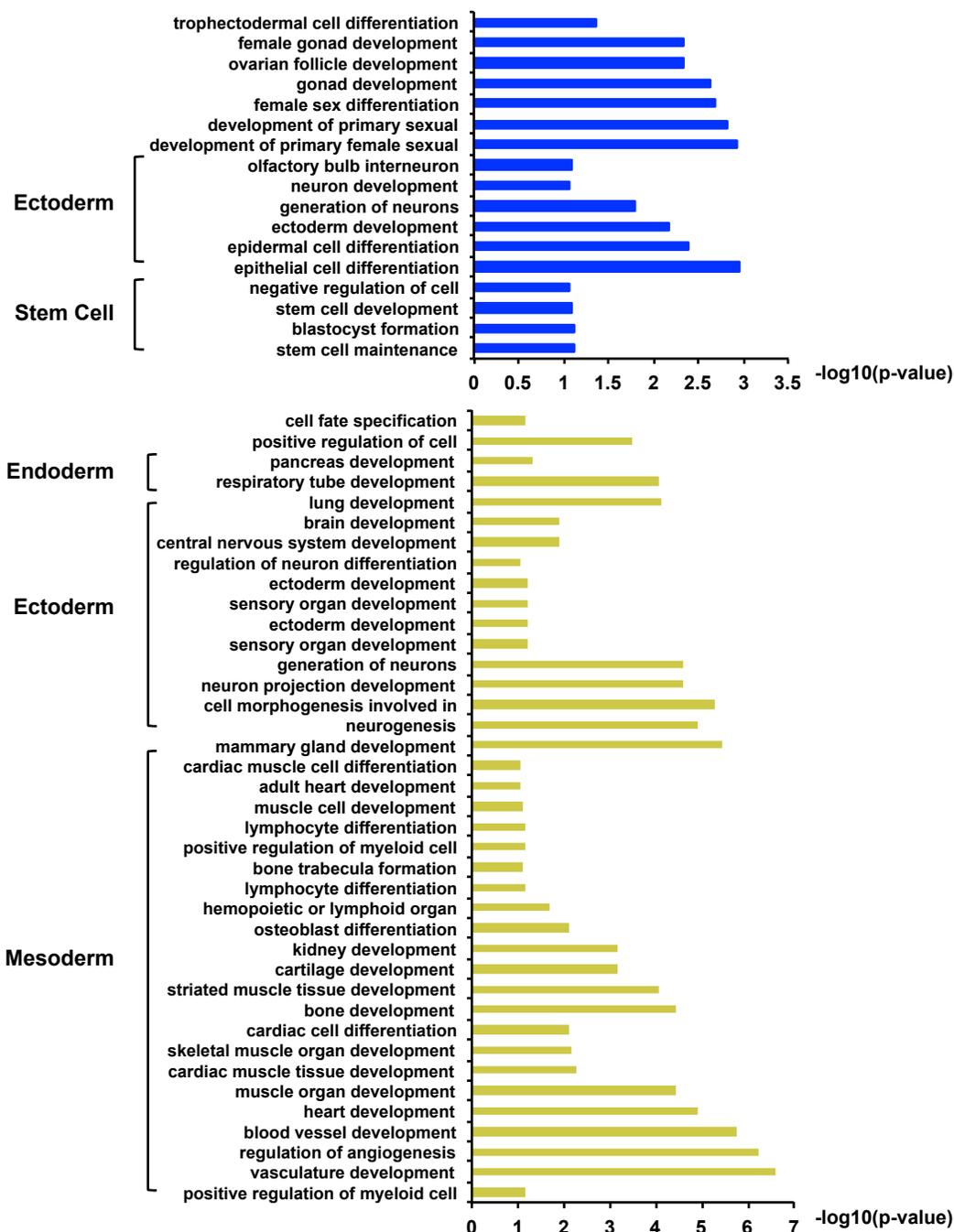


Figure 2.3. Gene expression changes in mESC upon Parp-1 depletion.

Gene ontology of genes positively or negatively regulated by Parp-1 using the DAVID bioinformatics tool.

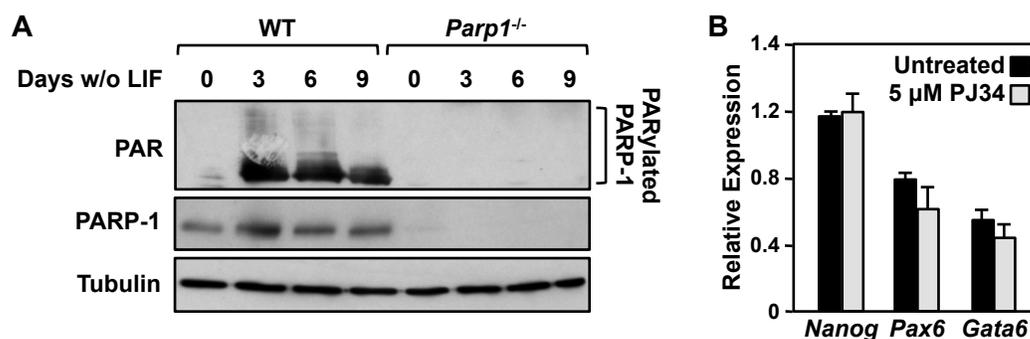


Figure 2.4. The catalytic activity of PARP-1 is not required for the regulation of gene expression in undifferentiated mESCs.

A) The level of PARP-1-mediated PARylation is very low in undifferentiated mESCs and increases upon differentiation. Western blots of poly(ADP-ribose) (PAR) and PARP-1 showing their relative levels in mESCs during a 9 day time course of differentiation upon LIF removal.

B) The expression of *Nanog* (a pluripotency-associated gene), as well as *Pax6* and *Gata6* (differentiation-associated genes), in undifferentiated ('Day 0') mESCs is not affected by treatment with the PARP inhibitor PJ34. RT-qPCR was performed using total RNA isolated from mESCs treated with 5 μ M PJ34 for 24 hrs. The expression levels were standardized to the expression of the *Gapdh* gene. Each bar represents the mean plus the SEM, $n \geq 3$. The small differences are not significant.

2.3.2. PARP-1 globally co-localizes and interacts with Sox2.

To determine how PARP-1 might regulate gene expression in mESCs, we performed a series of genomic analyses, including chromatin immunoprecipitation-sequencing (ChIP-seq) for PARP-1. Previous genomic study of PARP-1 using ChIP-chip in human breast cancer cell line MCF7 cells showed that PARP-1 preferentially binds to actively transcribed gene promoters^{6,7}. Consistent with this, when we aligned PARP-1 ChIP-seq signals to gene transcription start sites (TSS) and transcriptional termination sites (TTS) (Fig. 2.5C), we observed an enrichment of PARP-1 signal on gene TSS regions. Moreover, when we ranked genes based on their transcription level as shown by RNA-seq, we observed a preferential binding of PARP-1 on the TSSs of highly transcribed genes, while repressed genes were associated with very low levels of PARP-1 (Fig. 2.5D). Using MACS¹⁶, we identified 23,661 high-confidence PARP-1 binding sites ($P < 0.001$). Surprisingly, among these significantly PARP-1 enriched sites, less than 20%

were within 5kb of gene TSSs. More than 80% of PARP-1 peaks were distal binding sites more than 5kb away from their nearest gene TSSs (Fig. 2.5A-B).

We tested which transcription factor has the most similar genomic localization pattern to PARP-1 by performing hierarchical clustering of chromatin binding patterns? Comparing PARP-1 and a panel of chromatin-/transcription-related factors in embryonic stem cells. Strikingly, PARP-1 binding highly correlated with Sox2 and Oct4 (Fig. 2.6D). Moreover, when we aligned the Sox2 binding sites with PARP-1 binding sites using our ChIP-seq data sets, a significant correlation between Sox2 and PARP-1 occupancy was observed. Co-occupancy between PARP-1, Sox2 and Oct4 was observed in multiple genomic locations (Fig. 2.6E-G). Furthermore, we observed an interaction between PARP-1 and Sox2 by co-immunoprecipitation, but not between PARP-1 and Oct4, indicating PARP-1 directly interacts with Sox2, but not Oct4 (Fig. 2.6F). Together, these results indicate a physical association and colocalization between PARP-1 and Sox2 across the genome in mESCs.

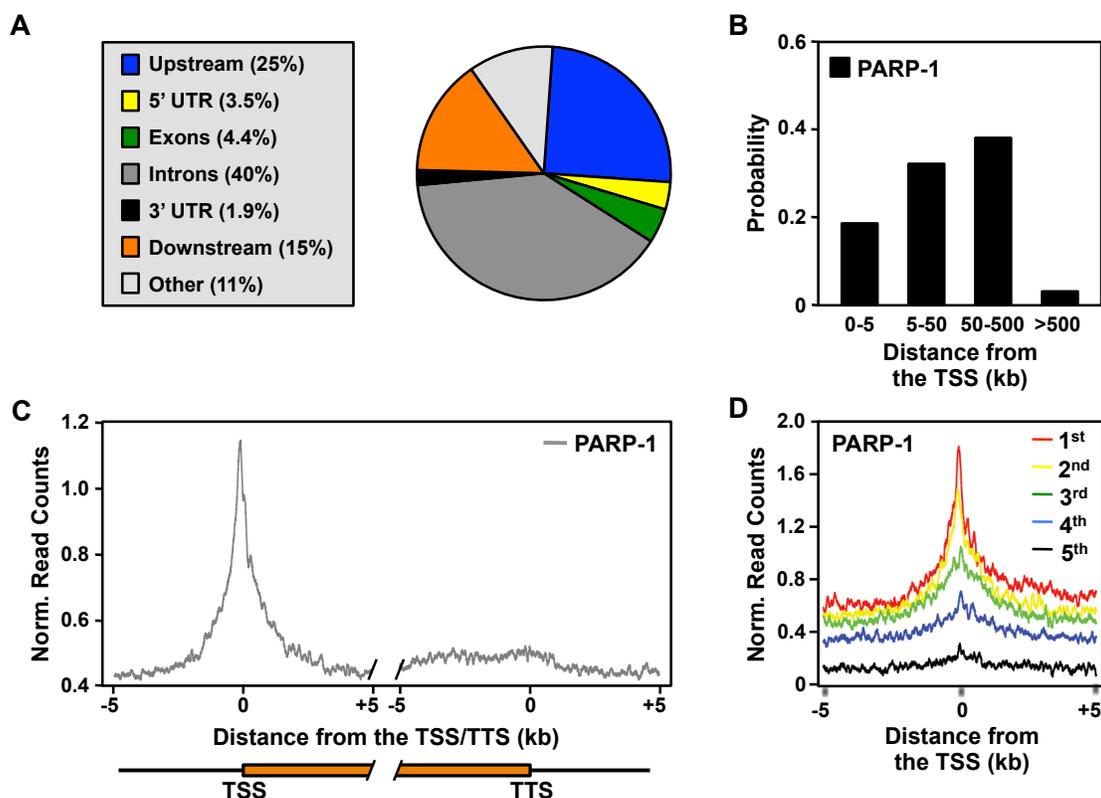


Figure 2.5. PARP-1 is enriched at the promoters of highly expressed genes.

A) Distribution of significant peaks of PARP-1 binding across genomic features in mESCs from PARP-1 ChIP-seq data.

B) Distribution of significant peaks of PARP-1 binding relative to the TSSs of all RefSeq genes determined using the web-based bioinformatics tool GREAT¹⁷.

C) PARP-1 is enriched at gene promoters in mESCs. Metagene plot showing the average normalized read counts of PARP-1 ChIP-seq reads around the transcription start site (TSS) and transcription termination site (TTS) for all RefSeq genes. The data were compiled for regions ± 5 kb relative to the TSS and the TTS.

D) Metagene plot showing the average normalized read counts of PARP-1 ChIP-seq reads around the TSS for genes ranked based on expression level as determined by RNA-seq. Five pentiles: 1st = highest 20% of expression; 5th = lowest 20% of expression.

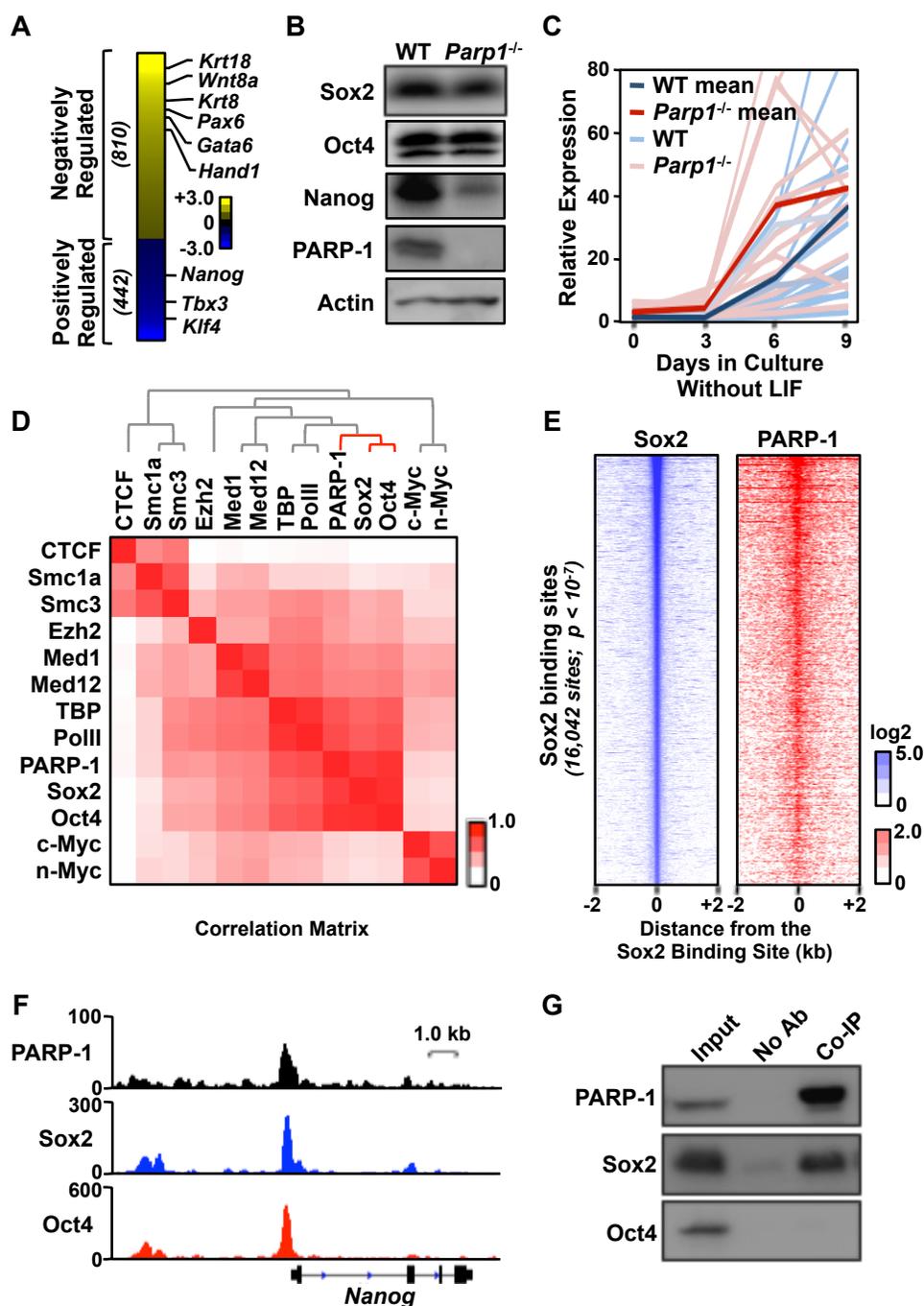


Figure 2.6. PARP-1 colocalizes with Sox2 genome-wide and helps to maintain a pluripotency gene expression program in mouse embryonic stem cells.

A) Effect of *Parp1* knockout on gene expression in mESCs as determined by RNA-seq. The heatmap shows the relative expression levels of genes whose expression significantly ($FDR < 5\%$) increases upon *Parp1* knockout (“Negatively Regulated” by PARP-1) or

decreases upon *Parp1* knockout (“Positively Regulated” by PARP-1). The data are $\log_2(\text{Parp1}^{-/-} \text{RPKM}/\text{WT RPKM})$.

B) Western blots showing the relative levels of three pluripotency factors (Sox2, Oct4, and Nanog) and PARP-1 in WT and *Parp1*^{-/-} mESCs. Actin is used as an internal loading control.

C) Analysis of mRNA expression for a panel of 19 differentiation-associated genes in mESCs during a 9 day time course of differentiation upon LIF removal. The expression levels of each mRNA are normalized to *Gapdh* mRNA levels. Light lines represent the relative expression levels of individual genes. Dark lines represent the mean relative expression levels of all 19 genes combined.

D) PARP-1 colocalizes with Sox2 genome-wide. Correlation matrix of genome-wide enrichment for chromatin- and transcription-related factors from ChIP-seq data in mESCs. The data are organized and ordered using hierarchical clustering.

E) *Left*, Heatmap of high confidence Sox2 peaks in mESCs from ChIP-seq data ($n = 16,042$; $p\text{-value} < 10^{-7}$) centered on the Sox2 binding sites (± 2 kb) and ordered top to bottom by signal intensity. *Right*, PARP-1 ChIP-seq signals associated with the corresponding Sox2 binding sites.

F) Genome browser tracks of PARP-1, Sox2 and Oct4 ChIP-seq data around the *Nanog* gene in WT mESCs.

G) PARP-1 binds to Sox2, but not Oct4. Flag-tagged PARP-1 was ectopically expressed in HEK293T cells with Sox2 and Oct4, and immunoprecipitated using a Flag antibody. Western blots showing the relative levels of PARP-1, Sox2, and Oct4 in the input and the co-immunoprecipitated (co-IP) material.

2.3.3. PARP-1 is required to stabilize the chromatin association of Sox2 to a subset of genomic locations.

To explore the functional link between Sox2 and PARP-1 in more detail, we performed Sox2 and Oct4 ChIP-seq in *Parp1* knockout and wild type ES cells. Knockout of *Parp1* caused a significant reduction in Sox2 binding across the genome. Out of 16,042 Sox2 peaks that we detected, we identified 1,606 sites that were lost (‘PARP-1-dependent’ Sox2 binding sites) and 4,556 sites that were unchanged (‘PARP-1-independent’ Sox2 binding sites) upon *Parp1* knockout (Fig. 2.10B). Notably, loss of Sox2 occupancy correlated with decreased gene expression. Specifically, PARP-1 depletion caused a decrease of Sox2 binding on the *Nanog* gene promoter, while not affecting Sox2 occupancy on the *Pou5f1* gene (Fig. 2.10A). In addition to the *Nanog* gene, *Klf4* as well as *Tbx3* genes positively regulated by PARP-1 were also found to be associated with Sox2 binding sites that are dependent on PARP-1 (Fig. 2.7A). To investigate whether the changed gene expression upon PARP-1 loss is explained by the decrease of Sox2 occupancy genome-wide, we associated each Sox2 binding sites with its highly likely

regulated genes defined based on having Sox2 peak-TSS distance smaller than 10 kb¹⁸. Gene set enrichment analysis (GSEA) revealed that the expression of genes associated with PARP-1-dependent Sox2 binding sites are significantly decreased in *Parp1*^{-/-} mESCs compared to WT mESCs based on RNA-seq (Fig. 2.7B). Furthermore, the genes positively regulated by PARP-1 were enriched for PARP-1-dependent Sox2 binding sites near their TSSs (Fig. 2.7C). Importantly, although some PARP-1-dependent Sox2 binding sites were co-occupied by Oct4 (e.g., Fig 5G, Fig. 8 and Fig. 10 B), the effects of PARP-1 on Sox2 binding were not mediated through Oct4. This is illustrated by the observation that Sox2 binding sites without Oct4, but not Oct4 binding sites without Sox2, exhibited PARP-1-dependent binding (Fig. 2.8). Together, these results suggest that decreased expression of pluripotency genes upon *Parp1* knockout in mESCs (e.g., Fig. 2.1A) are more likely to be a direct effect of the PARP-1-Sox2 pathway than increased expression of differentiation genes.

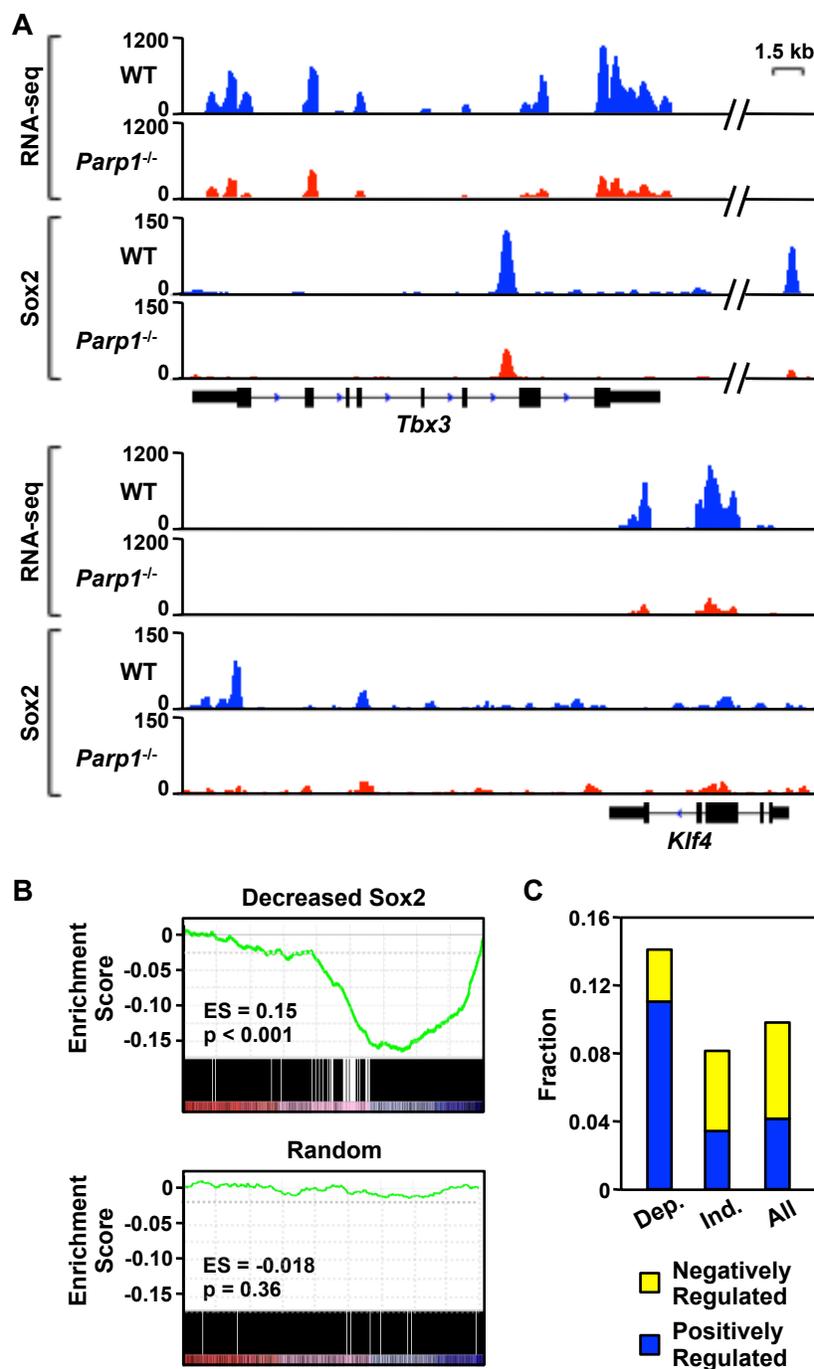


Figure 2.7. Decreased gene expression in mESCs upon *Parp1* knockout correlate with decreased Sox2 binding.

A) Knockout of PARP-1 in mESCs is associated with reduced Sox2 binding at some genomic loci. Genome browser tracks of RNA-seq data (*upper tracks in each group*) and Sox2 ChIP-seq data (*lower tracks in each group*) in WT and *Parp1*^{-/-} mESCs at the *Tbx3* (*top*) and *Klf4* (*bottom*) loci.

B) GSEA analysis showing the relationship between PARP-1-regulated Sox2 binding sites ($n = 1,606$) and gene expression changes upon *Parp1* knockout in mESCs. *Top*, The expression of genes associated with PARP-1-regulated Sox2 binding sites is significantly decreased in *Parp1*^{-/-} mESCs compared to WT mESCs (p -value < 0.001) based on RNA-seq. *Bottom*, A randomly selected and equally sized set of genes ($n = 1,606$) is shown as a control (p -value = 0.36).

C) Fraction of Sox2 binding sites associated with PARP-1-regulated genes. Dep. = PARP-1-dependent Sox2 binding sites; Ind. = PARP-1 independent Sox2 binding sites; All = All Sox2 binding sites. *Yellow*, Genes whose expression significantly increases upon *Parp1* knockout (“Negatively Regulated” by PARP-1). *Blue*, Genes whose expression significantly decreases upon *Parp1* knockout (“Positively Regulated” by PARP-1).

The results shown here suggest that decreased expression of pluripotency and related genes upon *Parp1* knockout in mESCs (e.g., Fig. 2.1A) is more likely to be a direct effect of the PARP-1-Sox2 pathway than increased expression of differentiation genes (e.g., Fig. 2.1, A and C).

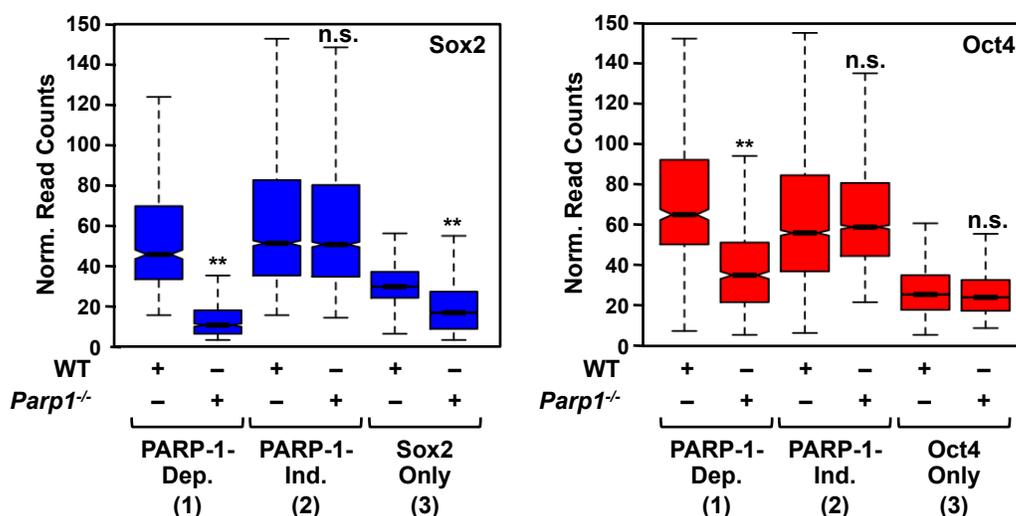


Figure 2.8. Sox2 binding sites without Oct4, but not Oct4 binding sites without Sox2, exhibit PARP-1-dependent binding.

Normalized ChIP-seq read counts for Sox2 (*left*) and Oct4 (*right*) at (1) PARP-1-dependent (Dep.) sites, (2) PARP-1-independent (Ind.) sites, and (3) sites with Sox2 or Oct4 only. Asterisks indicate significant differences: Student’s t-test, p -value $< 2.2 \times 10^{-16}$.

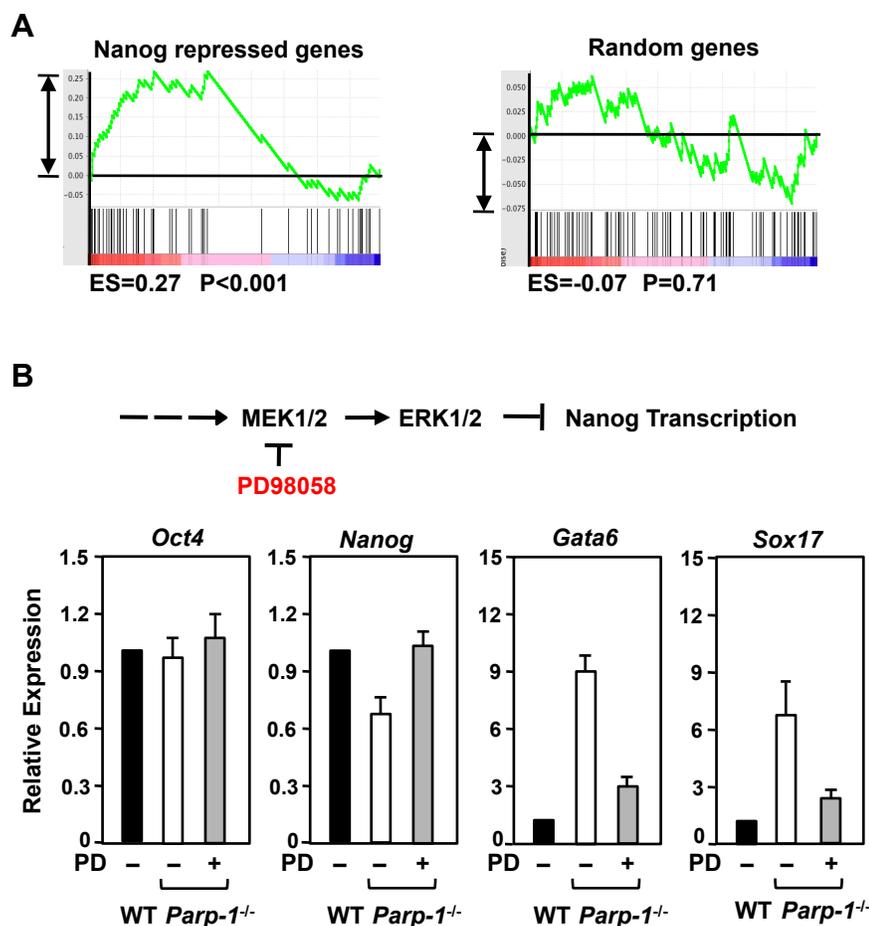


Figure 2.9. Increase of gene expression upon Parp-1 depletion is partially explained by decreased level of Nanog.

A) GSEA analysis showing the relationship between Nanog repressed genes and gene expression changes upon Parp-1 knockout. Nanog repressed genes are significantly decreased in Parp-1^{-/-} RNA-seq sample compared to WT (p-value < 0.001). Randomly selected equal number of genes are shown as control (p-value = 0.71).

B) Rescue of Nanog expression level using MEK inhibitor PD98058 partially rescues expression changes of differentiation genes in Parp-1^{-/-} mESCs.

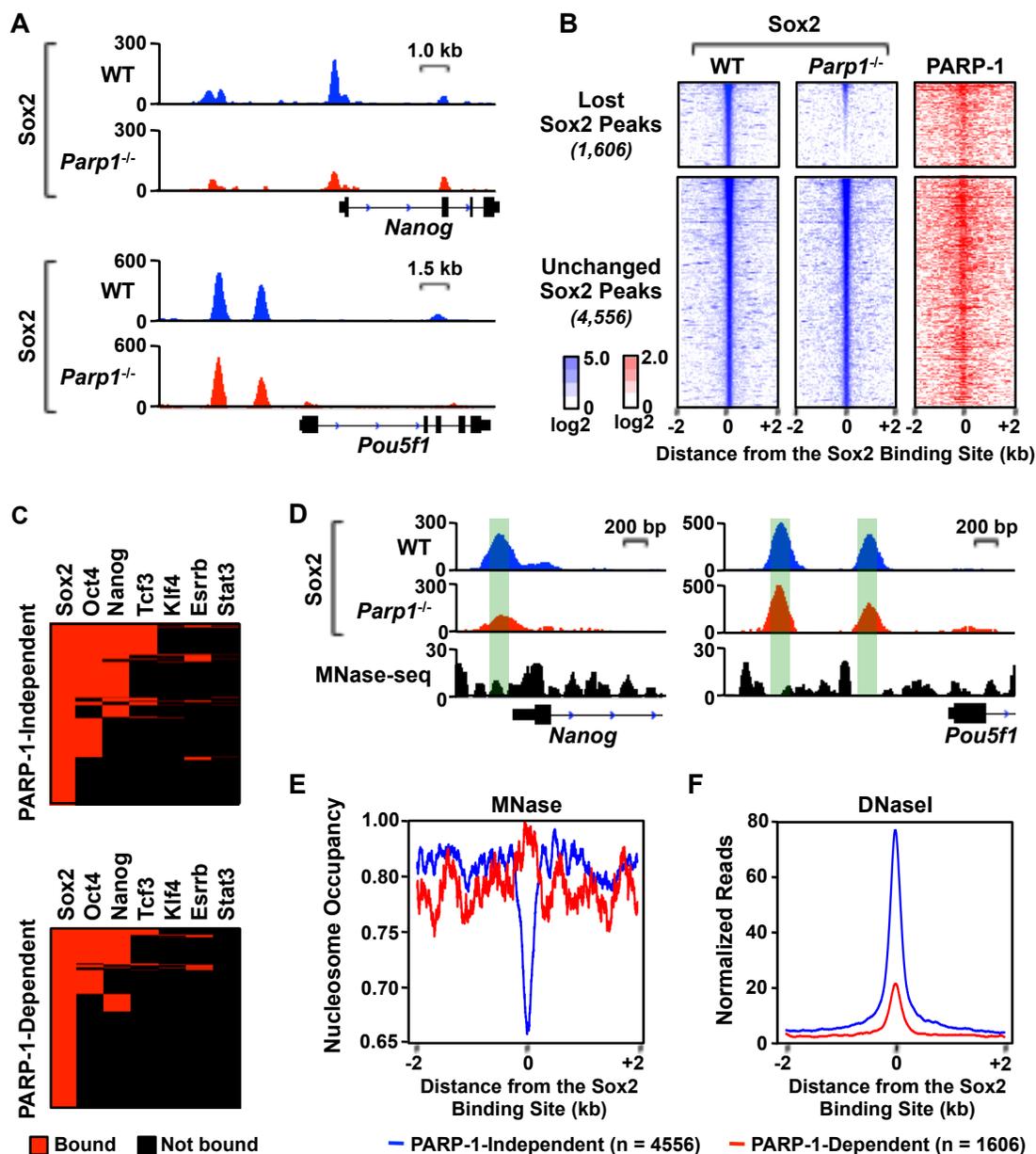


Figure 2.10. PARP-1 is required for the binding of Sox2 to a subset of its genomic sites in mESCs.

A) Genome browser tracks of Sox2 ChIP-seq data around the *Nanog* and *Pou5f1* (encoding Oct4) genes in WT and *Parp1^{-/-}* mESCs.

B) *Left and middle*, Heatmaps of Sox2 ChIP-seq signals in WT and *Parp1^{-/-}* mESCs centered on the Sox2 binding sites (± 2 kb) and ordered top to bottom by signal intensity. *Left*, PARP-1 ChIP-seq signals in WT mESCs associated with the corresponding Sox2 binding sites. PARP-1-dependent Sox2 binding sites are defined as those sites whose ChIP-seq signals are significantly decreased upon PARP-1 knockout (p-value < 0.01; n =

1,606). PARP-1-independent Sox2 binding sites show no change upon PARP-1 knockout (n = 4,556).

C) PARP-1-independent Sox2 binding sites have a higher level of co-occupancy by other transcription factors in mESCs. Heatmaps showing the binding of other transcription factors at Sox2 binding sites. *Top*, Results for PARP-1-independent Sox2 binding sites. *Bottom*, Results for PARP-1-dependent Sox2 binding sites. *Red*, significant binding; *Black*: no significant binding.

D) Genome browser tracks of Sox2 ChIP-seq data and MNase-seq data around the *Nanog* gene in WT and *Parp1*^{-/-} mESCs. The green shading highlights the relationship between Sox2 binding and nucleosome occupancy.

E) Average MNase-seq signals surrounding PARP-1-dependent (*red*) and PARP-1-independent (*blue*) Sox2 binding sites in mES cells. The data are centered on the Sox2 binding sites determined by ChIP-seq (± 2 kb).

F) Average DNase-seq signals surrounding PARP-1-dependent (*red*) and PARP-1-independent (*blue*) Sox2 binding sites in mES cells. The data are centered on the Sox2 binding sites determined by ChIP-seq (± 2 kb).

2.3.4. PARP-1-dependent Sox2 binding sites are associated with closed chromatin conformation.

Notably, both PARP-1-dependent and -independent Sox2 binding sites were associated with PARP-1. We asked what determines the PARP-1 dependency for Sox2. Transcription factors frequently co-occupy with other factors across the genome, and cooperatively stabilize each other's binding to chromatin^{19 20}. Interestingly, analysis of two distinct classes of Sox2 binding sites with differential PARP-1 dependency showed that although both classes of Sox2 sites were associated with *sox* binding motifs, a significantly lower fraction of PARP-1-dependent Sox2 sites contained optimal *sox/pou* motifs allowing strong interaction between Sox2 and Oct4²¹ (Fig. 2.12A). This indicates that PARP-1-dependent Sox2 binding sites may have lower cooperativity with Oct4. Indeed, Oct4 co-localized to a smaller fraction of PARP-1-dependent Sox2 binding sites (Fig. 2.12 B-C). In addition to Oct4, 5 other transcription factors we checked including Nanog, Tcf3, Klf4, Esrrb and Stat3 also showed significantly lower co-occupancy with PARP-1-dependent Sox2 sites compared to PARP-1-independent binding sites (Fig. 2.10C, Fig. 2.12B-C, Fig 2.13). Nucleosomes function as a barrier for transcription factor binding to chromatin²². Previous studies showed that Sox2 is able to bind to nucleosome-enriched regions during somatic cell reprogramming¹³. Consistent with this observation, we found that Sox2 binds to both nucleosome-depleted and -occupied chromatin regions (Fig. 2.10

D-E, Fig. 2.11). However, to our surprise, when we compared nucleosome occupancy between PARP-1-dependent and -independent Sox2 binding sites, PARP-1-dependent Sox2 sites showed significantly higher nucleosome occupancy (Fig. 2.10 D-E). Consistent with this, DNase-seq, which assesses chromatin accessibility, shows that PARP-1-dependent Sox2 binding sites are associated with closed chromatin structure (Fig. 2.10F). Collectively, our genomic analyses revealed a class of Sox2 binding sites that require PARP-1 for binding and have a specific set of features that distinguish them from PARP-1-independent Sox2 binding sites (Fig. 2.14).

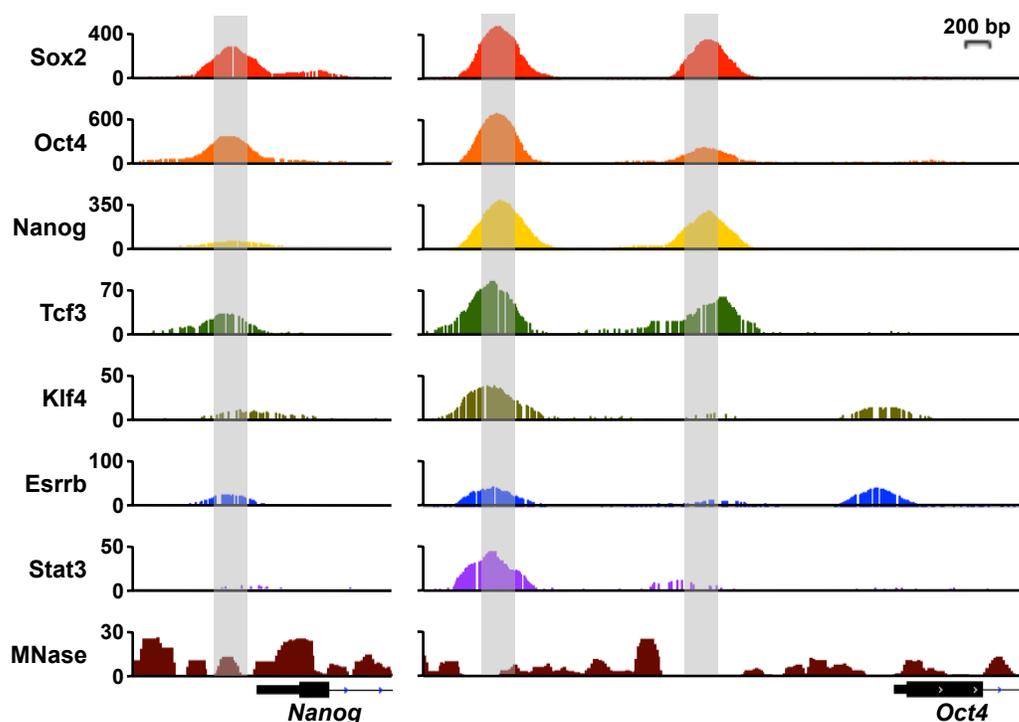


Figure 2.11. PARP-1-dependent Sox2 binding sites exhibit higher nucleosome occupancy and lower co-occupancy by other transcription factors than PARP-1-independent Sox2 binding sites.

Genome browser tracks of ChIP-seq data for various transcription factors and MNase-seq data from mESCs at the *Nanog* and *Oct4* genomic loci.

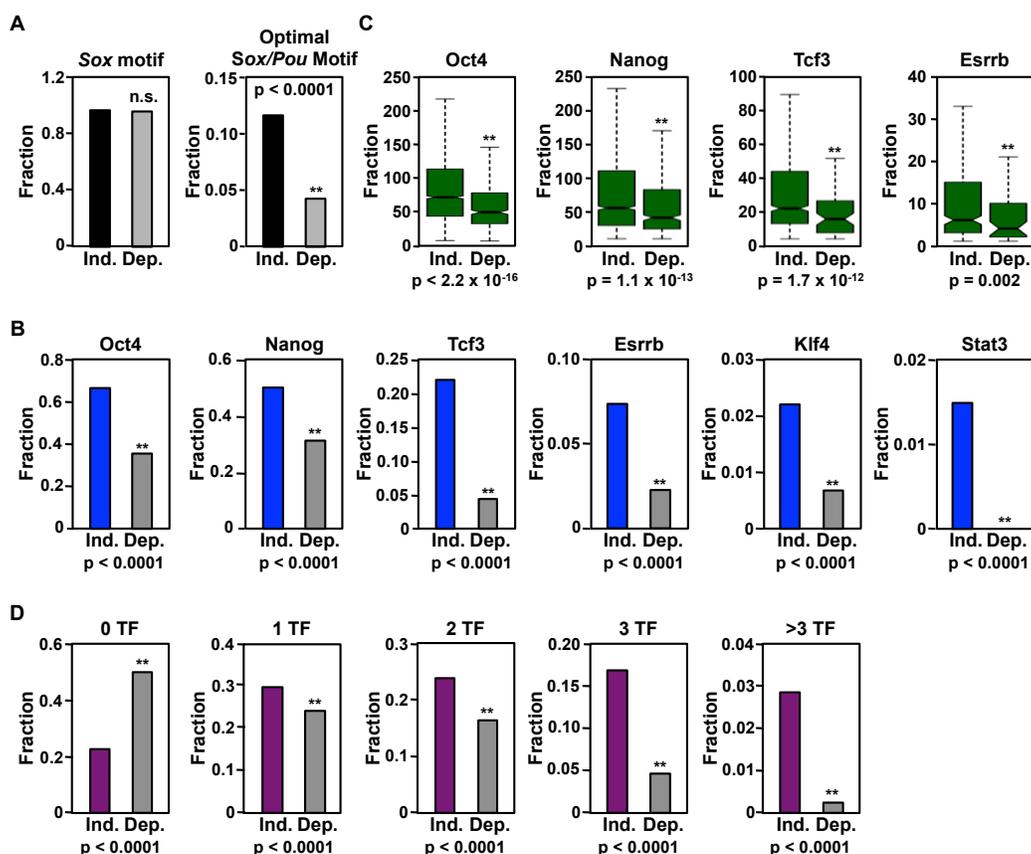


Figure 2.12. PARP-1-dependent Sox2 binding sites exhibit lower co-occupancy by other transcription factors than PARP-1-independent Sox2 binding sites.

A) Fraction of PARP-1-independent (Ind.) and PARP-1-dependent (Dep.) Sox2 binding sites associated with a *Sox* motif (*left*) or an optimal *Sox/Pou* motif (*right*) based on ChIP-seq in mESCs. Asterisks indicate significant differences: Fisher's exact test, p -value < 0.0001 .

B) Fraction of PARP-1-independent (Ind.) and PARP-1-dependent (Dep.) Sox2 binding sites from ChIP-seq data associated with the transcription factors indicated, based on ChIP-seq from mESCs. Asterisks indicate significant differences: Fisher's exact test, p -value < 0.0001 .

C) Normalized ChIP-seq read counts for the transcription factors indicated at PARP-1-independent (Ind.) and PARP-1-dependent (Dep.) Sox2 binding sites based on ChIP-seq in mESCs. Asterisks indicate significant differences: Student's t-test at the p -values indicated.

D) Fraction of PARP-1-independent (Ind.) and PARP-1-dependent (Dep.) Sox2 binding sites associated with the specified number transcription factors (TFs; 0, 1, 2, 3, >3) based on ChIP-seq in mESCs. Asterisks indicate significant differences: Fisher's exact test, p -value < 0.0001 .

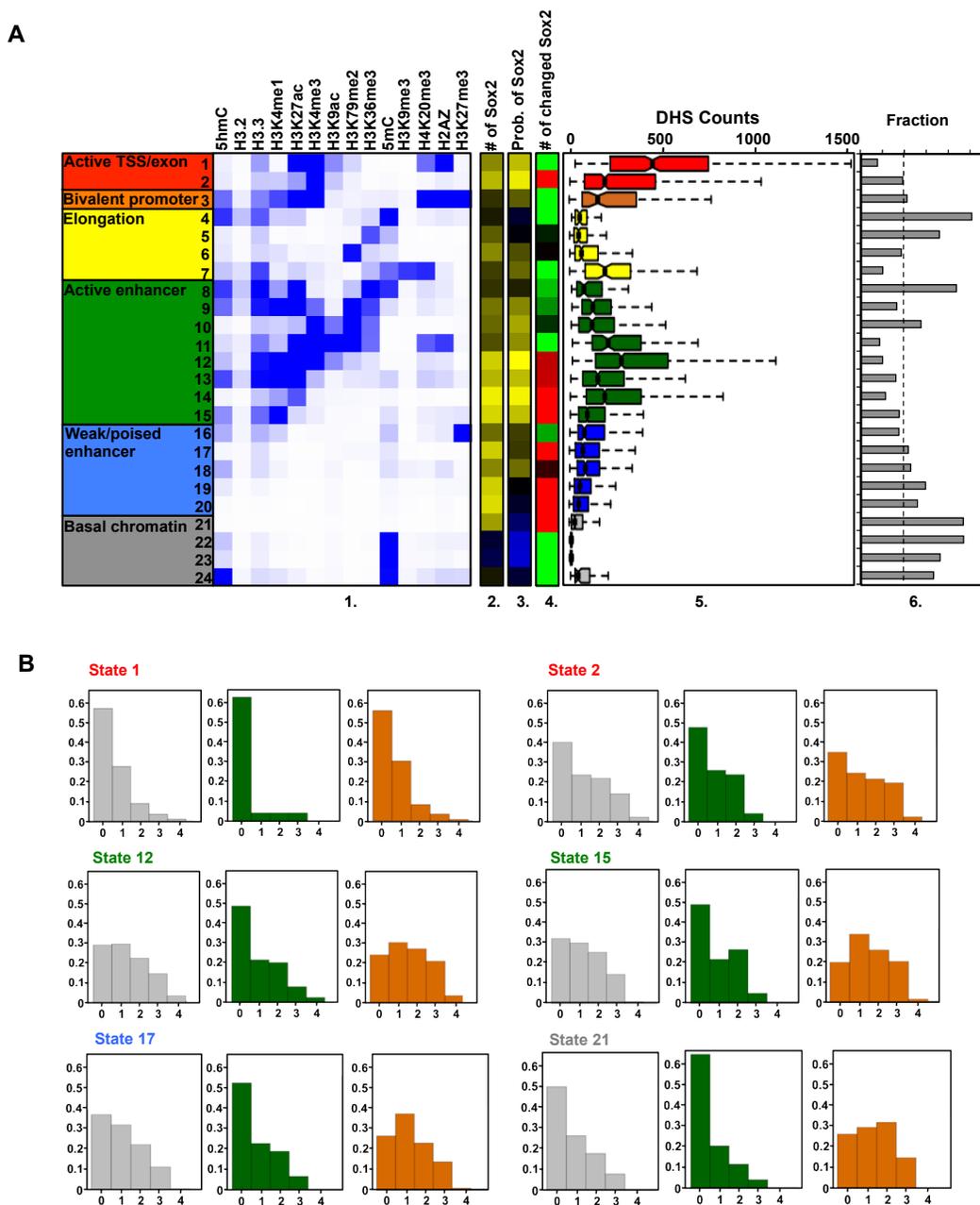


Figure 2.13. Parp-1 dependent Sox2 binding sites have less cooperation with other transcription factors.

A) (1) Fractionation of genome into 24 states using ChromHMM using available histone marks and DNA methylation data. (2) Number of Sox2 binding sites associated with each chromatin state. (3) Number of Sox2 binding sites associated with each chromatin state normalized to fraction of the respective chromatin state. (4) Number of Parp-1 dependent Sox2 binding sites in each chromatin state. (5) Normalized DNaseI-seq read counts around Sox2 binding sites associated with each chromatin state. (6) Fraction of Parp-1 dependent Sox2 binding sites associated with each chromatin state.

B) Fraction of Sox2 binding sites associated with various numbers of transcription factors in different chromatin states. (Grey) Total Sox2 binding sites; (Green) Parp-1 dependent Sox2 binding sites; (Orange) Parp-1 independent Sox2 binding sites.

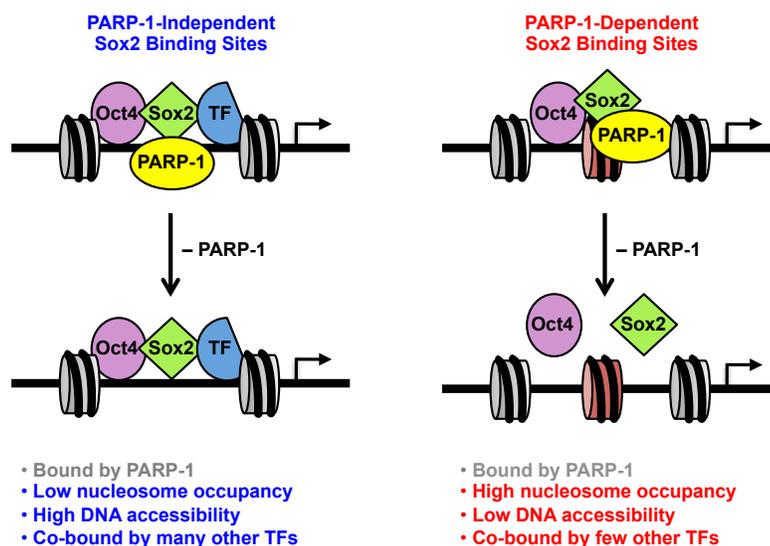


Figure 2.14. Summary of genomic features for PARP-1-independent and PARP-1-dependent Sox2 binding sites

Schematic representation of the genomic features for PARP-1-independent and PARP-1-dependent Sox2 binding sites based on Figs. 1 through 12.

2.3.5. PARP-1 stabilizes Sox2 binding to nucleosomes in vitro.

The observation that PARP-1 is specifically required for stabilizing Sox2 binding to closed chromatin regions with higher nucleosome occupancy led us to ask whether PARP-1 is able to stabilize Sox2 binding to nucleosomes. To answer this question, we biotin labeled the end of 601 nucleosome positioning DNA and assembled it into nucleosomes. In order to allow sequence-specific Sox2 binding, a 15 base pair sequence in the center of the 601 nucleosome positioning element (NPE) was replaced with a *Sox*-binding motif. Altering the sequence of 601 NPE preserved its ability to position the nucleosome (Fig. 2.15 B-C). We then immobilized the biotin-labeled nucleosome/DNA to streptavidin-coated magnetic beads and incubated it with recombinant Sox2 in the presence or absence of PARP-1 protein. Sox2 strongly bound to naked DNA, and adding

PARP-1 did not affect Sox2 binding (Fig. 2.16 A, lane 2 versus lane 4). On the other hand, Sox2 bound weakly to nucleosomes, with an affinity more than 20 fold less than to naked DNA. Strikingly, adding PARP-1 significantly enhanced Sox2 binding to nucleosomes (Fig. 2.16A, lane 10 versus lane 12). Interestingly, Sox2 also bound non-specifically to 601 nucleosomes without a Sox motif, although this non-specific interaction was inhibited by adding PARP-1 (Fig. 2.16A, lane 6 versus lane 8), indicating that PARP-1 may block non-specific nucleosome interactions of Sox2 by competing with it.

To gain insight into whether PARP-1 enhances direct interaction between Sox2 and nucleosomal DNA, we performed a DNaseI footprinting assay²³. Consistent with in vitro pull-down assay results, PARP-1 did not change Sox2 binding to naked DNA, while significantly promoting Sox2 interaction with DNA when it was assembled into nucleosomes (Fig. 2.16 B and C). Notably, point mutants of PARP-1 protein, which abolishes PARP-1 interaction with DNA⁵, failed to stabilize Sox2 binding to nucleosomes (Fig. 2.18 B and C), suggesting DNA binding of PARP-1 is required for enhancing Sox2 association with nucleosomes. Taken together, these results suggested that PARP-1 is able to stabilize Sox2 interaction with DNA in a nucleosomal context in vitro.

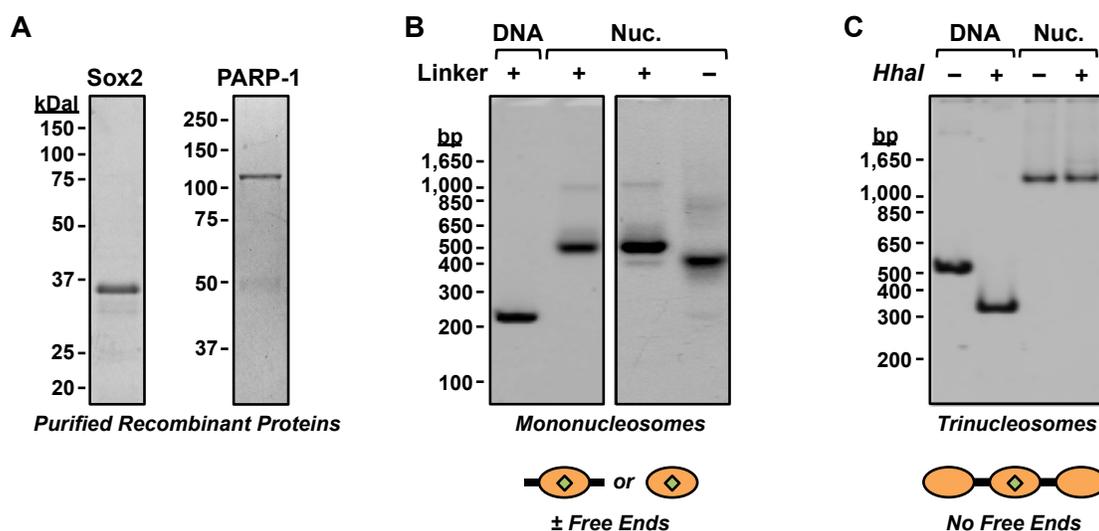


Figure 2.15. Recombinant proteins and nucleosomes used in biochemical assays exploring functional interactions between PARP-1 and Sox2.

A) Coomassie-stained SDS-PAGE gels showing purified recombinant Sox2 (*left*) and PARP-1 (*right*). Molecular weight markers in kDal are shown for comparison.

B) Native polyacrylamide gel showing biotinylated 601 nucleosome positioning element (NPE) DNA (“DNA”) or the same DNA assembled into mononucleosomes (“Nuc.”). The histone-DNA complex of the mononucleosome migrates more slowly than the free DNA. Mononucleosomes were assembled with (shown) or without (not shown) linker DNA and free DNA ends extending beyond the nucleosome core for use in nucleosome binding assays and DNaseI footprinting assays. The green diamond indicates the location of Sox motif.

C) Native polyacrylamide gel showing free 3x 601 nucleosome positioning element (NPE) DNA (“DNA”) or the same DNA assembled into a trinucleosome (“Nuc.”) with or without digestion by *HhaI*, which cleaves within the nucleosome. The histone-DNA complex of the mononucleosome migrates more slowly than the free DNA and is resistant to cleavage by *HhaI*. The trinucleosomes were used in DNaseI footprinting assays. The green diamond indicates the location of the Sox motif, which is in the middle nucleosome.

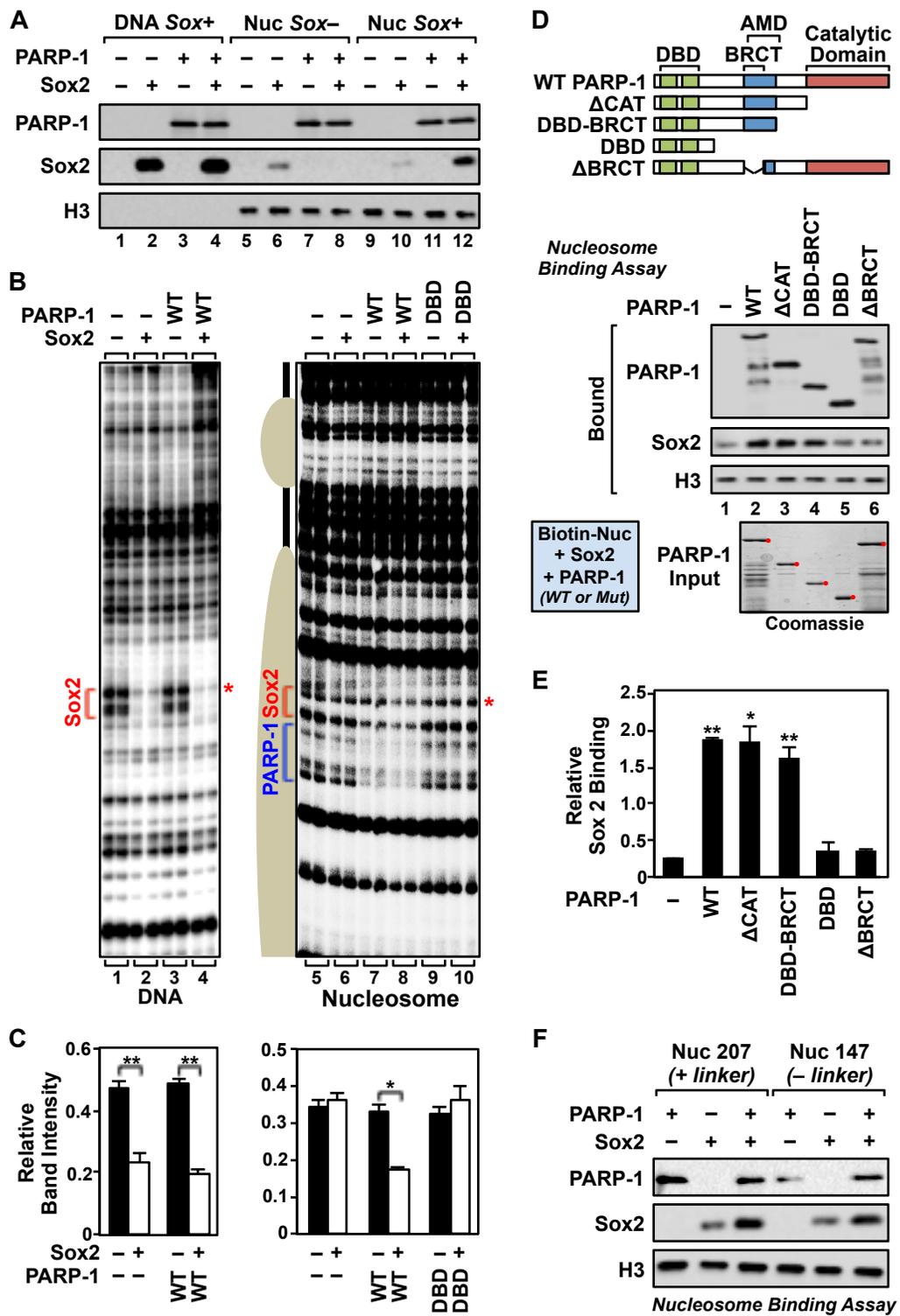


Figure 2.16. PARP-1 stabilizes Sox2 binding to nucleosomes.

A) In vitro nucleosome binding assays show PARP-1-dependent binding of Sox2 to nucleosomes, but not naked DNA. Biotin-labeled 601 NPE DNA with or without a *Sox* motif (*Sox*⁺ or *Sox*⁻, respectively), or the same DNA assembled into a mononucleosome (Nuc), was immobilized on streptavidin beads and incubated ± recombinant Sox2 and PARP-1, as indicated. After washing, the bound proteins were analyzed by Western blotting. Sox2 binds in a PARP-1-independent manner to naked DNA containing a *Sox* motif (lane 2 versus 4). PARP-1 reduces non-specific Sox2 binding to nucleosomes lacking a *Sox* motif (lane 6 versus 8), but enhances Sox2 binding to nucleosomes containing a *Sox* motif (lane 10 versus 12).

B) In vitro DNase I footprinting assays show PARP-1-dependent binding of Sox2 to nucleosomes, but not naked DNA. *Left*, Footprinting assay with naked 3x 601 NPE DNA containing a *Sox* motif. Addition of Sox2 protects the DNA template from digestion with or without the addition of wild-type (WT) PARP-1 (lane 1 versus 2 and 3 versus 4; see the band marked with a red asterisk). *Right*, Footprinting assay with a trinucleosome containing a *Sox* motif located at the dyad axis of the middle nucleosome. Addition of WT PARP-1, but not a DNA binding domain mutant PARP-1 (DBD), enhances Sox2 binding as indicated by protection from DNAase I digestion (lane 7 versus 8 and 9 versus 10; see the band marked with a red asterisk). PARP-1 binding to the nucleosome is also evident (lanes 5 and 6 versus 7 and 8).

C) Quantification of the DNase I footprinting assays shown in panel B. The bands marked with a red asterisk in (B) were quantified using a phosphorimager. Each bar represents the mean plus the SEM, n = 3. Asterisks indicate significant differences: Student's t test, **p < 0.01 and *p < 0.05.

D) The DNA binding and BRCT domains of PARP-1 are required for enhancement of Sox2 binding to nucleosomes. *Top*, PARP-1 deletion mutants used in the nucleosome binding assays. *Bottom*, In vitro nucleosome binding assays using various PARP-1 deletion mutants, as in panel A.

E) Quantification of the nucleosome binding assay shown in panel D. The bands were quantified by densitometry. Each bar represents the mean plus the SEM, n = 3. Asterisks indicate significant differences: Student's t test, **p ≤ 0.01 and *p < 0.02.

F) PARP-1 stabilizes Sox2 binding to nucleosomes with or without linker DNA. In vitro nucleosome binding assay, as in panel A, using 601 NPE mononucleosomes with (Nuc 207) or without (Nuc 147) linker DNA.

2.3.6. Formation of a PARP-1-Sox2 complex on the nucleosome stabilizes the interaction between Sox2 and the nucleosome

We then tested the possibility that PARP-1-Sox2-nucleosome form a complex, which stabilizes association of Sox2 with nucleosome. We first asked whether PARP-1 and Sox2 directly interact with each other by performing an in vitro pulldown assay. Indeed, Sox2 efficiently co-immunoprecipitated with Sox2, indicating a strong direct interaction between the two proteins (Fig. 2.17, A-B). Interaction between PARP-1 and Sox2 is mediated by the BRCT domain of PARP-1, since deleting the BRCT domain

abolishes the interaction between PARP-1 and Sox2 (Fig. 2.17, C-E). Notably, the BRCT domain of PARP-1 is also required to stabilize Sox2 interaction with nucleosomes as demonstrated by nucleosome binding assay (Fig. 2.16, D-E). Therefore, PARP-1 directly interacts with Sox2 as well as nucleosomes. Abolishing the interaction between PARP-1 and Sox2 or PARP-1 and nucleosomes by BRCT deletion or DBD mutation disrupts the formation of PARP-1-Sox2-nucleosome complex, which further interrupts the enhancement of Sox2 binding to nucleosome by PARP-1.

Interestingly, although PARP-1 is known to interact with linker DNA when binding to nucleosomes^{5,24}, binding assays using mononucleosomes without linker DNA showed that linker DNA is not required for PARP-1 to enhance the binding of Sox2 (Fig. 2.16F). In fact, interactions with Sox2 on nucleosomes may redirect PARP-1 from the linker DNA to the adjacent sites of Sox2 binding on the nucleosome (Fig. 2.18). In the absence of Sox2, PARP-1 binds to linker DNA of the nucleosome, as indicated by increased protection from/resistance to cleavage by DNaseI (Fig. 2.18A, a; compare lanes 1 and 3). Adding PARP-1 enhances the binding of Sox2 to the *Sox* motif, as indicated by increased protection from/resistance to cleavage by DNaseI (Fig. 2.18A, c; compare lanes 2 and 4). Moreover, the addition of Sox2 increases the interaction of PARP-1 in regions adjacent to the *Sox* motif (*b*), while decreasing the interaction of PARP-1 with the linker DNA DNaseI (Fig. 2.18A, a; compare lanes 3 and 4).

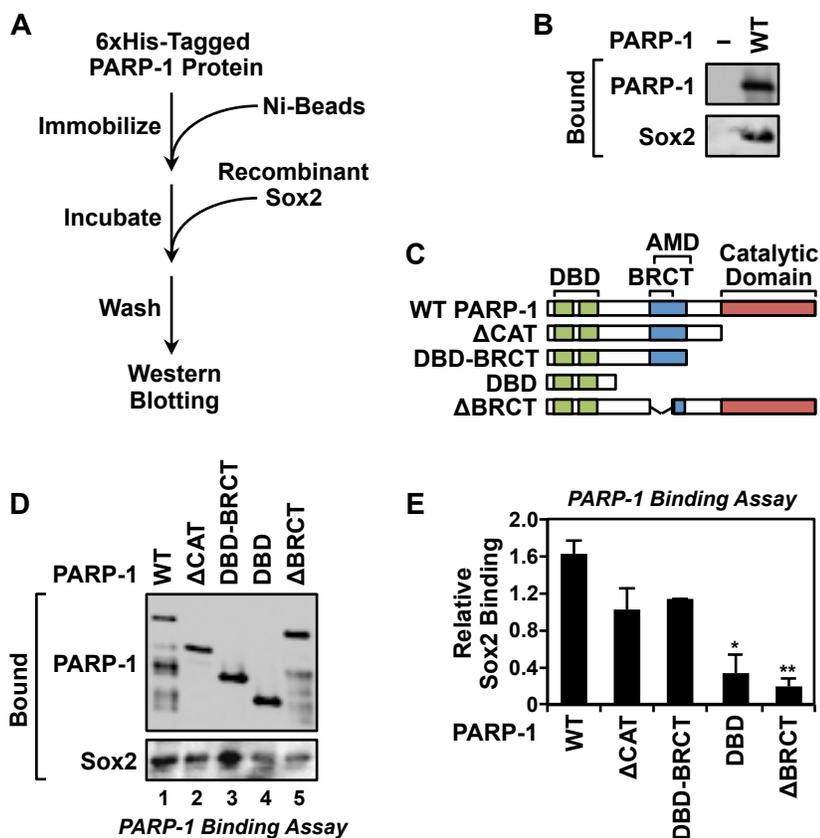


Figure 2.17. The PARP-1 BRCT domain mediates interactions between PARP-1 and Sox2.

A) Flow chart for the binding assay used to examine interactions between PARP-1 and Sox2.

B) PARP-1 interacts with Sox2 protein in vitro. Sox2 was incubated with immobilized PARP-1 as described in panel A. Bound material was analyzed by Western blotting for PARP-1 and Sox2, as indicated.

C) Schematic diagrams of the PARP-1 deletion proteins used in the in vitro binding assays shown in panels D and E.

D) The PARP-1 BRCT domain mediates interactions between PARP-1 and Sox2. Sox2 was incubated with immobilized wild-type or mutant PARP-1, as described in panel A. Bound material was analyzed by Western blotting for PARP-1 and Sox2, as indicated.

E) Quantification of the binding assays from panel D. The Sox2 bands were quantified by densitometry and the signals were normalized to the molarity of the PARP-1 protein immobilized on the beads. Each bar represents the mean plus the SEM, $n = 3$. Asterisks indicate significant differences: Student's t-test, * $p = 0.018$; ** $p = 0.005$.

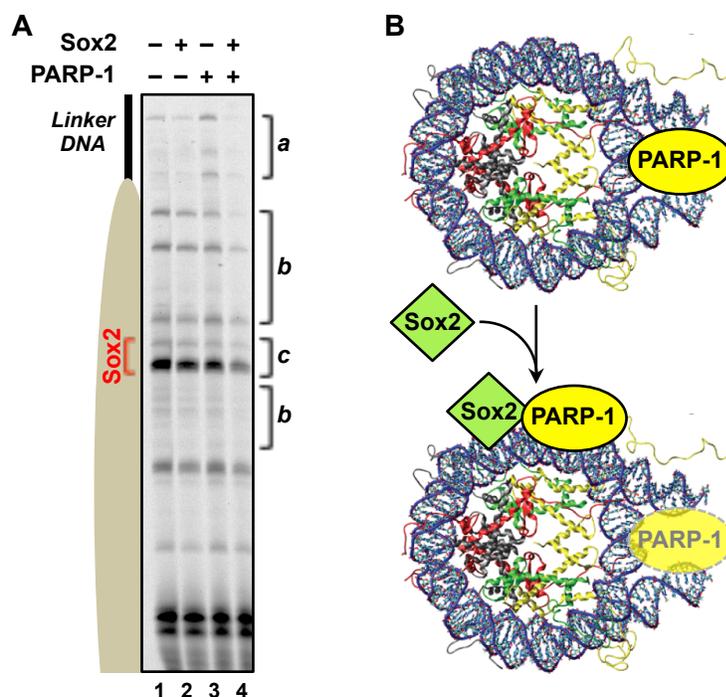


Figure 2.18. Sox2 redirects PARP-1 binding on the nucleosome.

A) DNaseI footprinting assay using a 601 NPE mononucleosome containing a *Sox* motif. The location of the nucleosome and the Sox2 binding site are shown. The results show that: (1) PARP-1 binds to linker DNA in the absence of Sox2, as we have shown previously⁵, as indicated by increased protection from/resistance to cleavage by DNaseI (*a*; compare lanes 1 and 3); (2) The addition of PARP-1 enhances the binding of Sox2 to the *Sox* motif, as indicated by increased protection from/resistance to cleavage by DNaseI (*c*; compare lanes 2 and 4); (3) The addition of Sox2 increases the interaction of PARP-1 in regions adjacent to the *Sox* motif (*b*), while decreasing the interaction of PARP-1 with the linker DNA (*a*; compare lanes 3 and 4). These results suggest that PARP-1 stabilizes Sox2 binding to nucleosomes by forming a complex with Sox2. In addition, these results suggest that Sox2 may redirect PARP-1 binding from the linker DNA to adjacent to the Sox2 binding site.

B) Schematic representation of the results from panel A, showing Sox2 redirecting PARP-1 binding from the linker DNA to the Sox2 binding site.

2.3.7. Regulation of Sox2 interaction with nucleosome by the position of the *Sox* motif

The structural features of nucleosomal DNA are determined by the rotational phasing of the DNA relative to the histone octamer. In this regard, we asked whether the position of the *Sox* motif along the length of the nucleosomal DNA affects the dependency of Sox2 on PARP-1 for binding to the nucleosome. To answer this question, we created a set of ten nucleosomal templates, each with the *Sox* motif located at a different position in

the DNA (Fig. 2.20A). Some of the *Sox* motifs were located in narrow minor grooves facing toward the histone octamer, while others were located in wide minor grooves facing away from the histone octamer (Fig. 2.20A). We observed a striking correlation of Sox2 binding to the nucleosome with the structural features of the nucleosomal DNA in which the *Sox* motif was located. When the first three bases of the *Sox* motif, which are likely to mediate interactions with Sox2 in the nucleosomal context¹⁴, were located in a narrow minor groove, Sox2 bound in a PARP-1-independent manner (Fig. 2.20B, constructs 14, 27, 56, and 64; Fig. 18C). In contrast, when the first three bases of the *Sox* motif were located in a wide minor groove, Sox2 bound in a PARP-1-dependent manner (Fig. 2.20B, constructs 0, 2, 10, 32, 44, and 62; Fig. 2.20C). A similar effect of PARP-1 on the binding of Oct4 was not observed (Fig. 2.10B). Interestingly, previous studies have shown that the binding of arginine residues in the DBD to narrow minor grooves in DNA is an important mode of protein-DNA recognition for many transcription factors^{25,26}, a result that may explain our observations with Sox2. Together, our results suggest that DNA shape can play important role in directing transcription factor binding to nucleosomes.

These results show that the enhancement of Sox2 binding to nucleosomes by PARP-1 is limited to those sites located in wide minor grooves facing away from the histone octamer. Furthermore, they indicate that PARP-1 acts to counteract the effects of minor groove width on Sox2 binding (Fig. 2.20C). Interestingly, a similar effect can be discerned for native nucleosomes examined across the genome of mESCs. We identified *Sox* motifs under Sox2 ChIP-seq peaks located within nucleosomes defined by MNase-seq (Fig. 2.20D). We then calculated the nucleosome rotational positioning (NRP) score, which indicates the orientation of DNA with respect to the histone octamer, around each nucleosomal *Sox* motif that was identified (Fig. 2.20D). A higher NRP score indicates a greater tendency for the DNA sequence to face away from the histone octamer²⁷. Native nucleosomal PARP-1-dependent Sox2 binding sites exhibited significantly higher NRP scores than PARP-1-independent Sox2 binding sites, notably at the first three bases of the *Sox* motif (Fig. 2.20E). These results provide a strong link between our observations about the effects of PARP-1 on Sox2 binding to nucleosomes in vitro and with our observations in cells.

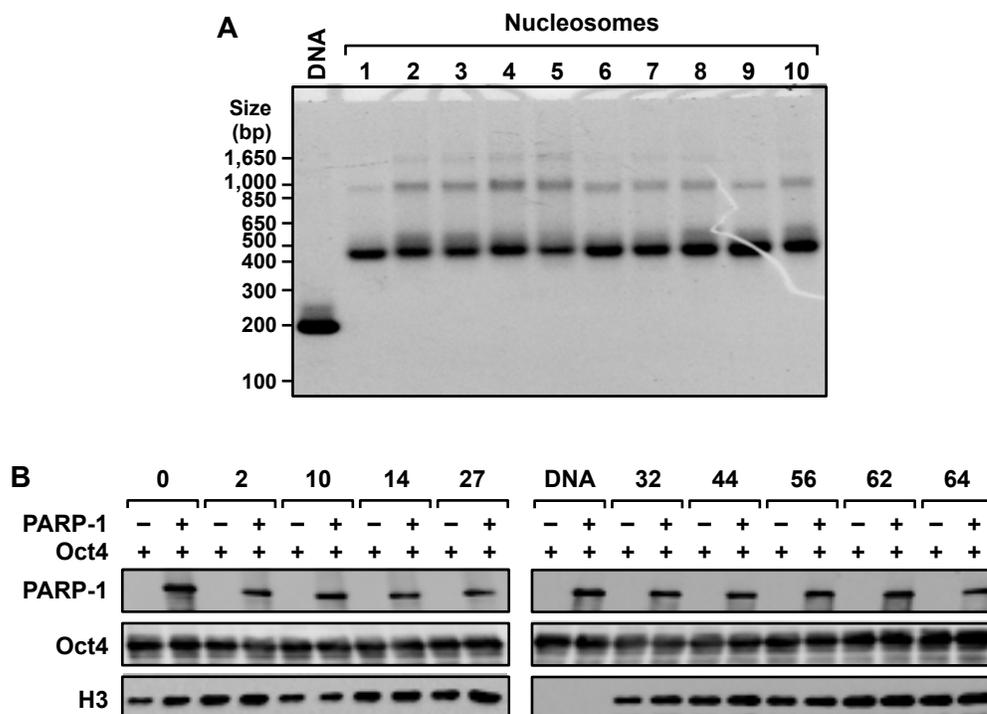


Figure 2.19. Assembly of mononucleosomes containing a natural *Pou/Sox* motif from the *Nanog* promoter and binding by Oct4 to the mononucleosomes.

A) Assembly of biotinylated mononucleosomes used in the nucleosome binding assays shown in Fig. 4B, as well as panel B below. The nucleosomes were assembled using DNA containing the 601 NPE, as well as a natural combined *Pou/Sox* motif (TTTGCATTACAATG; red, *Pou*; blue, *Sox*) from the *Nanog* promoter located at the following positions relative to the dyad axis: (1) 0, (2) 2, (3) 10, (4) 14, (5) 27, (6) 32, (7) 44, (8) 56, (9) 62, (10) 64 (numbering is relative to the “C” in the Sox motif). Native polyacrylamide gel showing biotinylated 601 nucleosome positioning element (NPE) DNA (“DNA”) or the same DNA assembled into mononucleosomes. The histone-DNA complex of the mononucleosome migrates more slowly than the free DNA.

B) Oct4 binds to mononucleosomes independent of PARP-1 and irrespective of the location of the *Oct* motif. The assays were performed as described in Fig. 4B using purified recombinant Oct4. Biotin-labeled 601 NPE DNA, or the same DNA assembled into a mononucleosome, was immobilized on streptavidin beads and incubated \pm recombinant Oct4 and PARP-1, as indicated. After washing, the bound proteins were analyzed by Western blotting.

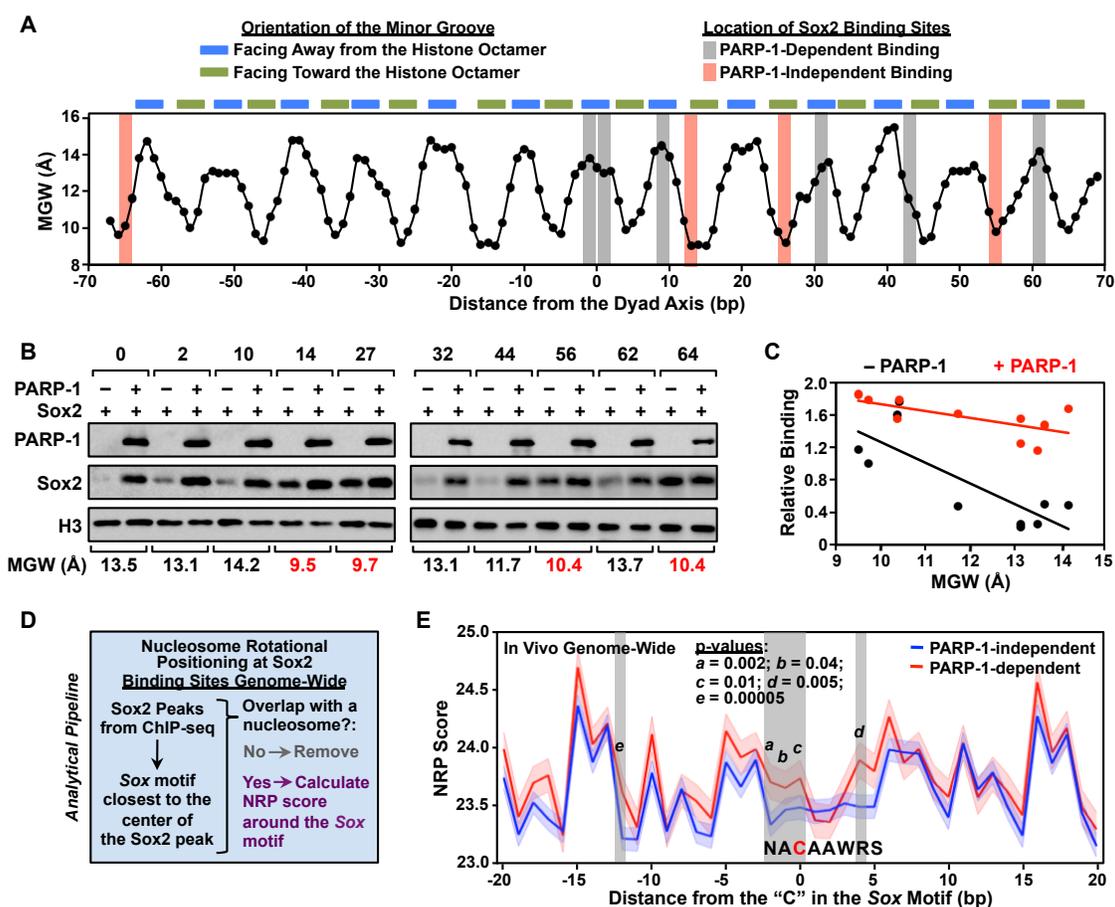


Figure 2.20. PARP-1 facilitates Sox2 binding to nucleosomes at Sox sites located in wide, outward-facing minor grooves.

A) Minor groove width (MGW) along the DNA of the 601 NPE nucleosome, as determined from high resolution X-ray crystal structures (PBD: 3LZ0, 3LZ1, 3MVD). The position of the dyad axis is set to 0. Blue and green bars above the graph indicate locations across the nucleosome where the minor groove is facing away from the histone octamer (*blue*) or toward the histone octamer (*green*). Vertical grey/red shading indicates the position of the “TAC” in the first half of the *Sox* motif (TACAATG) in the various nucleosome templates tested in panel B. *Grey*, *Sox* motifs located at these positions (large MGW, generally facing away from the histone octamer) support PARP-1-dependent binding of Sox2; *Red*, *Sox* motifs located in these positions (small MGW, generally facing toward the histone octamer) support PARP-1-independent binding of Sox2.

B) The location of the *Sox* motif determines PARP-1-dependence of Sox2 binding to nucleosomes in vitro. In vitro nucleosome binding assays, as in Fig 3A, using a set of nucleosome templates with a *Sox* motif placed in different locations across the nucleosome, as indicated in panel A. The numbering is based on the distance of the “C” in the first half of the *Sox* motif to the dyad axis of the nucleosome. The average MGW in Å for the “TAC” in the first half of the *Sox* motif of each nucleosome template is indicated.

C) Correlation between Sox2 binding and the average MGW for the “TAC” in the first half of the *Sox* motif. The data were quantified from a minimum of three nucleosome binding assays like those shown in (B). Sox2 binding was normalized to histone H3. *Black*, Sox2 alone; *Red*, Sox2 in the presence of PARP-1.

D) Pipeline for calculating the nucleosome rotational positioning (NRP) scores of *Sox* motifs associated with Sox2 binding sites located in nucleosomes, as determined by ChIP-seq in mESCs.

E) NRP scores associated with PARP-1-dependent and PARP-1-independent Sox2 binding sites in mESCs. The x-axis represents position in nucleotides of the *Sox* motif associated with the Sox2 binding site, as determined by ChIP-seq, with the “C” in the first half of the *Sox* motif set to 0. The y-axis represent average NRP score associated with each nucleotide position (dark lines), with the SEM range indicated by lighter shading. Vertical grey shading highlights regions exhibiting a significant difference in NRP score between PARP-1-dependent and PARP-1-independent Sox2 binding sites. Higher NRP scores indicate a greater tendency for the sequence to face away from the histone octamer.

2.4. Discussion

Our results demonstrate a potent effect of PARP-1 on the binding of Sox2 to a subset of its genomic binding sites, which are suboptimal with respect to the chromatin state and the location of the underlying *Sox* motif in the genomic DNA. The PARP-1-dependent binding of Sox2 to these sites drives the transcription of key genes whose expression helps to maintain the pluripotent phenotype of mESCs. Collectively, our results show that PARP-1 promotes the binding of Sox2 to inaccessible nucleosomes located in closed regions of chromatin co-occupied by few, if any, partner transcription factors (Fig. 2.21, A and B). Furthermore, our results show that PARP-1 helps Sox2 overcome the barrier to interact with *Sox* motifs with an orientation unfavorable to Sox2 binding (Fig. 2.21, C and D). These effects are mediated, in part, by the PARP-1 DBD, which allows PARP-1 to bind to nucleosomes, and the PARP-1 BRCT domain, which promotes interactions between PARP-1 and Sox2. We surmise that PARP-1 may act in two ways, which are not mutually exclusive: it may (1) fine tune the structure of the nucleosome to make *Sox* motif more accessible to Sox2 or (2) alter the structure of the Sox2 DBD to facilitate binding to *Sox* motifs located in nucleosomes. Our results illustrate how the actions of a chromatin-modulating protein, such as PARP-1, can act at the level of the nucleosome to produce global effects on transcription factor binding and biologically important gene expression outcomes.

2.4.1. Modulation of Oct4 and Sox2 DNA binding by nucleosome occupancy

Nucleosomes have long been considered as a barrier inhibiting transcription factor access to its target sequence^{28 29}. In order for transcription factors to bind to DNA, nucleosome needs to be either ejected or moved away by chromatin remodeling proteins such as SWI/SNF complex to expose the regulatory sequence^{28,30}. Recent studies suggested that certain transcription factors such as Oct4 and Sox2 have pioneer factor activity since they are capable of overcoming nucleosomal barrier and interact with their binding sites^{31 14}. Sox2 alone is able to bind to a Lin28 nucleosome, with binding affinity comparable to naked DNA¹⁴. In contrast, our study suggests that Sox2 alone cannot stably bind to nucleosome. Instead, it requires PARP-1 to interact with nucleosome in a specific and stable manner. One possible explanation to this discrepancy is that a more stringent binding condition, 150 mM NaCl and 0.1% NP-40, is used in this study, while the previous study used 10 mM KCl and no detergent¹⁴. In addition, since previous study used native Lin28 DNA sequence to assemble nucleosome, it is unknown where the Sox2 motif is positioned relative to nucleosome. In fact, our results indicate that when the *Sox* motif is positioned in certain areas of nucleosome, it directs stronger and PARP-1-independent Sox2 binding (Fig. 2.20). This is different from the case of Oct4, which binds to its motif irrespective of where the motif sequence is located in the nucleosome (Fig. 2.19).

2.4.2. Regulation of Sox2 binding to nucleosome by position of *Sox* motif in the nucleosome

The affinity of DNA binding sequence to transcription factor regulated by DNA shape feature has been characterized in a number of studies^{25 26}. DNA structure, specifically the minor groove width of the DNA sequence, acts as an independent factor in addition to the DNA sequence itself and determines the specificity and affinity of transcription factor association. Interestingly, in our study, we observed a strong correlation between the position of *Sox* motif and its affinity with Sox2. That is, when the first three nucleotides of *Sox* motif are positioned in narrow minor groove width region facing towards histones, Sox2 binding affinity is stronger. As an explanation for this, narrow minor groove of DNA may stabilize the interaction with the arginine residues in

the HMG domain of Sox2. This suggests that DNA shape may play important role in directing transcription factor binding in a nucleosomal complex.

One potential caveat for this model, however, is that when nucleosomal DNA has a narrow minor groove width, the same DNA segment also faces towards histone, which might hinder the interaction between transcription factor and nucleosomal DNA ³². On the other hand, one other interpretation for the position-dependent *Sox* motif binding affinity to Sox2 exists: the interaction between *Sox* motif and Sox2 protein is mainly determined by the second half part of the *Sox* motif. Therefore, when this part of the *Sox* motif faces outwards of histone, it is more accessible to Sox2. Although Soufi et al. showed using DNase footprinting assay that instead of the second half, it is the first half part of the *Sox* motif that mediates the interaction with Sox2 in a nucleosome ¹⁴, we still cannot rule out the possibility that the second half of the *Sox* motif is required for interaction with Sox2. Further analyses are needed to pin point the nucleotides mediating the interaction between *Sox* motif and Sox2 in a nucleosome in order to better understand the relation between *Sox* motif position in the nucleosome and its affinity to Sox2.

2.4.3. Role of PARP-1 PARylation in regulating chromatin association of Sox2

PARylation activity of PARP-1 is usually required for the function of PARP-1. For example, regulating KDM5B requires PARylation of KDM5B protein, which further inhibits the activity of KDM5B as well as blocking its association with chromatin ⁷. On the other hand, PARP-1 can also function in a PARylation activity-independent manner. One example for this is the regulation of transcription factor E2F1, where PARP-1 interacts with E2F1 independent of its DBD and catalytic domain and regulates the DNA binding of E2F1 ⁹. The PARylation activity-dependent or –independent mechanisms are not mutually exclusive. Instead, signal-induced changes of PARylation activity can trigger the switch of the mechanism used by PARP-1 to regulate the same proteins during different cell states. In this study, we identified the function of PARP-1 as a pioneer factor in undifferentiated embryonic stem cells, forming a complex with Sox2 and promotes the interaction between Sox2 and nucleosome. This does not require PARP-1 enzymatic activity, since 1) PARylation activity of PARP-1 is low in embryonic stem cells compared to differentiating ES cells, 2) inhibiting PARP-1 activity does not have the same gene expression changes as PARP-1 depletion, 3) deleting catalytic domain has no affect on

PARP-1's function in regulating Sox2 binding to nucleosome. On the other hand, PARP-1 PARylation activity dramatically increases when ES cells undergo differentiation (Fig 2.4A). Enhanced PARylation activity of PARP-1 may in turn use a different mechanism to regulate Sox2 in differentiating ES cells. In this regard, PARylation of Sox2 has been reported to promote Sox2 degradation in differentiating ES cells¹⁵. In addition, increased PARylation level of PARP-1 decreases the affinity between PARP-1 and chromatin, and increases the affinity between PARP-1 and Sox2, which was suggested to inhibit the Sox2 interaction with chromatin³³. Therefore, under various cell states, PARP-1 with different PARylation activity status can play different role to the same target protein. This adds one more layer to the regulatory functions of PARP-1 in response to different cellular cues.

2.4.4. Potential mechanisms of PARP-1 regulating somatic reprogramming

The function of PARP-1 in regulating iPSC generation has been reported in multiple studies. PARP-1 and its PARylation activity are required for efficiently reprogramming somatic cells into induced pluripotent cells². In addition, the authors showed that PARP-1 depletion causes decreased level of histone mark H3K4me2, increased H3K27me3, as well as decreased Oct4 chromatin binding². However, it is not clear whether the decreased transcription factor binding is due to changed histone mark level or it is the other way around, especially given the fact that pluripotency factors Oct4, Klf4, and Sox2 are able to access to closed chromatin during early stages of iPSC generation¹³. Our study has provided one potential explanation of how PARP-1 promotes reprogramming. That is, PARP-1 acts as a pioneer factor facilitating Sox2 access to closed chromatin region. However, one argument against this model is that since PARylation activity is not required for the pioneer factor activity of PARP-1, why does blocking PARylation activity also inhibit somatic reprogramming²? One explanation to this might be that during reprogramming, transcription factor binding profile undergoes dynamic changes. Instead of constantly stably binding to chromatin, OSKM factors are subject to various cellular signaling pathways and might progressively change their binding sites. This can be demonstrated by the dramatically different OSKM binding profile between early stage of somatic reprogramming and the end point of reprogramming – iPS cells¹³. Since PARylation can be a mechanism of regulating chromatin association of PARP-1 and its

target proteins, it is speculated that maintaining PARylation activity is also critical in order to allow the dynamic chromatin binding of OSKM factors during iPS generation.

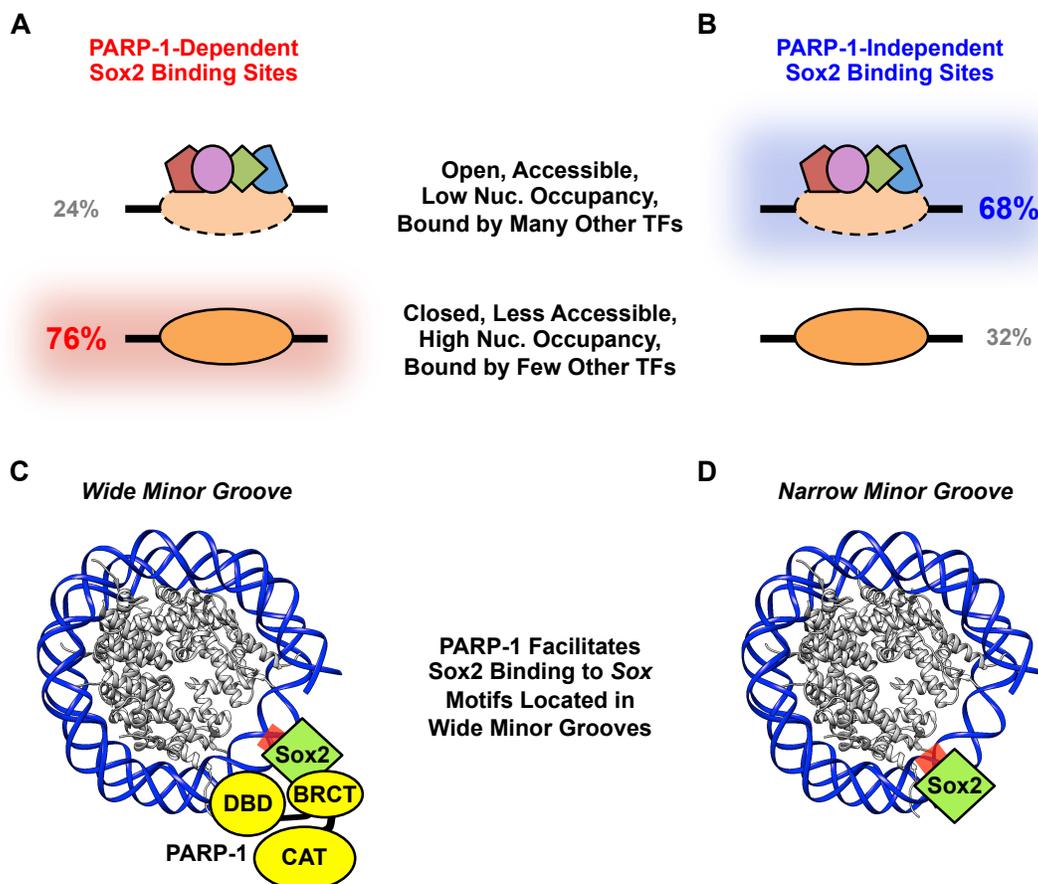


Figure 2.21. Model for PARP-1-dependent binding of Sox2 to nucleosomes.

A and B) Frequency of (A) PARP-1-dependent and (B) PARP-1-independent Sox2 binding sites with the features indicated.

C and D) Model for the mechanism of Sox2 binding at (C) PARP-1-dependent and (D) PARP-1-independent sites. The Sox motifs for PARP-1-dependent Sox2 binding sites typically located in wider minor grooves than the Sox motifs for PARP-1-independent Sox2 binding sites. PARP-1 interacts with nucleosomal DNA through its DBD and Sox2 through its BRCT motif.

2.5. Materials and Methods

Culture and Differentiation of Mouse Embryonic Stem Cells (mESCs)

Parp1^{+/+} (wild-type; WT) and *Parp1*^{-/-} (PARP-1 KO) mESCs were kindly provided by Dr. Zhao-Qi Wang (Leibniz Institute for Age Research, Fritz Lipmann Institute)¹⁵. The cells were maintained on a feeder layer in mESC growth medium: Dulbecco's modified Eagle's medium (Gibco; 11965) containing 2 mM L-glutamine (Sigma), supplemented with 15% (v/v) FBS (Atlanta Biologicals; S12450), 0.1 mM nonessential amino acids (Sigma; M7145), 1 mM sodium pyruvate (Invitrogen; 11360070), 1000 units/mL mLIF (Millipore; ESG1106), penicillin and streptomycin (Invitrogen; 15140), and 0.1 mM β -mercaptoethanol (Sigma). The cells were passaged twice to eliminate feeder cells before being used in experiments. To differentiate the mESCs into embryoid bodies, the cells were separated from the feeder layer before culturing in 10 cm diameter ultra low attachment plates at a density of 4×10^6 cells per plate in mESC growth medium containing 15% (v/v) FBS without mLIF. The medium was changed every other day prior to collection of the cells for experiments.

Antibodies

The antibodies used for Western blotting, co-IP, and/or ChIP were as follows: PARP-1 (previously characterized custom rabbit polyclonal antibody⁵; now available from Active Motif; 39561), PAR (Trevigen; 4335-AMC), Sox2 (Santa Cruz; sc-17320), Oct4 (Santa Cruz; sc-8628), Nanog (Abcam; ab8092), H3K4me3 (Active Motif; 3915), H3K27me3 (Millipore; 170622), β -actin (Sigma; A5316), Flag (Sigma; F3165), histone H3 (Abcam; ab1791).

Analysis of Protein and PARylation Levels by Western Blotting

Whole cell extracts (WCEs) were prepared from mESCs using cell lysis buffer [50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 1.0 mM EDTA, 1% NP-40, and 10% glycerol, 1x protease inhibitor cocktail (Roche)] containing 1 μ M PJ34 (to inhibits PARP activity) and 100 μ M tannic acid (to inhibit PARG activity). After clarification of the WCEs by centrifugation, aliquots containing equal amounts of total protein, as determined by a BCA assay (Pierce), were run on 8 to 10% polyacrylamide-SDS gels, transferred to nitrocellulose membrane, and subjected to Western blotting using the antibodies listed above and a chemiluminescent detection system (Thermo scientific).

RNA Extraction and RT-qPCR

RNA isolation and RT-qPCR were performed using a standard protocol, as previously described³⁴. Briefly, total RNA was extracted from mESCs using Trizol reagent (Life Technologies) following the manufacturer's instructions, reverse transcribed, and subjected to qPCR using the gene-specific primers listed below. Unless specified, all target gene expression levels were normalized to β -actin mRNA. All RT-qPCR assays were performed a minimum of three times to ensure reproducibility.

RNA-seq Library Preparation

Total RNA was isolated as described above. The integrity of the RNA was assessed and verified using an Experion Automated Electrophoresis System (Bio-Rad) before mRNA-seq libraries were prepared using methods described previously³⁵. Briefly, polyA+ RNA was enriched using Dynabeads oligo(dT)25 (Invitrogen), heat fragmented, and reverse transcribed using random hexamers in the presence of dNTPs. Second strand cDNA synthesis was performed with dNTPs, but replacing dTTP with dUTP. After end-repair, dA-tailing, ligation to adaptors containing barcode sequences, and size selection using AMPure beads (Agencourt), the synthesized second-strand was digested using uracil DNA glycosylase (Enzymatics). A final PCR reaction was performed using Phusion high-fidelity DNA polymerase (NEB). After library quality control assessment using a Bioanalyzer (Agilent), the samples were subjected to 50 bp sequencing using an Illumina HiSeq 2000 Sequencing System.

Analysis of mRNA-seq Data

mRNA-seq reads were subjected to quality-control using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), trimmed to remove adapter sequences, and aligned to the mouse reference genome (mm9) using Tophat³⁶ before transcript assembly using Cufflinks³⁷. The data were then converted to wiggle (WIG) file format using PeakRanger³⁸ for visualization using the Integrative Genomics Viewer (IGV2.3)³⁹. Cuffdiff³⁷ was used to identify genes that showed significant differential regulation upon PARP-1 depletion, with a false discovery rate (FDR) cutoff of 5%. The

expression data were visualized in heatmaps using Java TreeView⁴⁰, ranked in order based on the the log₂ of the PARP-1-1 KO to WT RPKM ratio.

The DAVID bioinformatics tool (was used for gene ontology analysis. Gene set enrichment analysis (GSEA) tools developed by broad institute were used for the GSEA analyses (<http://www.broadinstitute.org/gsea/index.jsp>).

Chromatin Immunoprecipitation and ChIP-qPCR

mESCs were passaged twice without a feeder layer and then grown in geletin-coated plates to ~70 to 80% confluence. The cells were cross-linked using 1% paraformaldehyde at room temperature for 10 min., followed by quenching in 125 mM glycine for 5 min. at 4°C. The crosslinked cells were collected, and the nuclei were released by gentle pipetting three times in lysis buffer [10 mM Tris-HCl pH 7.5, 2 mM MgCl₂, 3 mM CaCl₂, 0.5% NP-40, 10% glycerol, 1 mM DTT, 1x protease inhibitor cocktail (Roche)]. The nuclei were collected by gentle centrifugation and resuspended in sonication buffer [1x PBS, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 1x protease inhibitor cocktail (Roche)]. The nuclei were then incubated in sonication buffer on ice for 10 min., and sonicated at 4°C using a Biorupter (Diagenode) high setting, three cycles of 5 min. sonication (30 seconds on/30 seconds off) with 5 min. intervals) to generate genomic DNA fragments of 200 - 500 bp in length. The sonicated chromatin was clarified by centrifugation and pre-cleared with agarose beads.

Aliquots of the pre-cleared chromatin were immunoprecipitated with various antibodies at 4°C overnight, followed by collection of the immunoprecipitates using protein A (Millipore) or G (Invitrogen) agarose beads (protein A beads for H3K4me₃, H3K27me₃, and PARP-1 antibodies; protein G beads for Sox2 and Oct4 antibodies). The beads were collected by gentle centrifugation and washed in low salt wash buffer (20 mM Tris-HCl pH 7.9, 2 mM EDTA, 125 mM NaCl, 0.05% SDS, 1% Triton X-100, 1x protease inhibitor cocktail), high salt wash buffer (low salt wash buffer containing 500 mM NaCl), and LiCl wash buffer (10 mM Tris-HCl pH 7.9, 1 mM EDTA, 250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1x protease inhibitor cocktail). The immunoprecipitated genomic DNA was eluted and the crosslinks were reversed by incubation in elution buffer (100 mM NaHCO₃, 1% SDS) at 65°C overnight. The genomic DNA was then

deproteinized by digestion with proteinase K and extraction with phenol:chloroform:isoamyl alcohol, followed by precipitation with ethanol. The ChIPed DNA was then subjected to qPCR using the locus-specific primers listed below. All ChIP-qPCR assays were performed a minimum of three times to ensure reproducibility.

ChIP-seq Library Preparation

Approximately 50 ng of ChIPed DNA (quantified using a NanoDrop) was used to prepare each ChIP-seq library using methods described previously¹⁸. Briefly, the genomic DNA fragments were end-polished, dA-tailed, and ligated to Y-adaptors containing barcode sequences. After agarose gel-based size selection and purification, the DNA was amplified for 13 - 15 cycles by PCR using Phusion high-fidelity DNA polymerase (NEB). The final ChIP-seq libraries were subjected to quality control assessment using a Bioanalyzer (Agilent), followed by 50 bp sequencing using an Illumina HiSeq 2000 Sequencing System.

ChIP-seq Data Analysis

Read alignment and peak calling. ChIP-seq reads were subjected to quality-control using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), trimmed to remove adapter sequences, and aligned to the mouse reference genome (mm9) using Bowtie 0.12.7⁴¹, allowing for one mismatch during alignment. Significant peaks of PARP-1 enrichment in WT mESCs were called using MACS 1.4.2¹⁶, with a p-value cutoff of 0.001. Significant peaks of Sox2 and Oct4 enrichment in WT mESCs were called using PeakRanger 1.1³⁸ with a p-value cutoff of 1×10^{-7} . Genomic DNA isolated from sonicated chromatin without immunoprecipitation was used as an input control for identifying regions of enrichment. The data were then converted to wiggle (WIG) file format using PeakRanger³⁸ for visualization using the Integrative Genomics Viewer (IGV2.3)³⁹.

Generating a correlation matrix of transcription and chromatin factor enrichment. For each of the 13 transcription and chromatin factors included in this analysis (see Fig. 1D), significantly enriched regions were called using MACS and merged to generate a total number of 87,487 genomic loci showing enrichment for at least one factor. Read counts

for each factor at the 87,487 loci were determined and normalized to (1) the total number of aligned reads for that factor and (2) the length of each locus. Pearson correlation coefficients for all possible pairwise comparisons of factors were calculated based on the normalized reads, which were further converted to distance values. A larger distance value represents a lower correlation. Hierarchical clustering based on the distance values was performed using the “hclust” function in R, using the complete linkage method. The factors in the correlation matrix were ordered based on hierarchical clustering and the correlation matrix was visualized using Java TreeView⁴⁰.

Heatmaps. ChIP-seq read densities were visualized in heatmaps using Java TreeView⁴⁰. For the heatmaps of H3K4me3 and H3K27me3, we determined the read densities in a 10 kb window (± 5 kb) around the transcription start sites (TSSs) of the PARP-1 regulated genes, using a 20 bp moving window. For the heatmaps of Sox2 and PARP-1, we determined the read densities for Sox2 in a 4 kb window (± 2 kb) around the Sox2 peak summit using a 20 bp moving window. Corresponding PARP-1 read densities in the corresponding regions were determined for the same windows and plotted in the same order as the Sox2 data.

Box plots. Read intensities for H3K4me3 and H3K27me3 in a 1 kb window (± 0.5 kb) around the TSSs of RefSeq genes were determined and plotted using the box plot function in R.

Determining changes in Sox2 binding. Sox2 binding sites showing altered enrichment upon PARP-1 depletion were identified by mining the Sox2 peaks identified using PeakRanger in WT mESCs. The read counts across those peaks in both replicates of the Sox2 ChIP-seq data from the WT and PARP-1 KO samples were determined for a 400 bp region (± 200 bp) around the peak summit. Read counts were further normalized to the total reads in sample. Sox2 and Oct4 binding regions significantly changed upon PARP-1 depletion were identified using edgeR⁴² with a p-value cutoff of 0.01. To define PARP-1-independent Sox2 binding sites, Rc (Reads change) values of all Sox2 peaks were calculated using the following formula:

$$Rc = \log(Tp/Tw)$$

where T_p and T_w represent the binding intensity in the PARP-1 KO and WT samples, respectively¹⁸. Peaks with the highest tercile of R_c values were defined as PARP-1-independent binding sites.

Additional Genomic Data Analyses

Relationship between Sox2 binding and gene expression. To identify genes possibly regulated by adjacent Sox2 binding sites, each Sox2 binding site was assigned to the nearest RefSeq gene TSS within 10 kb (“associated genes”). For the GSEA analysis^{43,44}, we used genes associated with PARP-1-dependent Sox2 binding sites and their associated expression profiles from both replicates of WT and PARP-1 KO RNA-seq data. An equal number of randomly selected genes were used as control. Gene set enrichment analysis (GSEA) tools developed by broad institute were used for the GSEA analyses (<http://www.broadinstitute.org/gsea/index.jsp>).

Motif analyses. To identify the occurrence of *sox* motifs under Sox2 ChIP-seq peaks, we used the DNA sequences in 100 bp (\pm 50 bp) regions around the Sox2 peak summits as input in a directed search for the consensus *sox* motif ACAAWRS using FIMO in the MEME suite⁴⁵ with a p-value cutoff of 0.01 used to determine the association of a *sox* motif with the respective peak. When multiple ACAAWRS motifs were present in one peak, the motif with the lowest p-value and least number of mismatches was assigned to the peak. When multiple motifs with the same p-value and number of mismatches were found in one peak, the motif nearest to the peak summit was chosen.

To identify Sox2 peaks associated with the *Sox/Pou* motif, we performed a directed search for the *pou* motif ATGCWRA in regions 30 bp upstream and downstream of the *sox* motifs assigned to each Sox2 peak, with a p-value cutoff of 0.01 used to determine the association of a *pou* motif with the respective peak. The location of the *Sox/Pou* motif center within the same peak was then calculated. Optimal *Sox/Pou* motifs were defined as those *sox* and *pou* motifs are on opposite strands and are separated by 1 bp.

Analysis of transcription factors located near Sox2 binding sites. Nanog, Tcf3, Klf4, Esrrb, and Stat3 ChIP-seq data sets from mESCs were obtained from public data repositories (see below). The data were aligned and peaks called as described above for Sox2 and Oct4. Those transcription factors with peak summit distances <100 bp for a

Sox2 peak summit were defined as “colocalized” with Sox2 at that locus. For the heatmap showing the extent of colocalization with Sox2, we generated a binary matrix, where “0” indicates that there was no colocalization of the transcription factor with Sox2, while “1” indicates that there was colocalization. The Sox2 binding sites were ranked by hierarchical clustering using Cluster⁴⁶, with an uncentered correlation similarity metric using the single linkage method. Java TreeView⁴⁰ was used for visualization.

MNase-seq and DNase-seq data analysis. MNase-seq data⁴⁷ and DNase-seq data⁴⁸ from mESCs were obtained from public data repositories (see below). In both cases, the sequencing reads were aligned to the mouse reference genome (mm9) using Bowtie 0.12.7⁴¹. For the MNase-seq data, uniquely aligned reads were used to call nucleosome occupancy using DANPOS-2.1.2⁴⁹. Metaplots of nucleosome occupancy were generated by calculating the average nucleosome density for a 4 kb (± 2 kb) region around all Sox2 peak summits using a moving window of 20 bp. For the DNase-seq data, the metaplots were generated by calculating the average read counts for a 4 kb (± 2 kb) region around all Sox2 peak summits using a moving window of 20 bp.

Predicting nucleosome phasing of Sox motifs by calculating nucleosome rotational positioning (NRP) scores. To calculate NRP scores based on DNA sequences, Sox motifs associated with Sox2 binding sites from ChIP-seq data (within 20 bp of the Sox2 peak) were searched as described above. The Sox motifs located within a nucleosome, as determined from MNase-seq data, were then selected for further analysis. To calculate the NRP score of a particular nucleotide at position n , we took the DNA sequence from $n-73$ to $n+73$ into consideration. We then defined 14 minor-groove bending sites ($n\pm 4\sim n\pm 7$, $n\pm 14\sim n\pm 17$, $n\pm 24\sim n\pm 27$, $n\pm 34\sim n\pm 37$, $n\pm 45\sim n\pm 48$, $n\pm 56\sim n\pm 59$, $n\pm 66\sim n\pm 69$) and 12 major-groove bending sites ($n\pm 9\sim n\pm 12$, $n\pm 19\sim n\pm 22$, $n\pm 29\sim n\pm 32$, $n\pm 39\sim n\pm 42$, $n\pm 50\sim n\pm 53$, $n\pm 61\sim n\pm 64$), as described previously⁵⁰. The NRP score of a certain sequence motif x was calculated by summing the counts C_i for the motifs occurring at all minor groove or major groove bending sites:

$$S(n) = \sum_{i=1}^{26} w_i C_i,$$

where w_i is the weight of bending site i . NRP scores of sequence motifs WW, WWW, SS, SSS, YR, YYRR, RYRY were calculated and summed to determine the NRP score at position n ²⁷. The higher the NRP score of a particular nucleotide, the more likely it is to bend into major groove²⁷..

Protein Co-immunoprecipitation Assays

To explore potential interactions of PARP-1 with Sox2 and Oct4, HEK293T cells were transiently transfected with a plasmid for expressing Flag-tagged PARP-1 together with a plasmid for expressing either Sox2 or Oct4. The cells were collected and lysed by incubating on ice for 30 min in lysis buffer [25 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, 10% glycerol, 1x protease inhibitor cocktail (Roche)] with intermittent gentle inversion. After clarification of the lysates by centrifugation, aliquots containing equal amounts of protein were subjected to PARP-1 co-immunoprecipitation in the same buffer by adding 5 μ l Flag monoclonal antibody (Sigma) and incubating in 4°C overnight. The immunoprecipitates were collected by incubation with protein A/G agarose beads for 2 hours at 4°C. The beads were collected by gentle centrifugation, washed three times using wash buffer [25 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5% NP-40, 10% glycerol, 1x protease inhibitor cocktail (Roche)], and boiled in SDS-PAGE loading solution. The eluted immunoprecipitates were run on 8% polyacrylamide-SDS gels, transferred to nitrocellulose, and subjected to Western blotting using antibodies to PARP-1, Sox2, and Oct4 and a chemiluminescent detection system (Thermo Scientific).

Expression and Purification of Recombinant PARP-1 and Sox2

Purification of proteins expressed in insect cells. Flag tagged human PARP-1 proteins (wild-type and DBD point mutant) were expressed in insect cells using a baculovirus expression system and purified by using Flag-affinity purification, as described previously⁵. Briefly, Sf9 cells infected with a baculovirus vector for expression of recombinant PARP-1 were collected, resuspended in lysis buffer (20 mM Tris-HCl pH 7.5, 0.5 M NaCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 20% glycerol, 2 mM DTT, 1 mM PMSF, 20 μ g/ml leupeptin, 20 μ g/ml aprotinin), and homogenized by douncing 15 times on ice using a tight pestle. After clarifying the lysate by centrifugation, the supernatant was collected

and mixed with an equal volume of dilution buffer (20 mM Tris-HCl pH 7.5, 1.5 mM MgCl₂, 0.2 mM EDTA, 10% glycerol, 2 mM DTT, 1 mM PMSF, 20 µg/ml leupeptin, 20 µg/ml aprotinin). The diluted supernatant was mixed with α-Flag M2 affinity resin (Sigma) and incubated for 3 hours at 4°C. The resin was then collected by gentle centrifugation and washed four times in wash buffer (20 mM Tris-HCl pH 7.5, 1 M NaCl, 10% glycerol, 0.2 mM EDTA, 1.5 mM MgCl₂, 0.2% IGEPAL, 2mM DTT, 1 mM PMSF; the high salt concentration removes nucleic acids that are bound by the PARP-1 protein. Purified PARP-1 was eluted using elution buffer (20 mM Tris-HCl, 100 mM NaCl, 15% glycerol, 0.2 mM EDTA, 2 mM DTT, 0.2 mg/ml Flag peptide, 0.5 mg/ml recombinant human insulin), quantified using a BCA assay (Pierce), and stored in aliquots at -80°C. Flag-tagged human Sox2 was expressed in Sf9 cells and purified using the same method as described above for PARP-1.

Purification of proteins expressed in bacteria. His-tagged human PARP-1 proteins (wild-type and deletion mutants) were expressed in bacteria (BL21 cells) using a plasmid for IPTG-inducible expression⁵¹. Bacteria transformed with the expression plasmid were cultured to an OD of 0.4 and induced with 1 mM IPTG for 2 hours at 37°C. The cell pellets were collected by centrifugation and the cells were lysed using lysis buffer (10 mM Tris-HCl pH 7.5, 0.5 M NaCl, 0.5 mM EDTA, 0.1% NP-40, 10% glycerol, 10 mM imidazole; 1 mM PMSF, 2 mM DTT, 20 µg/ml leupeptin, 20 µg/ml aprotinin) with sonication. After clarification by centrifugation, the cell lysates were incubated with Ni-NTA agarose beads (Qiagen) at 4°C for 2 hours, followed by multiple washes with wash buffer (10 mM Tris-HCl pH 7.5, 1 M NaCl, 0.2% NP-40, 10% glycerol, 10 mM imidazole, 1 mM PMSF). The purified proteins were eluted using elution buffer (10 mM Tris-HCl pH 7.5, 200 mM NaCl, 0.1% NP-40, 10% glycerol, 500 mM imidazole, 1 mM PMSF, 2 mM DTT), dialysed into elution buffer without imidazole, and concentrated using a centrifugal concentrator (Millipore). The purified proteins were quantified using a BCA assay (Pierce) and stored in aliquots at -80°C.

PARP-1-Sox2 Interaction Assays

Three µg of 6xHis-tagged human PARP-1 protein purified from bacteria was incubated with 1 µg of human Sox2 protein (Abcam; Ab169843) in 1 mL of binding

buffer (20 mM HEPES pH 7.9, 150 mM NaCl, 0.2 mM EDTA, 20% glycerol, 0.1% NP-40, 0.1 ng/ μ L sonicated salmon sperm DNA, 1 mM DTT, and 1x complete protease inhibitor cocktail) for 2 h at 4°C in the presence of 30 μ l of Ni-NTA agarose beads (Qiagen). After 3 washes each with 1 mL of binding buffer, the beads were collected and boiled in 2x SDS loading dye. The eluted material was analyzed by Western blotting.

Nucleosome Assembly

Assembly of trinucleosomes. Core histones were prepared from HeLa cells as described previously⁵. To prepare a DNA template for the assembly of trinucleosomes, a pGEM1-based plasmid carrying three tandem copies of the 601 nucleosome positioning element (NPE)⁵² with the middle NPE containing a natural composite *Pou/Sox* motif from the *Nanog* gene (sequence 5'-**TTTTCATTACAATG**-3'; red, *Pou*; blue, *Sox*), was amplified in *E. coli* and purified using the PureLink HiPure plasmid filter maxiprep kit (Invitrogen). The purified plasmid DNA was digested using the *EcoRV* restriction enzyme to release the three NPE cassette with blunt ends, followed by selective precipitation of the plasmid backbone using 5% PEG and 10 mM MgCl₂. The three NPE cassette was then precipitated using 10% PEG and 10 mM MgCl₂, collected by centrifugation, and dissolved in 1x TE. To prepare a DNA template for the assembly of mononucleosomes, a DNA fragment containing the 601 NPE with or without a Sox2 binding site plus 30 bp of flanking linker DNA sequence was amplified by PCR. The Amplified DNA was subjected to phenol-chloroform extraction, followed by ethanol precipitation. Trinucleosomes and mono nucleosomes were assembled by salt gradient dialysis as described²³. The efficiency of assembly and the quality of assembled nucleosomes were checked by running aliquots of the assembled product on native 4% PAGE gels run in 0.25x TBE buffer, followed by staining with SYBR Gold (Life Technologies).

Assembly of mononucleosomes. Mononucleosomes with 5' biotinylated DNA were assembled by salt gradient dialysis as described above using 601 NPE DNA (with or without a composite *Pou/Sox* motif) that was amplified by PCR with 5' biotinylated primers (Sigma).

DNase I Footprinting Assays

Footprinting with trinucleosomes. Sox2 binding to naked DNA or the reconstituted trinucleosome templates described above was assayed by DNase I footprinting. For these assays (see Fig. 3B), 46 ng of the three NPE cassette as naked DNA or reconstituted into trinucleosomes was incubated with 25 nM of Sox2 protein (purchased from Abcam; ab169843) in the presence or absence of 10 nM of purified PARP-1 protein in a 20 μ L reaction for 1 hour at 30°C under the following buffer conditions: 15 mM Tris-HCl pH 7.5, 0.3 mM EDTA, 0.2 mM DTT, 2% glycerol, 50 mM NaCl, 150 ng/ μ L BSA, 1x protease inhibitor cocktail (Roche). The DNA was digested by the addition of 0.1 unit (for naked DNA) or 1 unit (for nucleosomal DNA) of amplification grade DNase I (Invitrogen) for 5 min. at 25°C. Digestion was stopped by addition of an equal volume of DNaseI stop solution containing 20 mM Tris-Cl pH 7.5, 50 mM EDTA, 2% SDS, 0.2 mg/ml proteinase K, 300 ng/ μ L glycogen with incubation for 1 hour at 55°C. The digested DNA was extracted with phenol:chloroform:isoamyl alcohol, precipitated with ethanol, dissolved in 1x TE, and subjected to 10 cycles of primer extension using a 32 P end-labeled primer that anneals to the linker region (5'-CCATGGAAGCTTCAGGTCACAGTGCTCGAG-3'). The resulting DNA fragments were run on an 8% PAGE-urea buffer-gradient gel in TBE. The gel was dried, exposed to a phosphoimager screen, and visualized using PharosFX system (Bio-Rad).

Footprinting with mononucleosomes. Sox2 binding to the reconstituted mononucleosome templates described above was assayed by DNase I footprinting. For these assays (see Fig. S7), mononucleosomes corresponding to 375 ng of DNA were incubated with 300 ng of Sox2 protein with or without 600 ng of His-tagged human PARP-1 in 50 μ L of binding buffer (25 mM HEPES pH 7.5, 100 mM KCl, 20% glycerol, 0.1% NP-40, 10 μ M ZnSO₄, 1 mM DTT, 1x complete protease inhibitor cocktail) for 1 hour at 30°C, followed by the addition of 5 U of DNaseI (Worthington; DPFF grade, diluted in 50 μ L of digestion buffer containing 10 mM MgCl₂ and 5 mM CaCl₂) and incubated at 25°C for 5 min. The digestion reactions were stopped by the addition of 100 μ L of DNaseI stop solution (20 mM Tris-HCl pH 7.5, 50 mM EDTA, 2% SDS, 0.2 mg/ml proteinase K, 300 ng/ μ L glycogen) with incubation for 1 hour at 55°C. The digested DNA was then extracted with phenol:chloroform:isoamyl alcohol, precipitated with ethanol, dissolved in 1x TE, and subjected to 9 cycles of primer extension using an Alexa

488 fluorescent end-labeled primer that anneals near the edge of the nucleosome (5'-CTGGAGAATCCCGGTGCCGAGGCC-3'). The resulting DNA fragments were run on an 8% PAGE denaturing buffer-gradient gel in TBE. The gel was dried, exposed to a phosphoimager screen, and visualized using PharosFX system (Bio-Rad).

Sox2 Nucleosome Binding Assays

Nucleosome binding “pulldown” assays were performed as previously described⁵³, with slight modification. Briefly, biotinylated mononucleosomes corresponding to 500 ng of DNA were immobilized on 30 µl of Dynabeads Streptavidin MyOne T1 (Invitrogen) in reconstitution buffer (10 mM Tris-HCl pH 7.5, 250 mM KCl, 1 mM EDTA, 0.1 % NP-40, 1 mM DTT) and incubated with 300 ng of Sox2 protein purchased from Abcam (Abcam; Ab169843) or purified from Sf9 cells as describe above with or without 600 ng of His-tagged human PARP-1 in 1 ml of binding buffer (20 mM HEPES pH 7.9, 150 mM NaCl, 0.2 mM EDTA, 20% glycerol, 0.1% NP-40, 1 mM DTT, and 1x complete protease inhibitor cocktail) for 2 h at room temperature. After 3 washes each with 1 ml of binding buffer, the beads were boiled in 2x SDS loading dye and the proteins were analyzed by Western blotting.

Determining DNA Shape Parameters

The 3DNA web-based tool⁵⁴ was used to determine the DNA shape parameters associated with the *Sox* motif located at different positions in the 601 nucleosome (see Fig. 4A) from high resolution X-ray crystal structures. Minor groove widths (MGWs) of three 601 nucleosome structures (PBD: 3LZ0, 3LZ1, 3MVD^{55,56}), including two different orientations (3LZ0 and 3LZ1), were measured and compared. The MGWs determined from all three nucleosome structures were similar. The data in Fig. 4A were plotted based on the 3MVD structure.

Oligonucleotide Sequences for RT-qPCR and ChIP-qPCR

The oligonucleotide primers listed below were used for RT-qPCR and ChIP-qPCR assays, as well as amplifying DNA for assembly into mononucleosomes.

• **Primers for RT-qPCR** (listed 5' to 3')

<i>Nanog</i> Forward:	CACCCACCCATGCTAGTCTT
<i>Nanog</i> Reverse:	ACCCTCAAACCTCCTGGTCCT
<i>Sox2</i> Forward:	CCGTTTTTCGTGGTCTTGTTT
<i>Sox2</i> Reverse:	TCAACCTGCATGGACATTTT
<i>Oct4</i> Forward:	CCAGAAGGGCAAAGATCAA
<i>Oct4</i> Reverse:	CTGGGAAAGGTGTCCCTGTA
<i>Parp1</i> Forward:	TGTTTGCCTCTTGTGGTGAG
<i>Parp1</i> Reverse:	AGCGTTCCTTCCTTTGGTCT
β - <i>actin</i> Forward:	CACAGGGGAGGTGATAGCAT
β - <i>actin</i> Reverse:	CTCAAGTTGGGGGACAAAAA
<i>Cdx2</i> Forward:	AAGACCGGAATTGTTTGCTG
<i>Cdx2</i> Reverse:	AAGGCTTGTTTGGCTCGTTA
<i>Krt8</i> Forward:	ACTCACTAGCCCTGGCTTCA
<i>Krt8</i> Reverse:	TCTTCACAACCACAGCCTTG
<i>Krt18</i> Forward:	CTTGCTGGAGGATGGAGAAG
<i>Krt18</i> Reverse:	CTGCACAGTTTGCATGGAGT
<i>Pax6</i> Forward:	GGCGGTTAGAAGCACTTCAC
<i>Pax6</i> Reverse:	TACGCAAAGGTCCTTGGTTC
<i>Hand1</i> Forward:	CCCTCTTCCGTCCTCTTACC
<i>Hand1</i> Reverse:	ATTCAGCAACGAATGGGAAC
<i>Wnt8a</i> Forward:	TGCCTAGTTGCAGGACAGTG
<i>Wnt8a</i> Reverse:	CTACAGGCCAACCCTGTGAT
<i>Gdf1</i> Forward:	CCATCTCCGTGCTCTTCTTC
<i>Gdf1</i> Reverse:	TCCACCACCATGTCTTCGTA
<i>Punc</i> Forward:	GGGTTCTCAAACCACAGGAA
<i>Punc</i> Reverse:	AGGGAGATGCCTTGCTTGTA
<i>Cyp26a1</i> Forward:	AGCTGGCTAGGCACTGTGAT
<i>Cyp26a1</i> Reverse:	GTGGGGCTTGTCTTCATTGT
<i>Sox1</i> Forward:	CGCGAGATGATCAGCATGTA
<i>Sox1</i> Reverse:	GTAGTGCTGTGGCAGCGAGT

Noggin Forward: TGTACGCGTGGAATGACCTA
Noggin Reverse: GTGAGGTGCACAGACTTGGA
Foxl1 Forward: CTGGGGTGGTTTCTTTTCAA
Foxl1 Reverse: GACAGGGCAGATCACCCCTAA
Brachyury Forward: GGCAACAAGGGAGGACATTA
Brachyury Reverse: GTCCACCCCCTGTCCTACTT
Map2 Forward: GAGGGACTTGGGCTCAATCT
Map2 Reverse: TGGAAAAGGCAAATTCAGG
Gata6 Forward: TGGTACAGGCGTCAAGAGTG
Gata6 Reverse: CAACACAGTCCCCGTTCTTT
Sox7 Forward: CACCCCAGACATTTCCATTC
Sox7 Reverse: TCCAAGGGCAGACAAGAGTC
Gata4 Forward: TTGTCCTCATCACCCACAGA
Gata4 Reverse: CTCTGCTACGGCCAGTAAGG
Sox17 Forward: ATCTCGACACCTGCCAAAAG
Sox17 Reverse: TGGGAAGTGGGATCAAGACT
Cdh13 Forward: CACACACAGTGGCACAACA
Cdh13 Reverse: AGAGCAGCGTGTGGATCTCT

• **Primers for ChIP-qPCR** (*listed 5' to 3'*)

Nanog_promoter Forward: CCAATGTGAAGAGCAAGCAA
 Nanog_promoter Reverse: GACCTTGCTGCCAAAGTCTC

• **Primers for amplifying 601 NPE DNA for assembly into mononucleosomes** (*listed 5' to 3'*)

601 MN Forward: CCATGGAAGCTTCAGGTCACAGTGCTCGAG
 601 MN Reverse: GGCGGCCGGATCCCGCTGTTCAATACATGC

Genomic Data Sets

The new RNA-seq and ChIP-seq data sets generated specifically for this study can be accessed from the NCBI's Gene Expression Omnibus (GEO) repository

(<http://www.ncbi.nlm.nih.gov/geo/>) using the series accession numbers GSE74111, GSE74112, as well as the individual accession numbers listed below. In addition, all of the published RNA-seq and ChIP-seq data sets from mESCs used in this study can be accessed through the GEO repository or the ENCODE Project repository (<https://www.encodeproject.org>) using the accession numbers listed below.

<u>Data Set</u>	<u>Source/Reference</u>	<u>Accession Number(s)</u>
WT RNA-seq	This study	GSM1910630/GSM1910631
<i>Parp1</i> ^{-/-} RNA-seq	This study	GSM1910632/GSM1910633
H3K4me3 ChIP-seq	This study	GSM1910636/GSM1910637
H3K27me3 ChIP-seq	This study	GSM1910634/GSM1910635
PARP-1 ChIP-seq	This study	GSM1910638/GSM1910639
WT Sox2 ChIP-seq	This study	GSM1910640/GSM1910642
<i>Parp1</i> ^{-/-} Sox2 ChIP-seq	This study	GSM1910641/GSM1910643
WT Oct4 ChIP-seq	This study	GSM1910644/GSM1910646
<i>Parp1</i> ^{-/-} Oct4 ChIP-seq	This study	GSM1910645/GSM1910647
WT ChIP-seq Input	This study	GSM1910648/GSM1910650
<i>Parp1</i> ^{-/-} ChIP-seq Input	This study	GSM1910649/GSM1910651
CTCF ChIP-seq	Chen <i>et al.</i> , 2008	GSM288351
Smc1a ChIP-seq	Kagey <i>et al.</i> , 2010	GSM560341
Smc3 ChIP-seq	Kagey <i>et al.</i> , 2010	GSM560343
Ezh2 ChIP-seq	Mikkelsen <i>et al.</i> , 2008	GSM327668
Med1 ChIP-seq	Kagey <i>et al.</i> , 2010	GSM560347
Med12 ChIP-seq	Kagey <i>et al.</i> , 2010	GSM560346
TBP ChIP-seq	Liu <i>et al.</i> , 2011	GSM774944
PolIII ChIP-seq	Liu <i>et al.</i> , 2011	GSM774946
c-Myc ChIP-seq	Chen <i>et al.</i> , 2008	GSM288356
n-Myc ChIP-seq	Chen <i>et al.</i> , 2008	GSM288357
Nanog ChIP-seq	Marson <i>et al.</i> , 2008	GSM307140/GSM307141
Tcf3 ChIP-seq	Marson <i>et al.</i> , 2008	GSM307142/GSM307143
Klf4 ChIP-seq	Chen <i>et al.</i> , 2008	GSM288354

Esrrb ChIP-seq	Chen <i>et al.</i> , 2008	GSM288355
Stat3 ChIP-seq	Chen <i>et al.</i> , 2008	GSM288353
MNase-seq	Li <i>et al.</i> , 2012	E-MTAB-1302
DNase-Seq	ENCODE/UW, 2011	ENCSR000COH

REFERENCES

1. Krishnakumar, R. & Kraus, W.L. The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. *Mol Cell* **39**, 8-24 (2010).
2. Doege, C.A. *et al.* Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* **488**, 652-5 (2012).
3. Roper, S.J. *et al.* ADP-ribosyltransferases Parp1 and Parp7 safeguard pluripotency of ES cells. *Nucleic Acids Res* **42**, 8914-27 (2014).
4. Kraus, W.L. & Hottiger, M.O. PARP-1 and gene regulation: progress and puzzles. *Mol Aspects Med* **34**, 1109-23 (2013).
5. Kim, M.Y., Mauro, S., Gevry, N., Lis, J.T. & Kraus, W.L. NAD⁺-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1. *Cell* **119**, 803-14 (2004).
6. Krishnakumar, R. *et al.* Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. *Science* **319**, 819-21 (2008).
7. Krishnakumar, R. & Kraus, W.L. PARP-1 regulates chromatin structure and transcription through a KDM5B-dependent pathway. *Mol Cell* **39**, 736-49 (2010).
8. Martin-Oliva, D. *et al.* Crosstalk between PARP-1 and NF-kappaB modulates the promotion of skin neoplasia. *Oncogene* **23**, 5275-83 (2004).
9. Imbulan-Rosenthal, C.M. *et al.* PARP-1 binds E2F-1 independently of its DNA binding and catalytic domains, and acts as a novel coactivator of E2F-1-mediated transcription during re-entry of quiescent cells into S phase. *Oncogene* **22**, 8460-71 (2003).
10. Young, R.A. Control of the embryonic stem cell state. *Cell* **144**, 940-54 (2011).
11. Zhang, S. & Cui, W. Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World J Stem Cells* **6**, 305-11 (2014).
12. Odato, M.A. *et al.* SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. *PLoS Genet* **9**, e1003288 (2013).
13. Soufi, A., Donahue, G. & Zaret, K.S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994-1004 (2012).
14. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555-68 (2015).

15. Gao, F., Kwon, S.W., Zhao, Y. & Jin, Y. PARP1 poly(ADP-ribosyl)ates Sox2 to control Sox2 protein levels and FGF4 expression during embryonic stem cell differentiation. *J Biol Chem* **284**, 22263-73 (2009).
16. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
17. McLean, C.Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501 (2010).
18. Franco, H.L., Nagari, A. & Kraus, W.L. TNFalpha signaling exposes latent estrogen receptor binding sites to alter the breast cancer cell transcriptome. *Mol Cell* **58**, 21-34 (2015).
19. Siersbaek, R. *et al.* Transcription factor cooperativity in early adipogenic hotspots and super-enhancers. *Cell Rep* **7**, 1443-55 (2014).
20. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106-17 (2008).
21. Remenyi, A. *et al.* Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev* **17**, 2048-59 (2003).
22. Hayes, J.J. & Wolffe, A.P. Histones H2A/H2B inhibit the interaction of transcription factor IIIA with the *Xenopus borealis* somatic 5S RNA gene in a nucleosome. *Proc Natl Acad Sci U S A* **89**, 1229-33 (1992).
23. Utley, R.T., Cote, J., Owen-Hughes, T. & Workman, J.L. SWI/SNF stimulates the formation of disparate activator-nucleosome complexes but is partially redundant with cooperative binding. *J Biol Chem* **272**, 12642-9 (1997).
24. Clark, N.J., Kramer, M., Muthurajan, U.M. & Luger, K. Alternative modes of binding of poly(ADP-ribose) polymerase 1 to free DNA and nucleosomes. *J Biol Chem* **287**, 32430-9 (2012).
25. Rohs, R. *et al.* The role of DNA shape in protein-DNA recognition. *Nature* **461**, 1248-53 (2009).
26. Abe, N. *et al.* Deconvolving the recognition of DNA shape from sequence. *Cell* **161**, 307-18 (2015).
27. Cui, F. & Zhurkin, V.B. Structure-based analysis of DNA sequence patterns guiding nucleosome positioning in vitro. *J Biomol Struct Dyn* **27**, 821-41 (2010).

28. Petesch, S.J. & Lis, J.T. Overcoming the nucleosome barrier during transcript elongation. *Trends Genet* **28**, 285-94 (2012).
29. Barozzi, I. *et al.* Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Mol Cell* **54**, 844-57 (2014).
30. Tang, L., Nogales, E. & Ciferri, C. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. *Prog Biophys Mol Biol* **102**, 122-8 (2010).
31. Zaret, K.S. & Carroll, J.S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25**, 2227-41 (2011).
32. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251-60 (1997).
33. Lai, Y.S. *et al.* SRY (sex determining region Y)-box2 (Sox2)/poly ADP-ribose polymerase 1 (Parp1) complexes regulate pluripotency. *Proc Natl Acad Sci U S A* **109**, 3772-7 (2012).
34. Hah, N. *et al.* A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* **145**, 622-34 (2011).
35. Zhong, S. *et al.* High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harb Protoc* **2011**, 940-9 (2011).
36. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11 (2009).
37. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-5 (2010).
38. Feng, X., Grossman, R. & Stein, L. PeakRanger: a cloud-enabled peak caller for ChIP-seq data. *BMC Bioinformatics* **12**, 139 (2011).
39. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-6 (2011).
40. Saldanha, A.J. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**, 3246-8 (2004).
41. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).

42. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40 (2010).
43. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).
44. Mootha, V.K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-73 (2003).
45. Bailey, T.L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-8 (2009).
46. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**, 14863-8 (1998).
47. Li, Z. *et al.* Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. *Cell* **151**, 1608-16 (2012).
48. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
49. Chen, K. *et al.* DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res* **23**, 341-51 (2013).
50. Cui, F. & Zhurkin, V.B. Rotational positioning of nucleosomes facilitates selective binding of p53 to response elements associated with cell cycle arrest. *Nucleic Acids Res* **42**, 836-47 (2014).
51. Wacker, D.A. *et al.* The DNA binding and catalytic domains of poly(ADP-ribose) polymerase 1 cooperate in the regulation of chromatin structure and transcription. *Mol Cell Biol* **27**, 7475-85 (2007).
52. Lowary, P.T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* **276**, 19-42 (1998).
53. Bartke, T. *et al.* Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470-84 (2010).

54. Lu, X.J. & Olson, W.K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* **31**, 5108-21 (2003).
55. Vasudevan, D., Chua, E.Y. & Davey, C.A. Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J Mol Biol* **403**, 1-10 (2010).
56. Makde, R.D., England, J.R., Yennawar, H.P. & Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467**, 562-6 (2010)

CHAPTER 3

Regulation of Triple Negative Breast Cancer by Sox2 and PARP-1

3.1. Summary

Triple-negative breast cancer (TNBC) is a breast cancer subtype that imposes challenges for treatment. Patients with TNBC frequently have poor prognoses and outcomes. Therefore, identifying novel genes and pathways that cause that progression of TNBC will be of great help for designing therapies for treatment. Various studies have suggested that Sox2 might be responsible for mediating the progression and metastasis of TNBC. In this study, we investigated the function of Sox2 in regulating gene expression program in triple negative breast cancer cells. We found that Sox2 regulates the transcription of genes mediating cell migration by occupying distal regulatory regions of these genes. In addition, we explored the potential regulatory effect of Sox2 genomic localization by inhibiting PARylation activity in TNBC. Surprisingly, blocking PAR activity alters Sox2 genomic binding pattern as well as Sox2 mediated gene transcription program. Therefore, we have uncovered a functional link between PARP, PARylation activity and Sox2 in triple negative breast cancer cells.

3.2. Introduction

Triple-negative breast cancer (TNBC) is a subtype that imposes challenges for treatment, in part due to its lack of chemotherapeutic target proteins such as ER, PR, HER2¹. Patients with TNBC frequently have poor prognoses and outcomes. Therefore, identifying novel genes and pathways that cause that progression of TNBC will be of great help for designing therapies for treatment². TNBC tumors have been known to be highly heterogeneous. Various different TNBC sub-types with distinct gene expression profiles have been identified by analyzing gene transcription patterns using datasets from patient tumor samples³. In addition, TNBC tumors are composed of multiple different cell subpopulations. Among these subpopulations, a small population of stem-like cells, or cancer stem cells (CSC), have been proposed to be responsible for mediating metastasis as well as resistance to cancer therapy⁴.

The function of Sox2 in mediating tumorigenesis as well as cancer progression has been reported in various studies⁵. Sox2 controls tumor initiation as well as cancer stem-cell functions in squamous-cell carcinoma⁶. In addition, Sox2 has been found to be involved in multiple different cancers including small-cell lung cancer⁷, pancreatic cancer⁸, prostate cancer⁹, glioblastoma¹⁰, as well as breast cancer^{11,12}. Sox2 gene is either amplified or transcriptionally up-regulated in 9% of breast cancer patients in The Cancer Genome Atlas (TCGA) database^{13 14}. Its expression level greatly elevates in subpopulations with stem cell-like features, and significantly correlates with triple negative breast cancer as determined by immunohistochemistry in multiple breast cancer patient samples¹⁵. Moreover, Sox2 was identified to be over-expressed in lymph node-metastased breast cancer tumors as compared to primary tumors using expression microarray¹⁶. Therefore, Sox2 is a potential regulator of triple negative breast cancer. Sox2 expresses in a heterogeneous manner in TNBC. It was found to be highly expressed in a sub-population of cells containing surface marks CD44⁺/CD24^{neg/low} in triple negative breast cancer cell line as well as dissociated primary breast cancer cultures. These cells, compared to CD44⁺/CD24⁻ cells, are more tumorigenic and show greater motility. Consequently, a small population of Sox2⁺ cells might act as “cancer stem cells” and mediate tumor initiation as well as metastasis¹⁷.

As a transcription factor, Sox2 influences cell phenotypes through altering gene expression profile. It regulates multiple cellular pathways which in turn control cell growth, survival, differentiation, or migration⁵. As an example, Sox2 has been shown to inhibit the expression of multiple genes in the Wnt signaling pathway, which is required for the transition from self-renewing stem cell state into differentiated state in the osteoblast lineage¹⁸. In addition, Sox2 was also shown to promote cell proliferation through enhancing cyclinD3 (CCND3) expression in pancreatic cancer cells⁸. Furthermore, Sox2 is implicated as playing a role in regulating cell migration and invasion in melanoma, through regulating the expression of matrix metalloproteinase (MMP)-3¹⁹. In breast cancer, on the other hand, ectopic Sox2 expression was shown to promote the expression of β -catenin and enhance Wnt signals, which in turn increases cell migration¹¹. Despite the multifaceted roles of Sox2, it seems that Sox2 functions in a cell type-specific manner. For example, although Sox2 promotes cell growth in pancreatic cancer, it inhibits cell proliferation in gastric cancers²⁰. In addition, how Sox2 regulates certain signaling pathway such as Wnt signals is also dependent on specific cancer cell types. Therefore, it is important to determine the context-specific role of Sox2 in order to understand its function in a given cancer type.

The cell type-specific manner of Sox2 function is in part due to the distinct chromatin binding patterns of Sox2 in different cell types, which is further determined by the expression of cell type-specific transcription factors which cooperate with Sox2 binding to the genome²¹. Consequently, identifying breast cancer-specific transcription factor responsible for mediating Sox2 transcriptional regulation activity will greatly help to understand the specific function of Sox2 in triple negative breast cancer.

Due to the fact that only a small population of cells are Sox2 positive in many triple negative breast cancer cell lines¹⁷, it is difficult to directly study the function as well as to identify regulatory targets of Sox2. In order to overcome this difficulty and dissect the role of Sox2 in triple negative breast cancer, we ectopically expressed Sox2 in TNBC cell lines including MDA-MB-231 and HCC1143¹¹. We found that Sox2-expressing TNBC cells proliferate slower, and are more invasive. Gene transcription profile regulated by Sox2 is consistent with the cellular phenotype caused by Sox2 expression. ChIP-seq was also performed to analyze the genome-wide binding pattern of Sox2. Finally, we explored the

potential effect of inhibiting PARylation activity on Sox2 chromatin binding as well as Sox2-mediated gene transcription regulation.

3.3. Results

3.3.1. Sox2-expressing TNBC cells proliferate slower and are more invasive.

To explore the potential function of Sox2 in regulating breast cancer, we first examined the correlation between Sox2 expression level and distal metastatic free survival (DMFS) rate in breast cancer patients using the online-based Kaplan-Meier plotter (<http://kmplot.com/analysis/>)²². Interestingly, high expression of Sox2 correlates with poor prognosis outcome only in triple negative breast cancer patients, but not in ER/PR positive breast cancer patients (Fig 1A). The TNBC-specific correlation between Sox2 expression level and prognosis outcome led us to hypothesize that Sox2 may play a role in regulating tumor progression in triple negative breast cancer. To test this hypothesis, we created a doxycycline (Dox)-inducible Sox2 expression cell line in human triple negative breast cancer cells MDA-MB-231 as well as HCC1143. Sox2 protein level is undetectable in both cell lines without doxycycline treatment. Upon adding Dox to cell culture media, Sox2 expression is robustly induced (Fig 3.1B, Fig 2A). We first used these cell lines to examine the effect of ectopic Sox2 expression on cell proliferation. Sox2-expressing MDA-MB-231 and HCC1143 cells grow more slowly as compared to control cells (Fig 3.1C, Fig 2B). We then examined the cell migration using Sox2-expressing MDA-MB-231 cells. Consistent with previous report¹¹, ectopic Sox2 expression drives more cell migration in wound healing assay (Fig 3.1D). This is consistent with the clinical data in triple negative breast cancer patients, where higher Sox2 level correlates with lower distal metastatic free survival (DMSF) rate (Fig 3.1A).

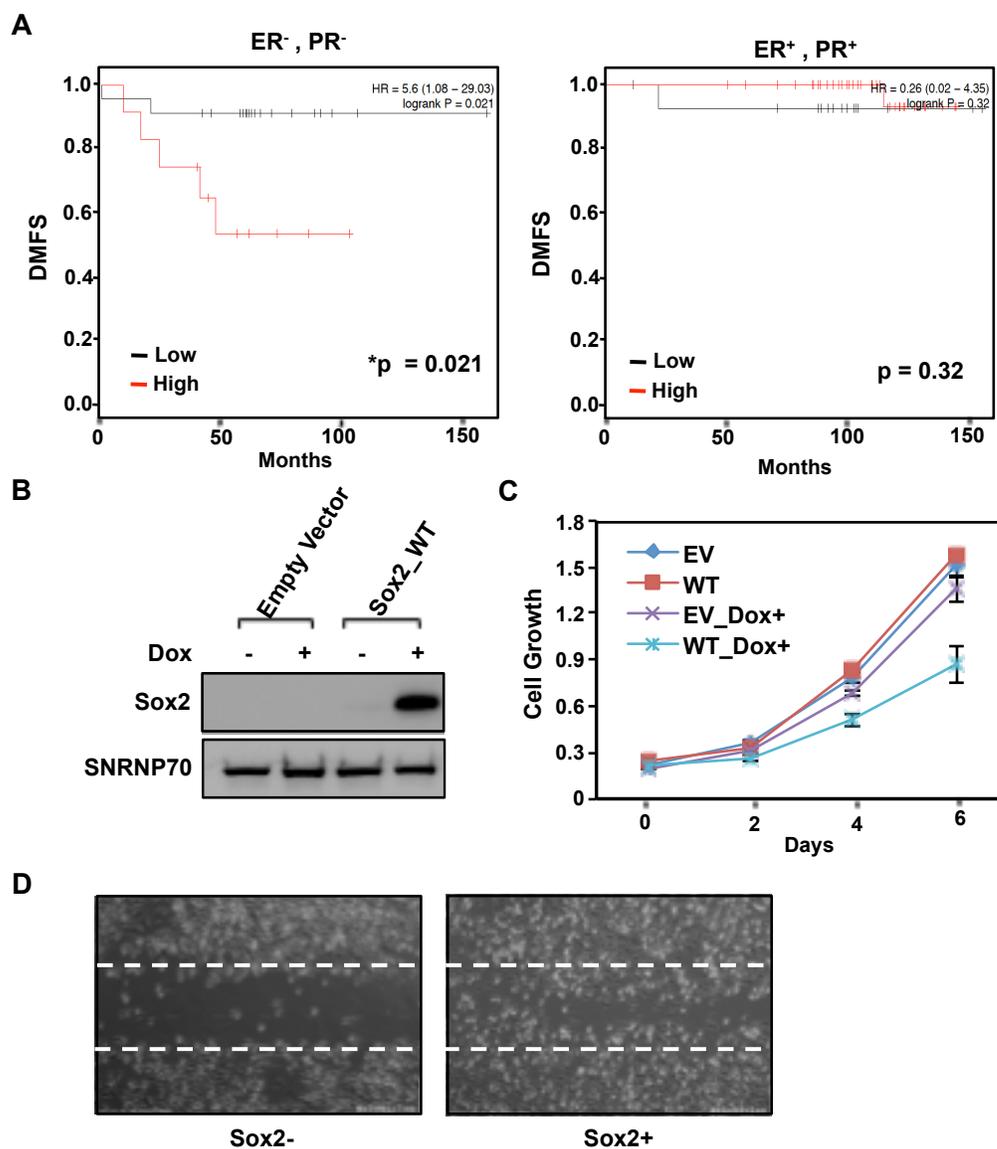


Figure 3.1. Sox2 promotes metastasis in triple negative breast cancer.

A) Kaplan-Meier survival analyses of the relation between Sox2 expression level and distal metastasis free survival (DMFS) rate in different breast cancer subtypes. The breast cancer outcome-related gene expression data was accessed using online tool Kaplan-Meier Plotter²².

B) Western Blotting analysis of Dox-induced Sox2 expression in MDA-MB-231 cells. SNRNP70 was used as an internal loading control.

C) MDA-MB-231 cell proliferation assay examining the effect of ectopic Sox2 expression on cell growth. Each data point represents the mean \pm SEM, n = 3. 250 ng/mL Dox was added to induce Sox2 expression. The difference in cell growth between Sox2-expressing cells and control cells are significant (t-test, p < 0.01).

D) Cell migration assay in MDA-MB-231 cells.

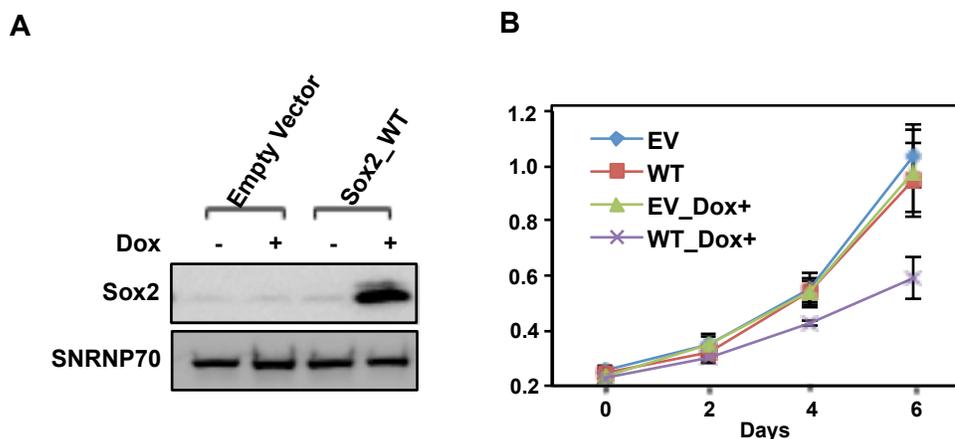


Figure 3.2. Ectopic Sox2 expression inhibits cell growth in HCC1143 cells.

A) Western Blotting analysis of Dox-induced Sox2 expression in HCC1143 cells. SNRNP70 was used as an internal loading control.

B) HCC1143 cell proliferation assay examining the effect of ectopic Sox2 expression on cell growth. Each data point represents the mean \pm SEM, $n = 3$. 250 ng/mL Dox was added to induce Sox2 expression. The difference in cell growth between Sox2-expressing cells and control cells are significant (t-test, $p < 0.01$).

3.3.2. Ectopic Sox2 expression alters gene transcription through chromatin association in MDA-MB-231 cells.

We hypothesize that the phenotype observed in Sox2-expressing cells is due to altered gene transcription profile regulated by Sox2. To test this hypothesis, we performed RNA-seq to identify genes that are regulated upon the induction of Sox2 expression. Following 72 hours of Dox-induced Sox2 expression, a total of 4,975 genes were significantly regulated in Sox2-expressing cells as compared to empty vector control cells ($FDR < 5\%$, $p\text{-val} < 0.01$), among which 2,302 genes were up-regulated, while 2,673 genes down-regulated (Fig 3.3A). We examined the identity of Sox2-regulated genes by performing gene ontology analysis^{23 24}. Consistent with cellular phenotype, the top gene ontology terms for up-regulated genes include genes involved in cell migration and cell motion pathways (Fig 3.3A), such as *FNI* and *MMP1*. On the other hand, down-regulated genes are enriched with genes regulating cell cycle and mitosis pathways (Fig 3.3A).

Genes that are responsible for mediating lung or brain metastasis in triple negative breast cancer have been identified by comparing gene expression between metastasized cancer cells and parental cells within MDA-MB-231 cells injected into immunodeficient

mice^{25 26}. To determine whether Sox2 regulates genes that mediate tumor metastasis, we performed gene set enrichment analysis (GSEA) using previously defined lung or brain metastasized tumor-enriched genes as input gene list²⁷. These genes are significantly up-regulated upon ectopic Sox2 expression in GSEA analysis, with a normalized enrichment score (NES) of 1.18, and p-value <0.001 (Fig 3.3B). Therefore, Sox2 causes an up-regulation of genes involved in cell migration and metastasis. This is consistent with observed cellular phenotype that cell migration increases upon Sox2 expression, as well as with the Kaplan-Meier survival analysis that a higher Sox2 expression level correlates with poor prognosis outcome in TNBC patients.

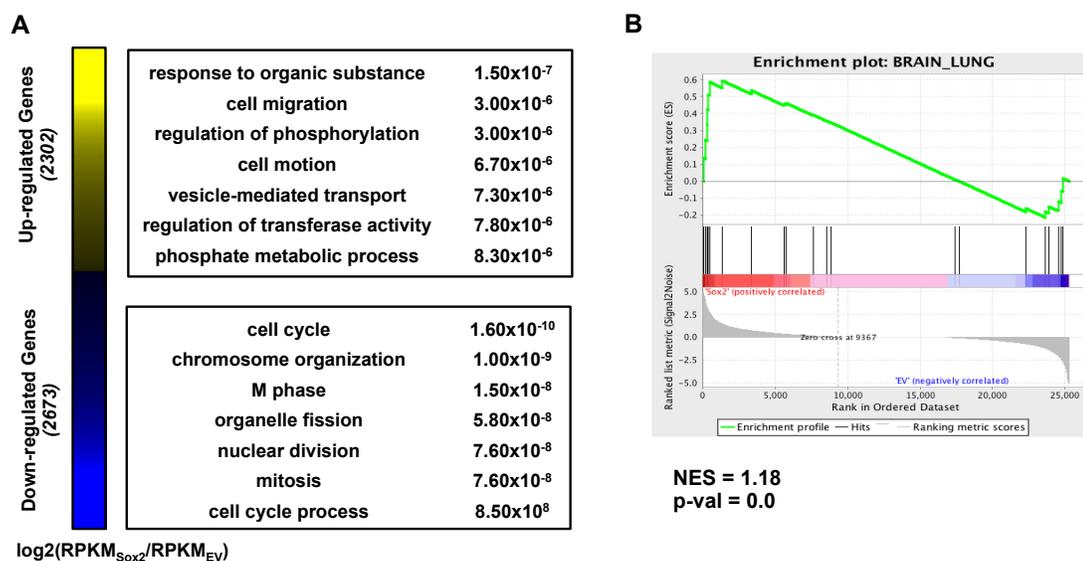


Figure 3.3. Ectopic Sox2 expression alters global gene expression in MDA-MB-231 cells.

A) (Left) Effect of Sox2 expression on gene expression in MDA-MB-231 cells as determined by RNA-seq. The heatmap shows the relative expression levels of genes whose expression significantly (FDR < 5%) increases upon Sox2 expression (Up-regulated genes) or decreases upon Sox2 expression (Down-regulated genes). The data are represented as log₂(Sox2 RPKM/EV RPKM). (Right) Top terms of gene ontology analysis for up-regulated and down-regulated genes upon Sox2 expression.

B) Gene Set Enrichment Analysis (GSEA) using genes that are enriched in brain or lung-metastasized MDA-MB-231 cells as input gene list. Genes were ranked based on fold change between Sox2 expression and empty vector control as determined by RNA-seq. Brain or lung metastasis-enriched genes are up-regulated upon Sox2 expression.

3.3.3. Sox2 regulates gene transcription by direct association with distal regulatory elements

To explore the mechanism of gene transcription regulation by Sox2, we performed Sox2 ChIP-seq following 72 hours of Dox-induced Sox2 expression in MDA-MB-231 cells. A total of 12,788 high confident Sox2 binding sites were identified ($p < 10^{-4}$). Similar to Sox2 binding profile in other cell types like embryonic stem cell (ESC) and neural progenitor cells (NPC)²¹, a small fraction of Sox2 bind to gene promoters, while majority of Sox2 bind to either gene body or intergenic regions (Fig 3.5A, Fig 3.4A). Canonical Sox2 binding DNA sequence is enriched in Sox2 binding sites (Fig 3.5B), indicating the Sox2 peaks identified in our ChIP-seq are biologically relevant. We then explored the chromatin states associated with Sox2 binding sites in MDA-MB-231 cells using chromHMM to annotate the genomic regions occupied by Sox2²⁸. ChIP-seq data sets for histone marks H3K27me3, H3K4me3, H3K27ac, H3K4me1, H3K36me3 were included in the annotation. Interestingly, although Sox2 preferably occupy active enhancer regions with high level of H3K27ac and H3K4me1 marks, majority of Sox2 binding sites are located in basal chromatin regions where active histone marks are lacking (Fig 3.5C). This is consistent with the previous report that during early stage of somatic reprogramming, pluripotency factors including Sox2 bind to closed chromatin regions²⁹.

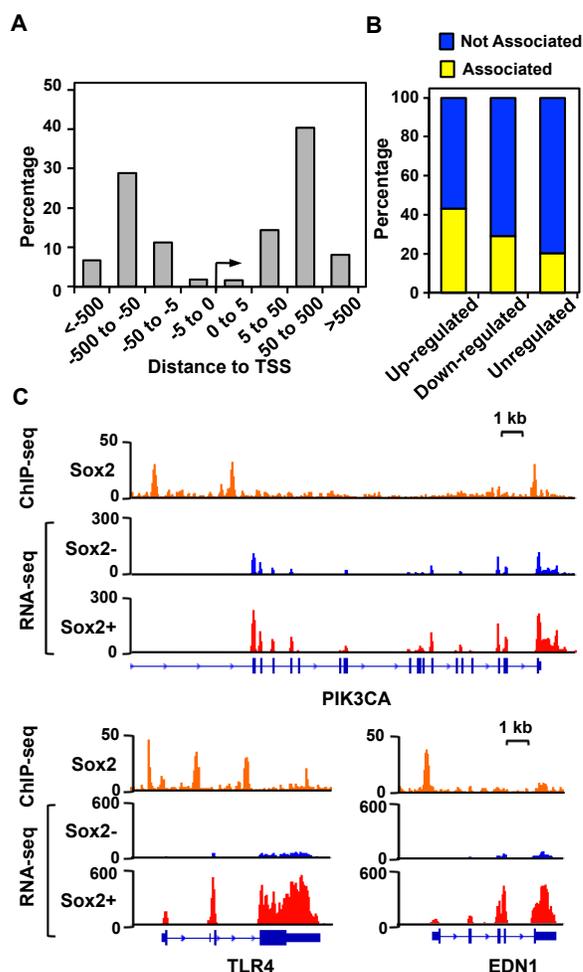


Figure 3.4. Sox2 directly binds to distal regions and regulates gene expression.

A) Distribution of Sox2 peaks relative to the TSSs of all RefSeq genes determined using the web-based bioinformatics tool GREAT³⁰.

B) Fraction of Sox2-regulated genes occupied by Sox2 within a 100 kb (\pm 50 kb) region surrounding TSSs. Up-regulated = genes whose expression significantly increases upon Sox2 expression; Down-regulated = genes whose expression significantly decreases upon Sox2 expression; Unregulated = genes whose expression does not significantly change upon Sox2 expression. *Yellow*: Genes with Sox2 binding. *Blue*: genes without Sox2 binding.

C) Genome browser tracks of RNA-seq and Sox2 ChIP-seq data showing examples of genes up-regulated upon Sox2 expression and are occupied by Sox2.

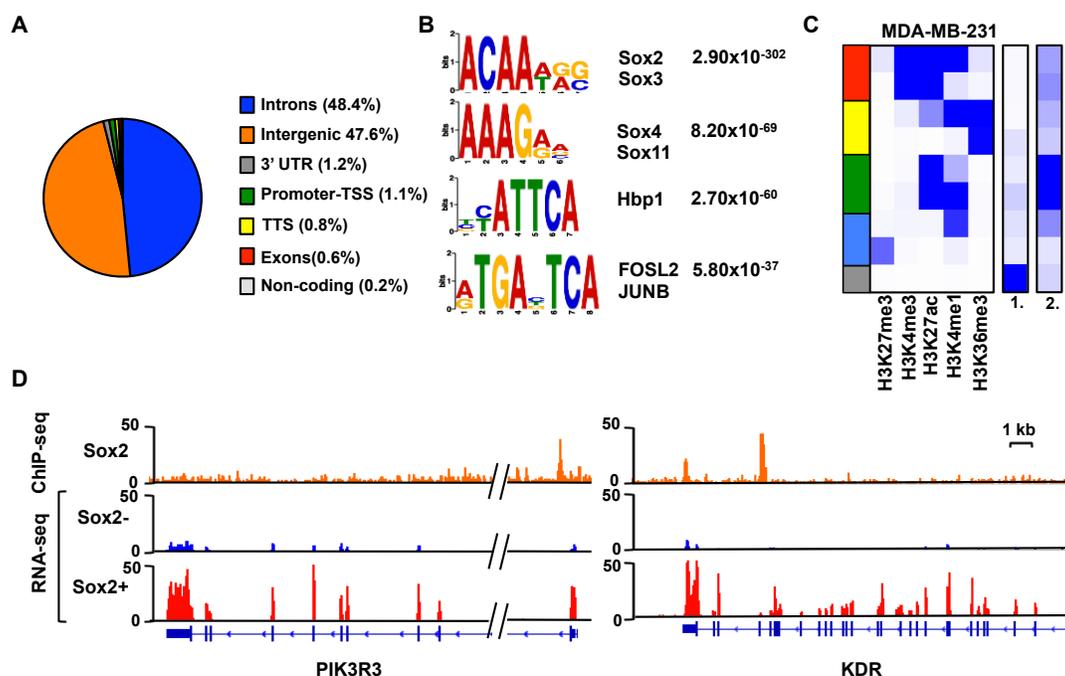


Figure 3.5. Characterization of chromatin features of Sox2 peaks in MDA-MB-231 cells.

A) Distribution of Sox2 peaks across genomic features in MDA-MB-231 cells from Sox2 ChIP-seq data.

B) Motif Analysis using MEME suite showing the enrichment of Sox2 motifs in Sox2 peaks in MDA-MB-231 cells.

C) Different chromatin states as determined by ChromHMM using histone marks in MDA-MB-231 cell. (1) Number of Sox2 binding sites associated with respective chromatin state. (2) Number of Sox2 binding sites associated with each chromatin state normalized to fraction of the respective chromatin state. Majority of Sox2 binding sites localize in inactive chromatin region in MDA-MB-231 cell.

D) Genome browser tracks of RNA-seq and Sox2 ChIP-seq data showing examples of genes up-regulated upon Sox2 expression and are occupied by Sox2.

To find direct Sox2 target genes whose expression is affected in response to Sox2 expression, we compared ChIP-seq and RNA-seq data sets in MDA-MB-231 cells. To do this, we searched for Sox2 binding within a 100 kb region (± 50 kb) of gene transcription start sites (TSSs) for genes whose expression is either up-regulate, down-regulated, or unregulated upon ectopic Sox2 expression. We then calculated the fraction of genes that are associated Sox2 binding, defined by ≥ 1 Sox2 peaks found within 100 kb TSS region. Compared to unregulated genes, a majority of up-regulated genes are associated with Sox2, while a smaller fraction of down-regulated genes are associated with Sox2 binding (Fig

3.4B-C, Fig 3.5D). Therefore, a large fraction of genes regulated by Sox2 can be explained by Sox2 occupancy. In addition, it seems that up regulation of gene expression is more likely a direct consequence of Sox2 binding.

3.3.4. Sox2 binds to a distinct cohort of genomic locations in triple negative breast cancer compared to embryonic stem cells

The repertoire of genes regulated by Sox2 in triple negative breast cancer cells is completely different from Sox2 regulatory targets in embryonic stem cells. We hypothesize that this is due to cell type specific genomic binding of Sox2. Genomic binding of Sox2 has been known to be highly cell type specific, which is due to association of Sox2 with distinct transcription factors expressed in a cell type specific manner. This phenomenon was demonstrated by comparison of Sox2 ChIP-seq data in embryonic stem cell and neural precursor cells²¹. We thus compared Sox2 binding profile between MDA-MB-231 cells and human embryonic stem cells³¹. Consistent with our hypothesis, a minimum number of overlapped Sox2 binding sites in both cell lines were observed. Out of a total of 16,852 Sox2 peaks that is observed in either hESC or MDA-MB-231 cells, only 198 sites were observed in both cell types. The rest were either ESC-specific ($n = 4,064$) or MDA-MB-231 cell specific ($n = 12,590$) (Fig 3.6A-B). Consequently, Sox2 binds to a distinct cohort of genomic locations in triple negative breast cancer cells and regulates breast cancer-specific gene expression program.

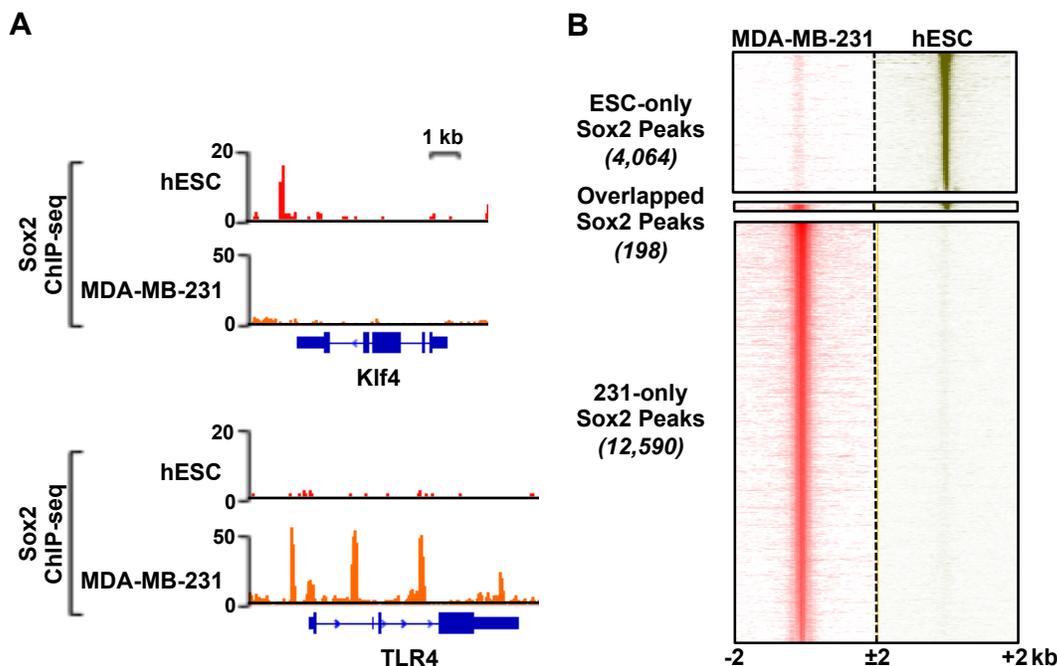


Figure 3.6. Sox2 binds to a distinct cohort of genomic locations in triple negative breast cancer compared to embryonic stem cells.

A) Genome browser tracks of Sox2 ChIP-seq in MDA-MB-231 cells and human embryonic stem cells showing cell type-specific Sox2 occupancy.

B) Heatmap showing cell type-specific Sox2 binding in MDA-MB-231 cells and hESC. ESC-only Sox2 peaks: Sox2 binding sites that is only present in hESC but not in MDA-MB-231 cells. Peaks are ordered top to bottom by signal intensity. Overlapped Sox2 peaks: Sox2 binding sites that is present in both hESC and MDA-MB-231 cells. 231-only Sox2 peaks: Sox2 binding sites that is only present in MDA-MB-231 cells but not in hESC.

3.3.5. Regulation of Sox2 chromatin association by PARP-1

Our previous study has identified a PARP-1-Sox2 regulatory pathway in mouse embryonic stem cells. PARP-1 is required to stabilize Sox2 binding to closed chromatin regions. In addition, PARP-1 also enhances Sox2 binding to nucleosome in biochemical assay. In TNBC, majority of Sox2 binding sites are in closed chromatin regions (Fig 3.5C). Therefore, we asked whether PARP-1 is also required to stabilize chromatin association of Sox2 in MDA-MB-231 cells. To answer this question, we used shRNA to knockdown PARP-1 in MDA-MB-231 cells. We then isolated different cellular fractions including nuclei plasma and chromatin, and analyzed the level of Sox2 associated with each fraction.

Although total level of Sox2 is unchanged upon PARP-1 knockdown, PARP-1 depletion significantly decreased the amount of Sox2 associated with chromatin fraction, while does not change the level of Sox2 in nuclei plasma fraction (Fig 3.7). Therefore, similar to embryonic stem cells, PARP-1 is required for stabilizing the chromatin association of Sox2 in triple negative breast cancer cells.

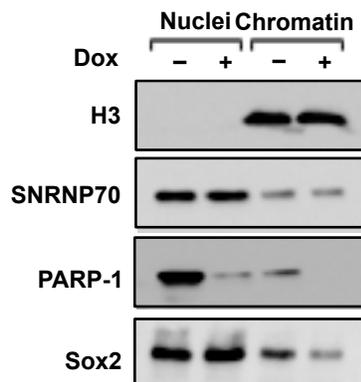


Figure 3.7. Sox2 association with chromatin requires PARP-1.

Western blotting analysis of Sox2 in different cell fractions upon PARP-1 knockdown. Sox2 in the chromatin fraction decreases with PARP-1 depletion.

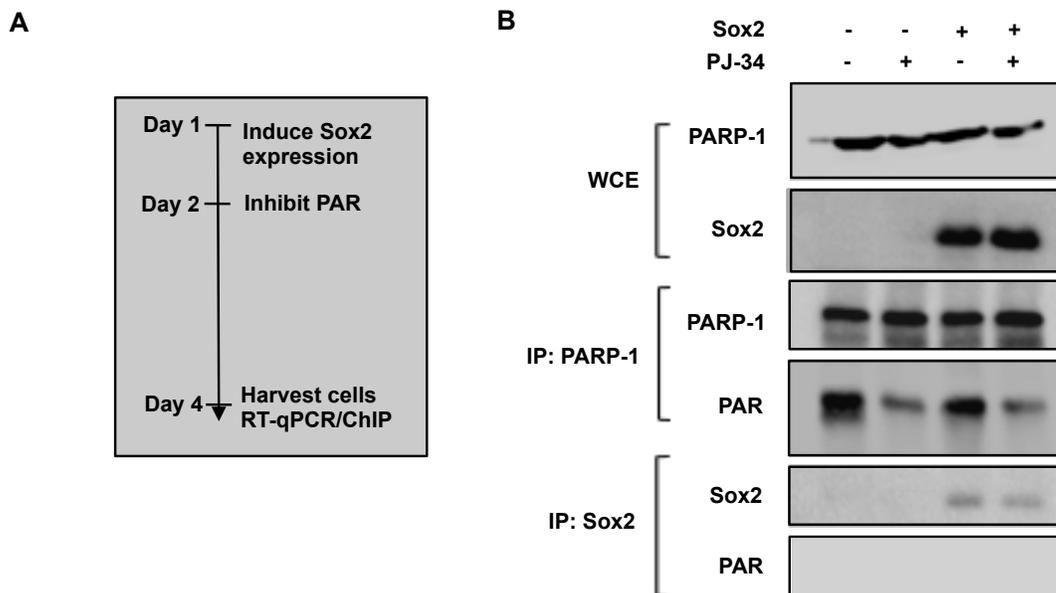


Figure 3.8. Inhibition of PARP-1 activity using PJ34.

A) Time course of inducing Sox2 expression and PJ34 treatment before cells were collected for experiments.

B) Western blotting analysis of PARylation level and PARP-1, Sox2 protein levels upon PJ34 treatment. PARP-1 and Sox2 protein were pulled down separately using PARP-1 or Sox2 antibody before loaded onto SDS-PAGE gel and analyze protein as well as PARylation level.

3.3.6. Blocking PARylation activity reshapes Sox2 genomic localization profile

Various clinical trials have been undergoing using a combination of PARP inhibitor and chemotherapy to treat triple negative breast cancer^{32 33}. Multiple studies have indicated a beneficial effect in inhibiting PARP enzymatic activity to TNBC patients³⁴. The interplay between PARP-1 and Sox2 lead us to question whether PAR inhibition can also have a regulatory effect on chromatin association of Sox2 as well as Sox2 mediated transcription regulation. To answer this question, we used PARylation inhibitor PJ-34 to treat MDA-MB-231 cells with or without Sox2 expression (Fig 3.8A-B). Blocking PARP-1 enzymatic activity does not change Sox2 expression level (Fig 3.8B). In addition, we did not detect PARylation signal in pulled down Sox2 protein (Fig 3.8B), indicating that in contrary to previously reported Sox2 PARylated by PARP-1 in embryonic stem cells³⁵, Sox2 is not modified by PARP-1 in MDA-MB-231 cells. We then examined whether PARylation activity inhibition alters chromatin association of Sox2 by analyzing the protein level of Sox2 in different cellular fractions. Strikingly, Sox2 level in the chromatin fraction dramatically increased upon PJ34 treatment, indicating chromatin association of Sox2 is enhanced (Fig 3.9A). We further performed Sox2 ChIP-seq with PJ34 treatment in MDA-MB-231 cells to analyze genomic localization profile of Sox2 upon PAR inhibition. Consistent with cellular fractionation results, PJ34 treatment causes Sox2 to gain access to almost twice as many binding sites compared to (12,083 gained peaks), compared to the original 12,783 Sox2 peaks without PAR inhibition.

In embryonic stem cells, PARP-1 specifically regulates chromatin association of Sox2 in closed chromatin regions. We wonder what chromatin features are associated with the newly gained Sox2 sites upon blocking PARylation activity. To answer this question, we compared FAIREseq signals between newly gained Sox2 binding sites and maintained Sox2 peaks in MDA-MB-231 cells³⁶. Interestingly, compared to maintained peaks, newly gained Sox2 binding sites are associated with lower FAIREseq signal, indicating that these Sox2 peaks are associated with closed chromatin regions (Fig 3.10A-B). Therefore,

blocking PAR activity surprisingly redirects Sox2 genomic localization, exposing tremendous numbers of novel Sox2 binding sites in the genome with closed chromatin conformation.

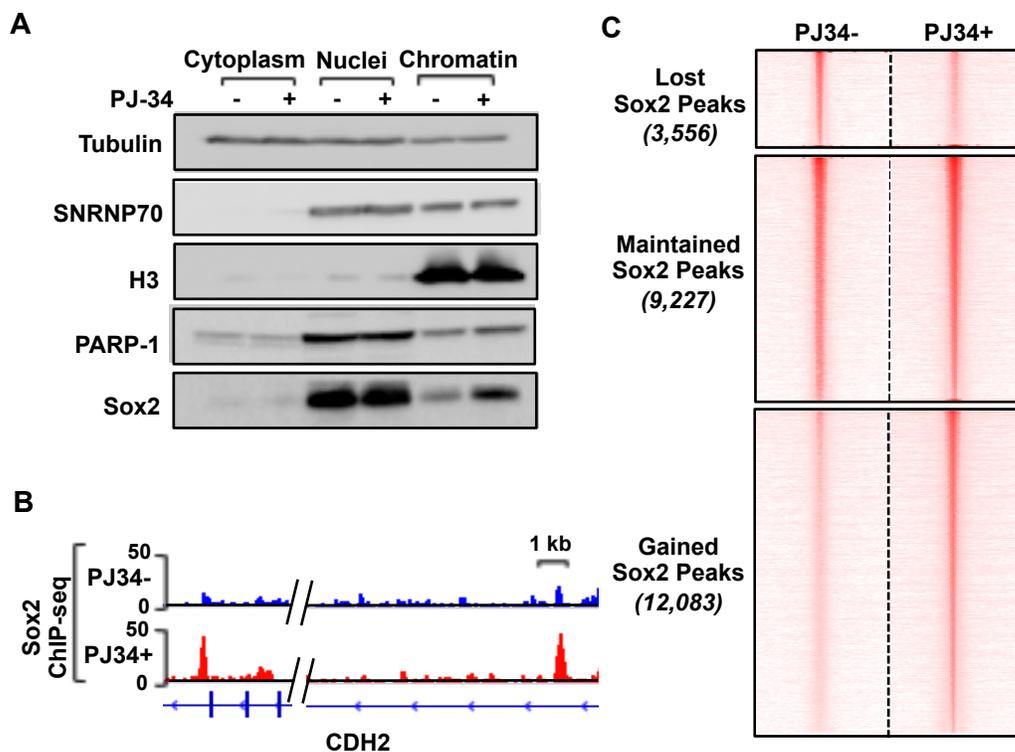


Figure 3.9. Blocking PARP-1 activity promotes chromatin association of Sox2.

A) Western blotting analysis of Sox2 in different cell fractions upon PAR inhibition. Sox2 in the chromatin fraction increases with blocking PARylation activity.

B) Genome browser tracks of Sox2 ChIP-seq data in MDA-MB-231 cells showing increased Sox2 binding upon PARylation inhibition.

C) Heatmap of high confidence Sox2 peaks in MDA-MB-231 cells from ChIP-seq data (p -value $< 10^{-4}$) centered on the Sox2 binding sites (± 2 kb) and segregated based on their change upon PARylation activity inhibition. Peaks are ordered top to bottom by signal intensity.

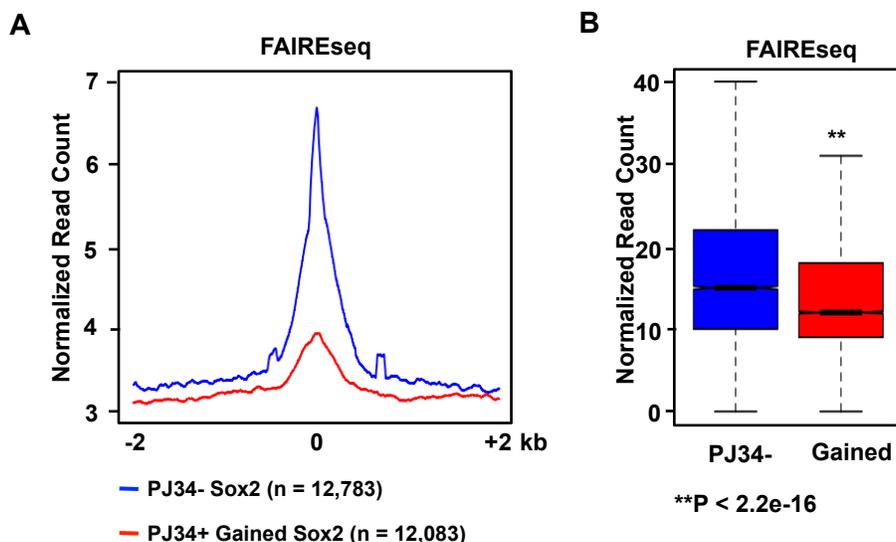


Figure 3.10. Newly gained Sox2 binding upon PAR inhibition are associated with closed chromatin.

A) Average FAIRE-seq signals surrounding newly gained (*red*) and maintained (*blue*) Sox2 binding sites upon PJ34 treatment in MDA-MB-231 cells. The data are centered on the Sox2 binding sites determined by ChIP-seq (± 2 kb).

B) Box plot showing the normalized FAIRE-seq signals surrounding newly gained (*red*) and maintained (*blue*) Sox2 binding sites upon PJ34 treatment in MDA-MB-231 cells in 500 bp window. Asterisks indicate significant differences: Student's t-test, p-value < 2.2×10^{-16} .

3.3.7. Modulation of Sox2-mediated gene transcription regulation through PAR inhibition.

The massive gained chromatin association of Sox2 upon PARylation inhibition led us to ask whether Sox2-mediated gene transcription profile is also reshaped by PJ34 treatment. We tackled this question by performing RNA-seq in MDA-MB-231 cells combining Sox2 expression and PJ34 treatment. Notably, the transcription profile regulated by Sox2 expression is extensively altered by PJ34 treatment, leading to the up- or down-regulation of a new set of genes whose expression were not changed by either Sox2 expression or PJ34 treatment alone, but significantly regulated upon the combination of the two conditions (1,708 newly regulated genes) (Fig 3.11 A-B). Importantly, this gained gene transcription regulation correlates with gained Sox2 occupancy upon PJ34 treatment, with more than 40% newly regulated genes associated with a newly gained

Sox2 binding within 100kb around their TSSs. Therefore, blocking PARylation activity caused Sox2 to access to new binding sites and regulate gene transcription profile.

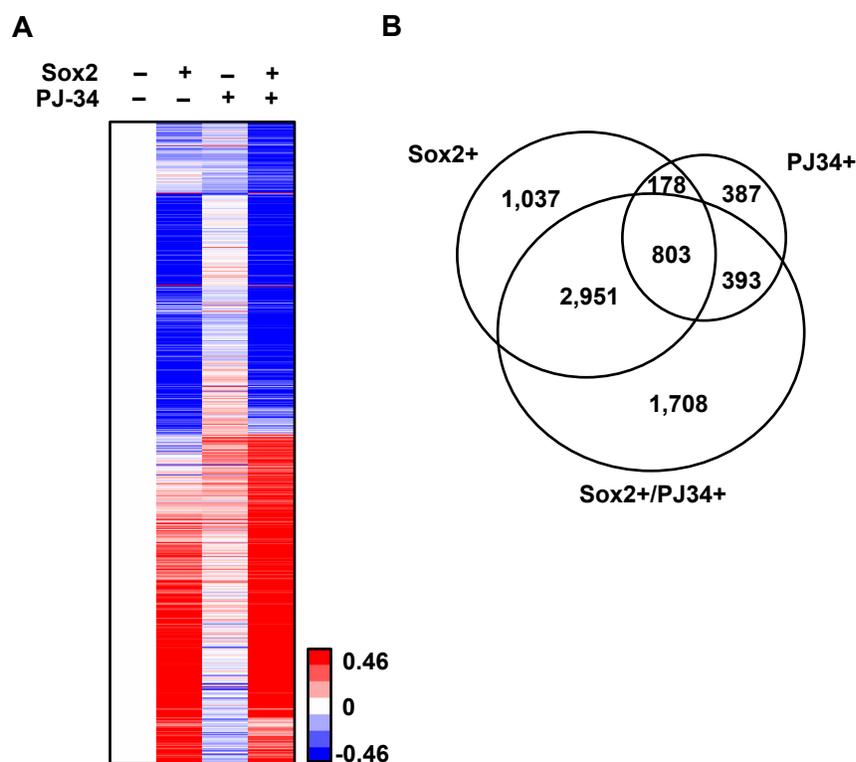


Figure 3.11. Modulation of Sox2-mediated gene transcription regulation through PAR inhibition.

A) Heatmap showing the effect of Sox2 expression and PJ34 treatment on gene expression in MDA-MB-231 cells as determined by RNA-seq. The changes in expression were centered relative to the control (empty vector, vehicle treatment) condition.

B) Venn diagram showing the number of regulated genes under different treatment conditions as compared to empty vector non-treatment control condition.

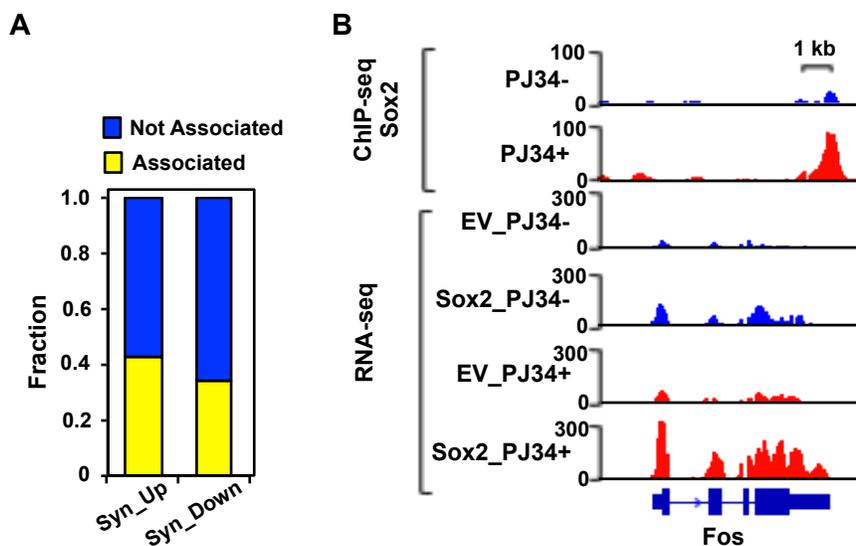


Figure 3.12. Modulation of Sox2-mediated gene transcription regulation through PAR inhibition.

A) Fraction of Sox2-regulated genes occupied by a newly gained Sox2 binding upon PAR inhibition within a 50 kb region surrounding TSSs. Syn-Up = genes whose expression is synergistically up-regulated by Sox2 and PARylation inhibition; Syn-down = genes whose expression is synergistically down-regulated by Sox2 and PARylation inhibition.

B) Genome browser tracks of RNA-seq and Sox2 ChIP-seq data showing examples of genes synergistically up-regulated by Sox2 and PAR inhibition which have newly gained Sox2 binding upon PAR inhibition.

Next, we explored the potential physiological effect by the combination of Sox2 expression and PAR inhibition through performing gene ontology analysis for genes that are synergistically regulated. Interestingly, the top GO terms of synergistically up-regulated genes include genes involved in cell apoptosis pathways and cell death pathways (Fig 3.13). On the other hand, synergistically down-regulated genes are enriched with cell cycle pathways (Fig 3.13). Blocking the PARylation activity thus exposes latent genomic binding sites of Sox2 and alters gene expression profile that may cause increased cell apoptosis, as well as slowed cell proliferation. This is a potential novel mechanism of the beneficial effects by using PAR inhibitor in treating Sox2 positive triple negative breast cancer.

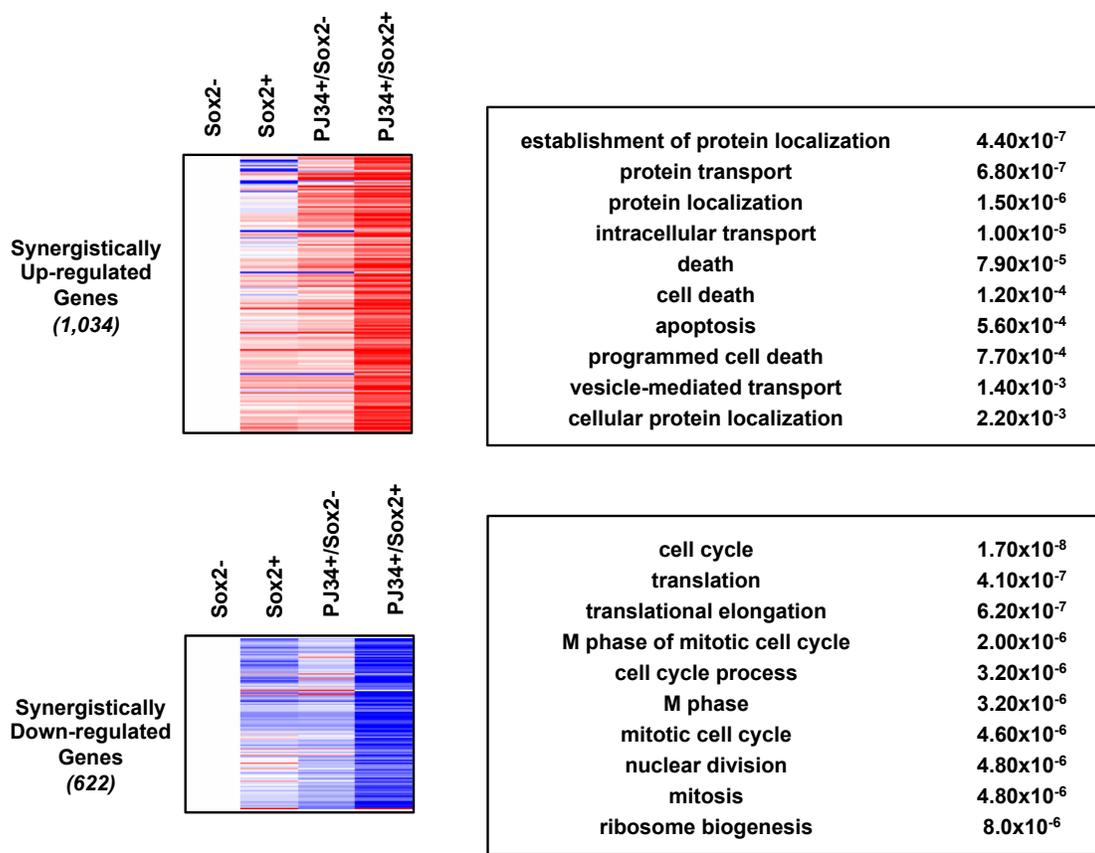


Figure 3.13. Modulation of Sox2-mediated gene transcription regulation through PAR inhibition.

(Left) Heatmap showing the effect of Sox2 expression and PJ34 treatment on gene expression in MDA-MB-231 cells as determined by RNA-seq. The changes in expression were centered relative to the control (empty vector, vehicle treatment) condition. Genes that are synergistically up-regulated (*Upper*) and down-regulated (*lower*) by Sox2 expression and PJ34 treatment are shown.

(Right) Top GO terms from gene ontology analysis for genes that are synergistically up-regulated (*Upper*) and down-regulated (*lower*) by Sox2 expression and PJ34 treatment.

3.4. Discussion

In this study, we have analyzed the function of Sox2 in driving triple negative breast cancer progression and metastasis. Interestingly, although Sox2-expressing TNBC cells grow slowly, they have higher potential to migration. This cellular phenotype of increased cell metastasis may explain the poor prognosis outcome for high Sox2 expression in triple negative breast cancer patients. Analysis of gene transcription profile

and genomic Sox2 binding profile helps us to gain better mechanistic insight of regulating TNBC by Sox2. Finally, by studying the interplay between PARylation inhibitor and Sox2 expression, we uncovered a potential novel mechanism of the beneficial role in using PAR inhibitor to treat TNBC patients.

3.4.1. Potential regulators mediating cell type-specific function of Sox2 in triple negative breast cancer

The mechanisms of cell type-specific genomic localization and function of Sox2 has been investigated in multiple studies. By associating with transcription factors that express in a cell type-specific manner, Sox2 achieves distinct chromatin association pattern in different cell types. This is demonstrated by the co-occupancy of Oct4 with Sox2 in embryonic stem cells, as well as BRN2 with Sox2 in neural progenitor cells²¹. In addition, p63 interacts with Sox2 at genetic loci in squamous cell carcinomas and is required for mediating Sox2-regulated gene transcription³⁷. In this study, we explored the cell type-specific genomic binding pattern of Sox2 by comparing Sox2 ChIP-seq in TNBC cells and human embryonic stem cells. Sox2 genomic localization in MDA-MB-231 cells is significantly different from in hES cells. Performing motif search in Sox2-enriched sites in TNBC provided several candidates that potentially co-occupy with Sox2 and regulated Sox2 genomic localization. Specifically, Sox4 motif and Hbp1 motif were identified to be the top most enriched motifs associated with Sox2 peaks. Interestingly, similar to Sox2, Sox4 has been found to play a critical role in mediating EMT process in breast cancer. Further studies are needed to uncover the functional link between Sox4 and Sox2 in triple negative breast cancer cells.

3.4.2. Interplay between PAR inhibition and Sox2 in gene transcription regulation

The regulatory roles of PARP-1 and PARylation on Sox2 function have been explored by multiple studies in embryonic stem cells. Our previous study identified a PARylation activity-independent function of PARP-1 as a pioneer factor stabilizing Sox2 bind to nucleosomes and regulate the transcriptional activity of Sox2 in undifferentiated ES cells. Upon differentiation, PARylation level of PARP-1 dramatically increases. Enhanced PARylation activity of PARP-1 was found to promote the interaction between PARP-1 and Sox2, as well as potentially decrease chromatin binding of PARP-1³⁵. This

leads to dissociation of Sox2 from chromatin, which thus decreases the activity of Sox2 in differentiating ES cells. In contrary, another group showed that PARP-1 is able to directly PARylate Sox2 during embryonic stem cell differentiation³⁸. This further decreases Sox2 protein stability, which is required for proper differentiation of ESC. These studies indicate a double-faced role of PARP-1 in regulating embryonic stem cell pluripotency depending on its enzymatic activity. That is, PARP-1 with a hypo-PARylation status in undifferentiated ES cells promotes Sox2 activity, while PARP-1 with a hyper-PARylation status in differentiating ES cells inhibits Sox2 function.

In this study, we explored the potential interplay between PARP-1 and Sox2 in the context of triple negative breast cancer. Consistent with the case in undifferentiated embryonic stem cell, depleting PARP-1 decreases Sox2 association with chromatin, suggesting that PARP-1 is able to promote Sox2 activity in TNBC. Interestingly, we observed a dramatic increase of Sox2 binding to a great number of genomic locations upon blocking PARylating activity in the same cell line. This may be due to increased chromatin association of PARP-1 by PARylation inhibition, which recruits Sox2 to novel binding sites. Consistent with this hypothesis, depleting PARP-1 protein abolishes the effect of PARylation inhibition, indicating that the effect of PJ34 on Sox2 chromatin binding is due to hypo-PARylated PARP-1 protein (Fig 3.14). This observation suggests an interesting feature of PARP-1 in regulating Sox2: in addition to double-faced function in different cell physiological conditions, PARP-1 and PARylation also play multifaceted roles in the same cell state. Local chromatin environment may cause different PARylation status of PARP-1, which in turn results in different response of Sox2 association on PARP-1 depletion and PARylation inhibition. Further studies are needed to uncover the signals that determine differential PARylation status in various chromatin environments in the same cells.

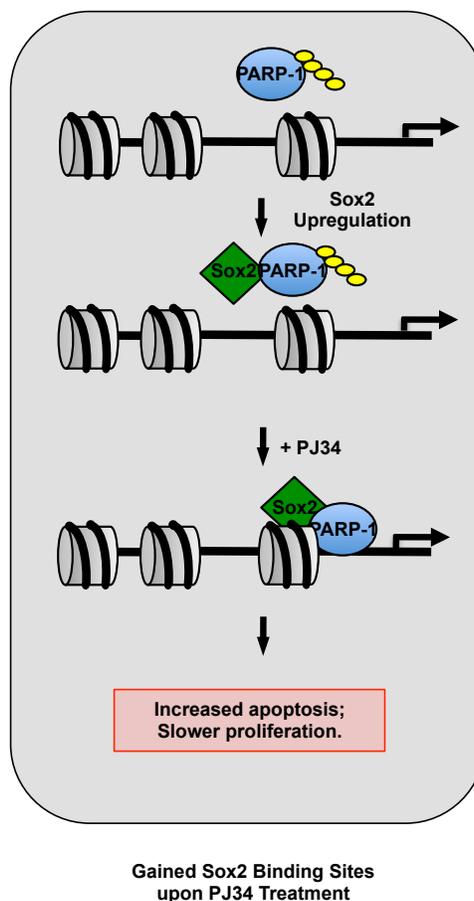


Figure 3.14. Model for regulation of Sox2 chromatin association by modulating PARylation activity.

3.4.3. Potential beneficial role of PAR inhibitor in treating Sox2-positive TNBC

Due to the well-studied functions of PARP-1 in mediating DNA damage response pathways^{39,40}, a large number of clinical studies have been conducted using PAR inhibitor in DNA repair-deficient cancers⁴¹. Treating these cancer patients with PAR inhibitor was found to sensitize the tumors to chemotherapy as well as radiation therapy⁴². Interestingly, many studies have shown that PARP inhibitor also inhibits the growth of cancers without BRCA1/2 mutations. This indicates a DNA damage pathway-independent manner of PARP inhibitor functioning in cancer^{41,43}. However, the underlying mechanism for this remains unclear. In this study, we used MDA-MB-231, a BRCA1/2 wild type triple

negative breast cancer cell line, as our model system. Surprisingly, we found a synergistic effect of PJ34 treatment and Sox2 expression on the gene expression profile in MDA-MB-231 cells through the PJ34 treatment-enhanced chromatin association of Sox2. Enhanced Sox2 chromatin association upon PJ34 treatment drives the “reprogramming” of gene expression profile, with increased expression of genes involved in cell death pathways and decreased expression of genes regulating cell cycle pathways. This may in turn reshape cell physiology, causing increased cell apoptosis and further slowed cell proliferation in TNBC cells. Further studies are needed to investigate the cellular phenotype in TNBC upon the combination of Sox2 expression and PJ34 treatment. If the effect of PAR inhibition on cell phenotype of Sox2-positive TNBC cells is consistent with gene expression changes, PAR inhibitor may have the potential to treat triple negative breast cancer using a novel mechanism, where it turns the transcription factor Sox2 into an inducer of cell apoptosis, promoting cell death in tumor cells.

Despite of this, caution is required when considering using PAR inhibitor to treat Sox2-positive cancers. More in-depth studies are needed in order to understand how PARylation inhibitor regulates Sox2 in a context-dependent manner. Specifically, since Sox2 is able to bind to and regulate genes responsible for cell migration and tumor metastasis, we need to have a better understanding of what determines the regulatory targets of Sox2. This is important for controlling the function of Sox2 using therapeutic drug like PAR inhibitor in treating breast cancer.

3.5. Materials and Methods

Kaplan-Meier Analysis

To examine the prognostic effect of Sox2 expression on breast cancer patients, Kaplan-Meier estimators were generated using the Kaplan-Meier Plotter (<http://kmplot.com/analysis/>)²².

Culturing Human Triple Negative Breast Cancer Cell Lines

Human triple negative breast cancer cell lines MDA-MB-231 and HCC1143 were obtained from ATCC. Both MDA-MB-231 and HCC1143 cells were maintained in RPMI-

1640 medium (Sigma; R5886) supplemented with 10% (v/v) FBS (Sigma) and penicillin and streptomycin (Invitrogen; 15140).

Inducible Ectopic Expression of hSox2 in Cells

A doxocycline (Dox)-inducible lentiviral vector for inducible expression of hSox2 in MDA-MB-231 and HCC1143 cells was generated. cDNA of hSox2 was inserted into pInducer20⁴⁴, and delivered into cells using lentivirus system. Cells stably infected were selected using neomycin/G418. Empty vector pInducer20 without hSox2 inserted was used as negative control. To induce ectopic expression of hSox2, 250 ng/mL Dox was added to cell culture medium. After 72 hours of induction, cells were collected and checked for Sox2 expression using Western Blotting.

Cell Proliferation Assay

MDA-MB-231 or HCC1143 cells were plated at a density of 2×10^4 cells per well in 6-well plates, with 250 ng/mL Dox added to culturing medium to induce Sox2 expression. The cells were collected every 2 days. Collected cells were washed with 1 x PBS, fixed with 10% formaldehyde, stained with 0.1% crystal violet containing 200 mM phosphoric acid, washed with water, and destained with 1 mL 10% acetic acid. The acetic acid destain was then collected and read at absorbance 595 nm. The results were expressed as cell growth⁴⁵.

Wound Healing Assay

MDA-MB-231 or HCC1143 cells were plated in 6-well plate, and cultured in the presence of 250 ng/mL Dox to induce Sox2 expression. 48 hours after Dox was added, wound was created by scratching the plate with a 200 μ L pipette tip. 24 hours after wound was created, cell migration was observed using bright-field microscope.

Antibodies

The antibodies used for Western Blotting, co-IP or ChIP were as follows: Sox2 (Santa Cruz; sc-17320), Hbp1 (Santa Cruz; sc-8488), SNRNP70, PARP-1 (previously

characterized custom rabbit polyclonal antibody; now available from Active Motif; 39561), PAR (home-made).

Analysis of Protein Levels by Western Blotting

Whole cell extracts (WCEs) were prepared from MDA-MB-231 cells using cell lysis buffer [50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 1.0 mM EDTA, 1% NP-40, and 10% glycerol, 1x protease inhibitor cocktail (Roche)]. When checking PARylation levels, 1 μ M PJ34 (to inhibits PARP activity) and 100 μ M tannic acid (to inhibit PARG activity) were added to lysis buffer. After clarification of the WCEs by centrifugation, aliquots containing equal amounts of total protein, as determined by a BCA assay (Pierce), were run on 8 to 10% polyacrylamide-SDS gels, transferred to nitrocellulose membrane, and subjected to Western blotting using the antibodies listed above and a chemiluminescent detection system (Thermo scientific).

Subcellular Fractionation and Analysis

MDA-MB-231 cells were grown in 15cm plate to a 75% confluence. Cells were harvested and resuspended in 1X isotonic lysis buffer (10 mM Tris-HCl pH 7.5, 2 mM MgCl₂, 3 mM CaCl₂, 0.3 M Sucrose, 1 mM DTT, and 1x protease inhibitor cocktail), followed by incubation on ice for 15 min.. NP-40 was then added to cell suspension to a final concentration of 0.6%, and vortexed vigorously for 10 seconds, followed by centrifuging. Supernatant containing the cytoplasmic fraction was then removed. Remaining ellet containing nuclei fraction was then extracted using buffer containing 20 mM HEPES pH 7.9, 1.5 mM MgCl₂, 0.42 M NaCl, 0.2 mM EDTA, 25% (v/v) Glycerol, 1 mM DTT, and 1x protease inhibitor cocktail, followed by vigorous vortexing. The supernatant containing nuclear extract was collected by centrifuging at maximum speed. The pellet was then resuspended in chromatin extraction buffer [1% SDS, 10 mM HEPES, 2 mM MgCl₂, 1 mM DTT, 1x protease inhibitor cocktail, and 50 U of benzonase (Sigma; E8263)], followed by incubation at 37°C for 1 hr with vigorous pipetting every 15 min.. The samples were then diluted in 1x SDS sample buffer and analyzed using Western Blotting.

RNA Extraction and RT-qPCR

RNA isolation and RT-qPCR were performed using a standard protocol, as previously described⁴⁶. Briefly, total RNA was extracted from MDA-MB-231 cells using TRI reagent (Sigma; T9424) following the manufacturer's instructions, reverse transcribed, and subjected to qPCR using the gene-specific primers listed below. Unless specified, all target gene expression levels were normalized to β -actin mRNA. All RT-qPCR assays were performed a minimum of three times to ensure reproducibility.

RNA-seq Library Preparation

MDA-MB-231 cells with both empty vector control and Dox-inducible Sox2 expression were cultured in the presence of 250 ng/mL Dox for 72 hours before cells were harvested. Total RNA was isolated using RNeasy Mini Kit (Qiagen, 74104) following the manufacturer's instructions. The integrity of the RNA was assessed and verified using an Agilent 2200 TapeStation system (Agilent Technologies) before mRNA-seq libraries were prepared using methods described previously. Briefly, polyA⁺ RNA was enriched using Dynabeads oligo(dT)25 (Invitrogen), heat fragmented, and reverse transcribed using random hexamers in the presence of dNTPs. Second strand cDNA synthesis was performed with dNTPs, but replacing dTTP with dUTP. After end-repair, dA-tailing, ligation to adaptors containing barcode sequences, and size selection using AMPure beads (Agencourt), the synthesized second-strand was digested using uracil DNA glycosylase (Enzymatics). A final PCR reaction was performed using Phusion high-fidelity DNA polymerase (NEB). After library quality control assessment using a Bioanalyzer (Agilent), the samples were subjected to 50 bp sequencing using an Illumina HiSeq 2000 Sequencing System.

Analysis of mRNA-seq Data

mRNA-seq reads were subjected to quality-control using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), trimmed to remove adapter sequences, and aligned to the human reference genome (hg19) using Tophat⁴⁷ before transcript assembly using Cufflinks⁴⁸. The data were then converted to wiggle (WIG) file format using PeakRanger for visualization using the Integrative Genomics Viewer

(IGV2.3)^{49,50}. Cuffdiff⁵¹ was used to identify genes that showed significant differential regulation upon Sox2 expression, with a false discovery rate (FDR) cutoff of 5%. The expression data were visualized in heatmaps using Java TreeView⁵², ranked in order based on the log₂ of the Sox2+ to Sox2- ratio.

The DAVID bioinformatics tool was used for gene ontology analysis²³. Gene set enrichment analysis (GSEA) tools developed by broad institute were used for the GSEA analyses (<http://www.broadinstitute.org/gsea/index.jsp>)²⁷.

Chromatin Immunoprecipitation and ChIP-qPCR

MDA-MB-231 cells with both empty vector control and Dox-inducible Sox2 expression were cultured in the presence of 250 ng/mL Dox for 72 hours before cells were harvested. The cells were cross-linked using 1% paraformaldehyde at room temperature for 10 min., followed by quenching in 125 mM glycine for 5 min. at 4°C. The crosslinked cells were collected, and resuspended in 300 µL lysis buffer [1% SDS, 10 mM EDTA, 50 mM Tris-HCl (pH 7.9), 1 mM DTT, 1x protease inhibitor cocktail (Roche)]. The resuspended cells were incubated on ice for 10 min. before sonicated at 4°C using a Biorupter (Diagenode) high setting, three cycles of 5 min. sonication (30 seconds on/30 seconds off with 5 min. intervals) to generate genomic DNA fragments of 200 - 500 bp in length. The sonicated chromatin was clarified by centrifugation and diluted 10 times by adding 2.7 mL dilution buffer (0.5% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH7.9, 150 mM NaCl, 1 mM DTT, 1x protease inhibitor cocktail) and pre-cleared with agarose beads.

Aliquots of the pre-cleared chromatin were immunoprecipitated with Sox2 or Hbp1 antibodies at 4°C overnight, followed by collection of the immunoprecipitates using protein G (Invitrogen) agarose beads. The beads were collected by gentle centrifugation and washed in low salt wash buffer (20 mM Tris-HCl pH 7.9, 2 mM EDTA, 125 mM NaCl, 0.05% SDS, 1% Triton X-100, 1x protease inhibitor cocktail), high salt wash buffer (low salt wash buffer containing 500 mM NaCl), and LiCl wash buffer (10 mM Tris-HCl pH 7.9, 1 mM EDTA, 250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1x protease inhibitor cocktail). The immunoprecipitated genomic DNA was eluted and the crosslinks were reversed by incubation in elution buffer (100 mM NaHCO₃, 1% SDS) at 65°C

overnight. The genomic DNA was then deproteinized by digestion with proteinase K and extraction with phenol:chloroform:isoamyl alcohol, followed by precipitation with ethanol. The ChIPed DNA was then subjected to qPCR using the locus-specific primers listed below. All ChIP-qPCR assays were performed a minimum of three times to ensure reproducibility.

ChIP-seq Library Preparation

ChIPed DNA was quantified using Qubit (Thermo). 10ng of ChIPed DNA was used to prepare each ChIP-seq library using methods described previously. Briefly, the genomic DNA fragments were end-polished, dA-tailed, and ligated to Y-adaptors containing barcode sequences. After agarose gel-based size selection and purification, the DNA was amplified for 13 - 15 cycles by PCR using Phusion high-fidelity DNA polymerase (NEB). The final ChIP-seq libraries were subjected to quality control assessment using a Bioanalyzer (Agilent), followed by 50 bp sequencing using an Illumina HiSeq 2000 Sequencing System.

ChIP-seq Data Analysis

Read alignment and peak calling. ChIP-seq reads were subjected to quality-control using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), trimmed to remove adapter sequences, and aligned to the human reference genome (hg19) using Bowtie 0.12.7⁵³, allowing for one mismatch during alignment. Significant peaks of Sox2 were called using PeakRanger-1.1 with a p-value cutoff of 1×10^{-4} . Genomic DNA isolated from sonicated chromatin without immunoprecipitation was used as an input control for identifying regions of enrichment. The data were then converted to wiggle (WIG) file format using PeakRanger⁵⁴ for visualization using the Integrative Genomics Viewer (IGV2.3)⁵⁰.

Annotating the Sox2 Peaks. To analyze the relation between Sox2 binding sites with RefSeq genes, bed file containing Sox2 peak location information was imported to the online ChIP data analysis tool ChIPseek (<http://chipseek.cgu.edu.tw>)⁵⁵. To examine the distance between Sox2 binding sites and the TSSs of RefSeq genes, the online

bioinformatics tool GREAT (<http://bejerano.stanford.edu/great/public/html/>)³⁰ was used. In order to explore the effects of PARylation inhibition on Sox2 binding, peak locations called by PeakRanger (Sox2 alone or Sox2 plus PJ34 treatment) were combined. The Sox2 peaks that were gained, maintained and lost upon PJ34 treatment were identified as described below. The reads under the combined pool of peaks for Sox2 were calculated for Sox2 alone (T1) and Sox2 plus PJ34 (T2). Rc was calculated using the following formula:

$$Rc = \log(T1/T2)$$

Where a larger Rc value means increased Sox2 binding upon PJ34-treatment, while smaller Rc values indicate decreased Sox2 binding. Median absolute deviation (MAD) was calculated for the Rc values and used as a cutoff to define gained ($Rc > xMAD$) and lost ($Rc < -xMAD$) peaks. A cut off of 4MAD was used to define changed ChIP-seq peaks⁵⁶.

Determining the Relation Between Sox2 Binding and Gene Expression Changes. To determine the relation between Sox2 occupancy and gene expression changes upon Sox2 expression, a 100 kb window (± 50 kb) surrounding the TSSs of RefSeq genes was searched for Sox2 binding as determined by ChIPseq. If at least one Sox2 peak is present within the 100 kb region, the gene is defined as associated with Sox2. The fraction of genes associated with Sox2 was then calculated.

Determining the Relation Between Sox2 Binding in hESCs and MDA-MB-231 cell. Significant peaks of Sox2 in hESC were called using PeakRanger-1.1 with a p-value cutoff of 1×10^{-4} . If the distance of Sox2 binding sites in hESCs and MDA-MB-231 cells is within 100 bp, we defined peaks as overlap between hESC and MDA-MB-231 cells.

Heatmaps. ChIP-seq read densities were visualized in heatmaps using Java TreeView⁵². For the heatmaps of Sox2, we determined the read densities for Sox2 in a 4 kb window (± 2 kb) around the Sox2 peak summit using a 20 bp moving window.

Determining the Chromatin States of Sox2 Binding Sites. ChromHMM²⁸ was used to determine chromatin states of MDA-MB-231 cell genome, using ChIPseq data sets of

histone marks H3K27me3, H3K4me3, H3K9ac, H4K8ac, H3K4me1, H3K27ac, and H3K36me3. We then defined the chromatin state that overlaps with the center of a Sox2 binding site as the chromatin state associated with the Sox2 binding site.

Motif Analysis. Motif analysis tool MEME-ChIP⁵⁷ was used to identify motifs enriched in Sox2 binding sites. 200 bp DNA sequence surrounding Sox2 ChIP-seq peak center was used as input to search for enriched DNA sequences.

REFERENCES

1. Hudis, C.A. & Gianni, L. Triple-negative breast cancer: an unmet medical need. *Oncologist* **16 Suppl 1**, 1-11 (2011).
2. Foulkes, W.D., Smith, I.E. & Reis-Filho, J.S. Triple-negative breast cancer. *N Engl J Med* **363**, 1938-48 (2010).
3. Lehmann, B.D. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* **121**, 2750-67 (2011).
4. Medema, J.P. Cancer stem cells: the challenges ahead. *Nat Cell Biol* **15**, 338-44 (2013).
5. Weina, K. & Utikal, J. SOX2 and cancer: current research and its implications in the clinic. *Clin Transl Med* **3**, 19 (2014).
6. Boumahdi, S. *et al.* SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. *Nature* **511**, 246-50 (2014).
7. Rudin, C.M. *et al.* Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. *Nat Genet* **44**, 1111-6 (2012).
8. Herreros-Villanueva, M. *et al.* SOX2 promotes dedifferentiation and imparts stem cell-like features to pancreatic cancer cells. *Oncogenesis* **2**, e61 (2013).
9. Kregel, S. *et al.* Sox2 is an androgen receptor-repressed gene that promotes castration-resistant prostate cancer. *PLoS One* **8**, e53701 (2013).
10. Berezovsky, A.D. *et al.* Sox2 promotes malignancy in glioblastoma by regulating plasticity and astrocytic differentiation. *Neoplasia* **16**, 193-206, 206 e19-25 (2014).
11. Li, X. *et al.* SOX2 promotes tumor metastasis by stimulating epithelial-to-mesenchymal transition via regulation of WNT/beta-catenin signal network. *Cancer Lett* **336**, 379-89 (2013).
12. Leis, O. *et al.* Sox2 expression in breast tumours and activation in breast cancer stem cells. *Oncogene* **31**, 1354-65 (2012).
13. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-4 (2012).
14. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1 (2013).

15. Abd El-Maqsoud, N.M. & Abd El-Rehim, D.M. Clinicopathologic implications of EpCAM and Sox2 expression in breast cancer. *Clin Breast Cancer* **14**, e1-9 (2014).
16. Suzuki, M. & Tarin, D. Gene expression profiling of human lymph node metastases and matched primary breast carcinomas: clinical implications. *Mol Oncol* **1**, 172-80 (2007).
17. Azzam, D.J. *et al.* Triple negative breast cancer initiating cell subsets differ in functional and molecular characteristics and in gamma-secretase inhibitor drug responses. *EMBO Mol Med* **5**, 1502-22 (2013).
18. Seo, E., Basu-Roy, U., Zavadil, J., Basilico, C. & Mansukhani, A. Distinct functions of Sox2 control self-renewal and differentiation in the osteoblast lineage. *Mol Cell Biol* **31**, 4593-608 (2011).
19. Santini, R. *et al.* SOX2 regulates self-renewal and tumorigenicity of human melanoma-initiating cells. *Oncogene* **33**, 4697-708 (2014).
20. Hutz, K. *et al.* The stem cell factor SOX2 regulates the tumorigenic potential in human gastric cancer cells. *Carcinogenesis* **35**, 942-50 (2014).
21. Lodato, M.A. *et al.* SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. *PLoS Genet* **9**, e1003288 (2013).
22. Gyorfyy, B. *et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* **123**, 725-31 (2010).
23. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
24. Huang da, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).
25. Minn, A.J. *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* **436**, 518-24 (2005).
26. Bos, P.D. *et al.* Genes that mediate breast cancer metastasis to the brain. *Nature* **459**, 1005-9 (2009).
27. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).

28. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-6 (2012).
29. Soufi, A., Donahue, G. & Zaret, K.S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994-1004 (2012).
30. McLean, C.Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501 (2010).
31. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-22 (2009).
32. Livraghi, L. & Garber, J.E. PARP inhibitors in the management of breast cancer: current data and future prospects. *BMC Med* **13**, 188 (2015).
33. Weil, M.K. & Chen, A.P. PARP inhibitor treatment in ovarian and breast cancer. *Curr Probl Cancer* **35**, 7-50 (2011).
34. von Minckwitz, G. & Martin, M. Neoadjuvant treatments for triple-negative breast cancer (TNBC). *Ann Oncol* **23 Suppl 6**, vi35-9 (2012).
35. Gao, F., Kwon, S.W., Zhao, Y. & Jin, Y. PARP1 poly(ADP-ribosyl)ates Sox2 to control Sox2 protein levels and FGF4 expression during embryonic stem cell differentiation. *J Biol Chem* **284**, 22263-73 (2009).
36. Rhie, S.K. *et al.* Nucleosome positioning and histone modifications define relationships between regulatory elements and nearby gene expression in breast epithelial cells. *BMC Genomics* **15**, 331 (2014).
37. Watanabe, H. *et al.* SOX2 and p63 colocalize at genetic loci in squamous cell carcinomas. *J Clin Invest* **124**, 1636-45 (2014).
38. Lai, Y.S. *et al.* SRY (sex determining region Y)-box2 (Sox2)/poly ADP-ribose polymerase 1 (Parp1) complexes regulate pluripotency. *Proc Natl Acad Sci U S A* **109**, 3772-7 (2012).
39. Dantzer, F. *et al.* Poly(ADP-ribose) polymerase-1 activation during DNA damage and repair. *Methods Enzymol* **409**, 493-510 (2006).
40. Wang, Z., Wang, F., Tang, T. & Guo, C. The role of PARP1 in the DNA damage response and its application in tumor therapy. *Front Med* **6**, 156-64 (2012).
41. Frizzell, K.M. & Kraus, W.L. PARP inhibitors and the treatment of breast cancer: beyond BRCA1/2? *Breast Cancer Res* **11**, 111 (2009).

42. Aly, A. & Ganesan, S. BRCA1, PARP, and 53BP1: conditional synthetic lethality and synthetic viability. *J Mol Cell Biol* **3**, 66-74 (2011).
43. Inbar-Rozensal, D. *et al.* A selective eradication of human nonhereditary breast cancer cells by phenanthridine-derived polyADP-ribose polymerase inhibitors. *Breast Cancer Res* **11**, R78 (2009).
44. Meerbrey, K.L. *et al.* The pINDUCER lentiviral toolkit for inducible RNA interference in vitro and in vivo. *Proc Natl Acad Sci U S A* **108**, 3665-70 (2011).
45. Sun, M., Gadad, S.S., Kim, D.S. & Kraus, W.L. Discovery, Annotation, and Functional Analysis of Long Noncoding RNAs Controlling Cell-Cycle Gene Expression and Proliferation in Breast Cancer Cells. *Mol Cell* **59**, 698-711 (2015).
46. Franco, H.L., Nagari, A. & Kraus, W.L. TNFalpha signaling exposes latent estrogen receptor binding sites to alter the breast cancer cell transcriptome. *Mol Cell* **58**, 21-34 (2015).
47. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11 (2009).
48. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-5 (2010).
49. Thorvaldsdottir, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-92 (2013).
50. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-6 (2011).
51. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**, 46-53 (2013).
52. Saldanha, A.J. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**, 3246-8 (2004).
53. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
54. Feng, X., Grossman, R. & Stein, L. PeakRanger: a cloud-enabled peak caller for ChIP-seq data. *BMC Bioinformatics* **12**, 139 (2011).

55. Chen, T.W. *et al.* CHIPseek, a web-based analysis tool for ChIP data. *BMC Genomics* **15**, 539 (2014).
56. Rao, N.A. *et al.* Coactivation of GR and NFκB alters the repertoire of their binding sites and target genes. *Genome Res* **21**, 1404-16 (2011).
57. Bailey, T.L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-8 (2009).