Title

Transcriptional and Epigenetic Regulation of Embryonic Stem Cell Pluripotency and Reprogramming

Author

Ho, Jaclyn J.

Publication Date

2016

Peer reviewed|Thesis/dissertation

Transcriptional and Epigenetic Regulation of Embryonic Stem Cell Pluripotency and Reprogramming

by

Jaclyn Je-ling Ho

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge
Professor Robert Tjian, Chair
Professor Richard Harland
Assistant Professor Dirk Hockemeyer
Associate Professor Danica Chen

Spring 2016

ABSTRACT

Transcriptional and Epigenetic Regulation of Embryonic Stem Cell Pluripotency and Reprogramming

Jaclyn Je-ling Ho

Chair of Committee:
Professor Robert Tjian
Department of Molecular and Cell Biology

Embryonic stem cells (ESCs), like all tissues, largely rely on precise transcriptional and epigenetic regulation for proper cell specification and differentiation. General transcription factors, transcriptional regulators, coactivators, and chromatin remodelers are often coordinated and expressed in a cell-type specific manner to ensure the integrity of this gene expression network. However, how ESCs are able to tightly regulate gene expression and remain highly malleable to external developmental cues is still an open biological question. The XPC DNA repair complex was recently identified as one of a set of critical stem cell coactivators (SCC) involved in driving robust OCT4/SOX2-dependent expression of pluripotent genes in ESCs. This dissertation will focus on the molecular mechanisms of the XPC complex and other SCCs in controlling cell identity and fate in pluripotent stem cells.

Chapter 1 will provide an introduction to the transcriptional and epigenetic regulation in pluripotent stem cells. Specifically, it will highlight the core transcriptional network, the mechanisms of reprogramming, and barriers to gene expression (e.g. DNA methylation). Chapter 2 will describe a novel role of the XPC DNA repair complex in regulating DNA methylation in both somatic and pluripotent cells, in part through regulating the activity of thymine DNA glycosylase, a major player in active DNA demethylation. Chapter 3 will describe the identification and characterization of two additional stem cell coactivators: the DKC1 ribonucleoprotein and ABCF1. As it is with the case of the XPC DNA repair complex, DKC1 and ABCF1 are multifunctional proteins with unanticipated roles in pluripotent gene transcription.

In summary, this dissertation will highlight the role of versatile proteins in pluripotent stem cells and their involvement in transcriptional regulation and beyond. By strengthening our understanding of the fundamental molecular mechanisms employed by pluripotent stem cells, we may gain a better insight into how cell identity is specified.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

My deepest gratitude goes out to a great number of individuals for their scientific, professional, and personal support along this journey that has brought me through graduate school to the completion of this dissertation.

I would like to thank my advisor, Prof. Robert Tjian, for his mentorship, support, and encouragement over the years. I have been amazingly fortunate to be given the freedom to explore my own biological questions, but at the same time the guidance to recover when projects faltered. It has been such a remarkable opportunity to be a member of the Tij Lab.

I also wish to thank the members of my thesis committee, Profs. Richard Harland, Dirk Hockemeyer, Danica Chen, and former committee member, Prof. Dave Schaffer, for all their feedback, advice, and guidance over the years. This dissertation would not have been possible without all of their support. I would also like to thank my rotation mentors, Profs. David Bilder and Lin He, who welcomed me into their labs during rotations and supported me during my graduate career.

I would like to thank my fellow labmates for all of their advice, stimulating discussions, camaraderie formed over late nights (and early mornings) in the lab, and for the occasional rescue from the coldroom when it was required. I am especially grateful to Prof. Yick Fong, Dr. Claudia Cattoglio, and Dr. Carla Inouye for the extensive training and scientific support they have provided during my graduate career. In particular, I would also like to thank lab members Elisa Zhang, Chiahao (Kevin) Tsui, Sheila Teves, David McSwiggen, Nikki Kong, Frank Xie, Mallory Haggart, Shuang Zheng, Gina Dailey, Janeen Lim, Lily Mirels, Claire Darzacq, and Prof. Xavier Darzacq. I would also like to thank all of my classmates and friends who have shared this journey with all of its ups and downs, including Nick Ellis, Alisha Ellis, Alex Wu, Jenn Cisson, Drew Friedmann, Mike Bond, Alex Wilkinson, Christine Lee, Mary Hong, Stephanie Chan, and Mailan Pham, among many others.

Last but not least, I would never have made it through graduate school and the completion of this dissertation without the unwavering support and encouragement from my family: my parents, David and Susan Ho, my siblings, Kathryn and Jonathan, as well as their spouses Chris and Maggie. I am eternally grateful for their unending love and support. And a special thanks to my nieces, Kylie and Charlotte – budding little scientists, for continuously reminding me that great things come in small, adorable packages.

The past 6 years have been an incredible journey and have shaped the scientist and person I am today. I anxiously await the next chapter and to what the future may bring.

CHAPTER ONE:

THE CROSSROADS OF TRANSCRIPTIONAL REGULATION AND DNA METHYLATION IN PLURIPOTENT STEM CELLS

Abstract

Proper cell fate specification is critical for eukaryotic development and is largely determined by gene regulatory networks and epigenetic signatures. These transcriptional regulatory networks require the coordination of RNA polymerase II, basal transcription machinery (e.g. TFIID), transcription factors, chromatin remodelers, and coactivators at cell-type specific gene promoters. Emerging evidence has recently highlighted the unanticipated diversification of multifunctional transcriptional complexes that serve as important transcriptional and epigenetic regulators. This chapter will dive into the molecular mechanisms that underpin the transcriptional network in embryonic stem cells and induced pluripotent stem cells, as well as the epigenetic forces (e.g. DNA methylation) that help maintain cell identity.

Part One: The Pluripotent Transcriptional Network

Pluripotent stem cells, such as embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs), have the unique ability to self-renew indefinitely and differentiate into specialized cell types. Thus, these cell types provide remarkable opportunities for regenerative medicine, disease-modeling, and drug discovery (Daley and Scadden, 2008; Yamanaka and Blau, 2010). While there have been breakthroughs and successful applications of iPSCs (Dimos et al., 2008; Hanna et al., 2007; Soldner et al., 2011; Wernig et al., 2008), they are currently limited in their use due to the low efficiency of reprogramming, genomic instability, and retention of epigenetic memory. By improving our understanding of the molecular basis of stem cell pluripotency and self-renewal, we may be able to overcome some of these barriers, which could impact their applications to basic research and therapeutics.

Cellular reprogramming strategies

Cellular reprogramming can be achieved through one of three strategies: nuclear transfer, cell fusion, or ectopic expression of transcription factors. Reprogramming is acutely sensitive to perturbations in factors required for the initiation or maintenance of the pluripotency program. Thus, although cellular reprogramming is largely an artificial process, much can be learned about pluripotency itself by understanding the mechanisms of its reacquisition.

Nuclear transfer involves the transplantation of a somatic nucleus into an enucleated oocyte. The success of nuclear transfer to create pluripotent cells challenged the existing dogma at the time that cell fate specification was irreversible (Gurdon, 1962). Furthermore, it suggested that cytoplasmic determinants in the oocyte were sufficient in reverting cell identity. Transcriptional profiling of embryos derived by nuclear transfer, artificial insemination, and in vitro fertilization demonstrate that embryos resulting from nuclear transfer undergo nearly complete reprogramming (Smith et al., 2005). However, nuclear transfer is a technically exceedingly difficult process and often results in many developmental abnormalities (Yang et al., 2007).

Reprogramming can also be achieved through the fusion of somatic and pluripotent cells. Both mouse and human pluripotent cells have the ability to reprogram somatic cells following cell fusion, resulting in the reactivation of silent pluripotent genes and changes to the epigenetic landscape (Cowan et al., 2005; Tada et al., 2001; Yu and Thomson, 2006). This process can be further enhanced by the overexpression of NANOG in ESCs (Silva et al., 2006). Transdifferentiation of cell identity is not unique to the cell fusion with pluripotent cells. Indeed, many fusions of different somatic cell types have also demonstrated cell fate plasticity from trans-acting factors. Furthermore, mixed species heterokaryons allows for the more comprehensive study of nuclear reprogramming given that gene expression changes in either nucleus can be monitored explicitly through species-specific differences in gene products (Blau et al., 1983).

Since reprogramming of heterokaryons occurs within 1-3 days following fusion, it allows for the rapid study of early reprogramming initiation events in the absence of cell division (Bhutani et al., 2010; Han et al., 2008; Pereira et al., 2008), but not DNA replication (Tsubouchi et al., 2013). Unfortunately, tetraploidy of the resulting hybrids makes reprogrammed cells by cell fusion unfit for many downstream applications.

Lastly, cellular reprogramming can be accomplished through the ectopic expression of transcription factors. Enforced expression of cell type-specific transcription factors (e.g. MYOD (Davis et al., 1987), C/EBPA (Xie et al., 2004), PAX5 (Cobaleda et al., 2007)) has been previously shown to result in transdifferentiation of one cell type to another. However until 2006, that had not been accomplished for pluripotent stem cells. Takahashi and Yamanaka showed that the expression of four factors – OCT4, SOX2, KLF4, and c-MYC (OSKM, also known as 'Yamanaka' factors) – could result in the formation of ESC-like iPSCs from mouse embryonic fibroblasts (MEFs) (Takahashi and Yamanaka, 2006). Just a year later, the same was shown for the human iPSCs (Takahashi et al., 2007; Yu et al., 2007). Since then, a plethora of different combinations of transcription factors (e.g. NANOG and LIN28 (Yu et al., 2007), UTF1 (Zhao et al., 2008), SALL4 (Tsubooka et al., 2009), TBX3 (Han et al., 2010), NR5A2 (Heng et al., 2010), TCL1 (Picanço-Castro et al., 2011), ESRRB (Buganim et al., 2012; Festuccia et al., 2012), TET1 (Gao et al., 2013)), miRNAs (Anokye-Danso et al., 2011; Judson et al., 2009), and small molecules (Bar-Nur et al., 2014; Federation et al., 2014) have been shown to be capable of inducing somatic cell reprogramming, highlighting the many facets and ways of reacquiring pluripotency.

The 'core' ESC regulatory network: NANOG, OCT4, and SOX2

Pluripotency and self-renewal is largely coordinated by a circuit of 'core' stem cell-specific transcription factors – namely NANOG, OCT4, and SOX2 (Chambers and Tomlinson, 2009; Jaenisch and Young, 2008). Together, these transcription factors co-occupy many of the same target genes and form both autoregulatory and feed-forward networks in ESCs (Boyer et al., 2005; Loh et al., 2006).

NANOG is a homeodomain transcription factor capable of maintaining ESC self-renewal in the absence of LIF (Chambers et al., 2003; Mitsui et al., 2003). Its expression can be detected in the morula and is exclusively confined to the cells of the inner cell mass (ICM) in the blastocyst. Ablation of NANOG in vivo leads to the failed specification of the ICM and differentiation into extraembryonic endoderm (Mitsui et al., 2003). However, conditional deletion of NANOG in ESCs demonstrates that NANOG is not required for self-renewal, suggesting that NANOG may play a larger role in stem cell specification rather than maintenance (Chambers et al., 2007). On the other hand, enforced expression of NANOG in ESCs results in sustained self-renewal and reduced differentiation potential (Chambers et al., 2003). Cell fusion-based reprogramming is increased by 200-fold when ESCs overexpress NANOG (Silva et al., 2006). NANOG is also critical in maintaining ground state, naïve pluripotency (Silva and Smith, 2008; Silva et al., 2009; Takashima et al., 2014).

The POU family transcription factor OCT4 (also known as POU5F1) is also critical for ESC specification. Zygotic OCT4 expression can be detected at the 8-cell stage in mouse embryos and is expressed exclusively in the ICM at the blastocyst stage (Palmieri et al., 1994). Ablation of OCT4 is embryonic lethal and results in the failure to form pluripotent cells (Nichols et al., 1998). A two-fold reduction in expression of OCT4 leads to the rapid differentiation of ESCs to trophectoderm, while a two-fold increase in expression leads to primitive endoderm, suggesting that the levels of OCT4 in ESCs is finely tuned (Niwa et al., 2000).

SOX2 is a high mobility group (HMG)-containing transcription factor that is known to interact with OCT4 in activating target genes (Ambrosetti et al., 2000; Kuroda et al., 2005; Rodda et al., 2005; Yuan et al., 1995). Traditional targeted deletion of SOX2 reveals a failure of the post-transplantation epiblast to maintain pluripotency (Avilion et al., 2003). Avilion and colleagues postulated that the delayed embryonic lethality in comparison to OCT4- and NANOG-null embryos could be explained by the persistence of maternally deposited SOX2, partial redundancy by other SOX family members, or the specific requirement of SOX2 in epiblast stem cells (EpiSCs). To address this, siRNAs against SOX2 were injected into 2-cell stage embryos, which revealed SOX2-depleted embryos arrested in the morula stage and failed to develop into blastocysts (Keramari et al., 2010). Similar to OCT4, depletion of SOX2 in ESCs also results in rapid differentiation to trophectoderm (Fong et al., 2008; Ivanova et al., 2006; Masui et al., 2007), which can be rescued by ectopic OCT4 expression (Masui et al., 2007).

Stem cell coactivators

In addition to these core stem cell-specific transcription factors, regulated pluripotent gene expression requires the orchestration of RNA polymerase II, general transcription factors (e.g. TFIID), ancillary transcriptional activators (e.g. ESRRB, TBX3, TCL1, STELLA, etc.), chromatin remodelers, and coactivator complexes at their cognate promoters (Näär et al., 2001). Even though OCT4 and SOX2 are critical in activating NANOG expression in ESCs (Kuroda et al., 2005; Rodda et al., 2005), combined with the abundant general transcriptional machinery available in differentiated cells, OCT4 and SOX2 are unable to activate the NANOG promoter in 293T and 3T3 cells (Rodda et al., 2005). This suggests other cofactors are required to potentiate NANOG transcription. A number of strategies have been used to identify additional players in the pluripotent network, including affinity-based purification and mass spectrometry (Kim et al., 2008; Liang et al., 2008; Wang et al., 2006), RNAi screens (Chia et al., 2010; Ding et al., 2009; Fazzio et al., 2008; Hu et al., 2009; Ivanova et al., 2006), transcriptional profiling (Ramalho-Santos et al., 2003; Richards et al., 2004; Sato et al., 2003; Sperger et al., 2003; Tanaka, 2002), and genome-wide ChIP-seq analyses (Boyer et al., 2005; Chen et al., 2008; Loh et al., 2006; Marson et al., 2008). To specifically address the role of coactivators in OCT4/SOX2-mediated transcription, an unbiased in vitro transcription-biochemical complementation assay was developed to identify novel stem cell coactivators (SCC) that could synergistically activate NANOG transcription with OCT4 and SOX2 (Fong et al., 2011). The screen identified three SCC

complexes: the XPC DNA repair complex (Fong et al., 2011), the dyskerin ribonucleoprotein (Fong et al., 2014), and the ATP-binding cassette (ABC) subfamily F member 1, ABCF1 (of which the characterization of the latter two SCCs will be discussed in Chapter 3).

The SCC/XPC complex is composed of three subunits: XPC, RAD23B, and CETN2. Together, the heterotrimeric complex plays a critical role in the global genome nucleotide excision repair (GG-NER) pathway, which targets helix-distorting pyrimidine dimers and bulky adducts for DNA repair (Mu et al., 1996; Reardon et al., 1996; Sugasawa et al., 1998, 2002). The XPC subunit is responsible for DNA damage sensing and recognition, though RAD23B and CETN2 stimulate activity in vivo and in vitro by stabilizing XPC (Araki et al., 2001; Ng et al., 2002; Okuda et al., 2004; Sugasawa et al., 1996). Once the DNA damage is recognized, XPC can sequentially recruit TFIIH, XPA, and endonucleases XPG and ERCC-XPF to excise the damaged region (Volker et al., 2001). Outside of DNA repair, RAD23B serves to shuttle ubiquitinated proteins to the proteasome (Chen and Madura, 2002); CETN2 is involved in centrosome duplication (Errabolu et al., 1994; Lee and Huang, 1993) and homologous recombination (Molinier et al., 2004). While XPC has an established role in GG-NER, its role as a transcriptional coactivator with OCT4 and SOX2 can be separated from its DNA repair activity (Fong et al., 2011).

Robust activation of NANOG by OCT4 and SOX2 requires SCC complexes, including the XPC complex. XPC is required for the maintenance of pluripotency in ESCs and for the reacquisition of pluripotency during cellular reprogramming. The ability of XPC to potentiate OCT4/SOX2-dependent transcription does not require DNA binding, as shown through the use of a DNA repair/binding-deficient XPC mutant in in vitro transcription assays (Fong et al., 2011), suggesting that the function of XPC as a DNA repair complex can be separated from its function as a transcriptional coactivator. Extensive biochemical and genomic occupancy analyses reveal the XPC complex can directly interact with OCT4 and SOX2 and co-occupies approximately 70% of OCT4/SOX2-bound promoter and gene regulatory elements (Cattoglio et al., 2015; Fong et al., 2011). Furthermore, XPC recruitment to OCT4/SOX2 binding sites is dependent on the presence of the activators, implying an active recruitment mechanism (Cattoglio et al., 2015).

Aside from its role in pluripotent gene transcription, XPC has also been identified as a transcriptional regulator in fibroblasts and HeLA cells (Le May et al., 2010). XPC binds at the RARβ2 gene promoter in response to retinoic acid treatment and facilitates the nucleation of additional NER complexes to coordinate gene expression, even in the absence of genotoxic stress (Le May et al., 2010). Depletion of XPC attenuates gene expression, though does not completely abolish it. The localization of NER complexes further results in local DNA demethylation and histone modifications through the recruitment of GADD45a and histone modifying enzymes, though the role of GADD45a in DNA demethylation remains controversial (Barreto et al., 2007; Jin et al., 2008; Le May et al., 2010). The role of XPC in this context is mechanistically distinct from the role XPC as a SCC, since additional factors are not required in the in vitro transcription assay in which it was discovered (Fong et al., 2011). However, it does highlight the

possibility that XPC may interact with a number of transcriptional activators in different cell types to promote transcription.

Part Two: Mammalian DNA Methylation

DNA methylation is a critical epigenetic modification that is stably propagated through cell divisions and functionally impacts gene expression. 5-methylcytosine (5mC) occurs at 70-90% of CpG dinucleotides in the mammalian somatic cell, which represents approximately 1% of all cytosines in the genome (Ehrlich et al., 1982). Naïve pluripotent cells, however, are hypomethylated and contain approximately 25% CpG methylation (Ficz et al., 2013; Guo et al., 2016; Leitch et al., 2013). Indeed, there is a sharp decline in methylation during embryonic development following fertilization, which is restored by de novo methylation at the time of implantation (Kafri et al., 1992; Monk et al., 1987). Because methylation can be faithfully maintained, methylation patterns set during implantation presumably establish an epigenetic barrier that reduces developmental potential and promotes cellular specification during development.

DNA methyltransferases

There are four DNA methyltransferase (DNMT) family members: DNMT1, DNMT3A, DNMT3B, and DNMT3L, which are predominantly responsible for both establishing and maintaining 5mC in the mammalian genome. DNMT3L is the only family with no methyltransferase enzymatic activity, but can modulate activity of DNMT3A/B (Kareta et al., 2006). DNMT1 is required for the genomic methylation maintenance, while DNMT3A/B are tasked with de novo methylation early in development. Correspondingly, DNMT1 is expressed in all somatic tissues, while DNMT3A and DNMT3B expression is limited to the oocyte/preimplantation embryo and epiblast respectively, which are responsible in establishing early methylation patterns during development and are downregulated in differentiated tissues (Okano et al., 1998; Watanabe et al., 2002).

DNMT1 maintains heritable methylation patterns by specifically targeting hemi-methylated DNA at replication foci and catalyzes the transfer of a methyl group from the donor S-adenosyl-L-methionine to the 5-position of an unmethylated cytosine (Chen et al., 1993; Leonhardt et al., 1992; Santi et al., 1983; Yoder et al., 1997). Autoinhibition by the N-terminus of the DNMT1 and association with URHF1 ensures DNMT1 binds with high fidelity to hemi-methylated DNA, preventing unscrupulous DNMT activity (Avvakumov et al., 2008; Bostick et al., 2007; Sharif et al., 2007; Song et al., 2012). Underscoring the importance of DNA methylation in development, DNMT1 knockout is embryonic lethal by E9.5-10.5 in mice, leading to decreased genomic methylation and severe developmental defects (Lei et al., 1996; Li et al., 1992).

Although the mechanism of methylation maintenance is conceptually simple, it has been reported that DNMT1 does not always copy the methylation on the daughter strand at the same position as the parental strand, but rather copies to nearby positions on the daughter strand (Silva et al., 1993; Stöger et al., 1997). This explains why methylation patterns can still be quite variable among clonal cells, though the overall methylated vs. unmethylated state of a region is maintained (Bird, 2002). There may additionally be supplementary mechanisms for the maintenance of methylation that have yet to be determined, which could potentially explain how some regions maintain methylation even in the absence of DNMT1 (Rhee et al., 2002).

De novo methylation, on the other hand, is primarily driven by DNMT3A and DNMT3B. The genetic ablation of either DNMT3A or DNMT3B leads to lethality in mice. DNMT3B-null mice exhibit embryonic lethality at E14.5-18.5, while DNMT3A-null mice appear normal at birth, but quickly become runted and die at 4 weeks of age (Okano et al., 1999). Homozygous double knockout embryos arrest shortly after gastrulation and exhibit similar developmental and morphological defects to DNMT1 knockout embryos (Lei et al., 1996; Okano et al., 1999). No de novo methylase activity is reported in DNMT3A and DNMT3B double knockouts, indicating that both may share some functional redundancy and together are essential for de novo methylation.

DNA methylation and gene silencing

DNA methylation has long been understood as an epigenetic gene silencing mark. Interestingly, most examples of gene silencing occurs before DNA methylation, suggesting that DNA methylation may act not to silence transcribed genes, but to irrevocably silent genes that have already been downregulated (Turker, 2002). Indeed, DNA methylation has been shown to be necessary for complete X-chromosome inactivation (Csankovszki et al., 2001; Sado et al., 2000).

The mechanism of gene silencing by DNA methylation likely occurs through the combination of two pathways that suppress transcription. First, methylated DNA can prevent the binding of some transcription factors to their consensus sequence (Figure 1). Of note, CTCF binding on the maternally inherited Igf2 locus prevents the H19 enhancer from activating Igf2 transcription. However, the paternally inherited locus, which contains a methylated CTCF site between the H19 enhancer and Igf2 promoter, is transcriptionally active due to the loss of CTCF binding (Bell and Felsenfeld, 2000; Hark et al., 2000; Szabó et al., 2000). Recently, DNA methylation-sensitive CTCF binding was also shown to prevent activation of PDGRA from an upstream enhancer, which is disrupted in IDH mutant gliomas (Flavahan et al., 2015). Though there are many examples supporting CTCF as a methylation-sensitive binding factor, it is unlikely that methylation-sensitive CTCF binding is a universal phenomenon. Only 1.5% of CTCF sites genome-wide appear to have differential binding of CTCF when cells are artificially hypomethylated (Maurano et al., 2015), suggesting DNA methylation may drive locus-specific effects with regards to CTCF. Additionally the binding of some

transcription factors such as CTF and SP1 are methylation-insensitive, indicating additional mechanisms for gene silencing are required.

In contrast to the inhibition of transcription factor binding by methylation, gene silencing can also be accomplished through DNA methylation by the recruitment of repressive methyl-binding proteins, which associate with histone deacetylases (HDAC) and chromatin remodelers (Figure 1). There are currently seven known mammalian methyl-binding proteins that repress transcription: the methyl-CpG-binding domain (MBD) proteins MeCP2 (Lewis et al., 1992), MBD1, MBD2 (originally identified as a component of the MeCP1 complex) (Meehan et al., 1989), MBD3 (Hendrich and Bird, 1998), MBD4 (Kondo et al., 2005), and the Kaiso-like proteins ZBTB4 and ZBTB38 (Filion et al., 2006). All of these methyl-binding proteins have been shown to associate with chromatin remodelers, leading to histone deactylation and chromatin compaction (Filion et al., 2006; Jones et al., 1998; Kondo et al., 2005; Nan et al., 1998; Ng et al., 1999; Wade et al., 1999; Zhang et al., 1999). In fact, treatment with a specific HDAC inhibitor, trichostatin A (TSA), can lead to derepression of MeCP2-mediated silenced reporters, further implicating the role of HDACs in gene silencing. Interestingly, MBD3 is the only MBD family member whose loss leads to embryonic lethality (Hendrich et al., 2001), potentially due to ESCs failing to silence pluripotent genes, thus leading to defective differentiation during development (Kaji et al., 2006).

Passive vs. active demethylation

The removal of DNA methylation can occur through passive (replication-dependent) or active (replication-independent) mechanisms. Passive demethylation relies on repeated rounds of DNA replication that leads to the gradual loss and dilution of 5mC when DNMTs are either downregulated, excluded from the nucleus, or unable to recognize the substrate (in the case of oxidized products of 5mC). After fertilization, the maternal genome undergoes slow, gradual loss of 5mC that occurs over several rounds of replication (Harland, 1982). Coincidentally, there is nuclear exclusion of oocyte-specific DNMT1 (DNMT1o), leading to passive demethylation (Cardoso and Leonhardt, 1999; Carlson et al., 1992). Additionally, the sequestration of URHF1 (also known as Np95) by Stella can prevent DNMT1 from associating with replication foci and thus inhibiting 5mC maintenance in ESCs (Funaki et al., 2014).

In contrast to passive demethylation, active demethylation does not require DNA replication, but instead relies on oxidizing and/or deaminating enzymes, followed by base excision repair (BER). One such classic example is the demethylation of the paternal genome following fertilization. Whereas the maternal genome undergoes a slow, passive loss of methylation, the paternal genome undergoes a rapid removal of 5mC within 6-8 hrs post fertilization (Mayer et al., 2000). Following fertilization, maternally deposited TET3 is recruited to the male pronucleus and catalyzes the conversion of 5mC to 5hmC (Gu et al., 2011), which can subsequently be further oxidized and removed by BER or lost through non-recognition of 5-hydroxymethylcytosine (5hmC) by DNMT1 (Inoue and Zhang, 2011). Recent data,

however, has revealed that demethylation following fertilization is more complex. Both DNA replication and TET3-mediated active demethylation play a large role in the demethylation of the maternal and paternal genomes in mice (Guo et al., 2014a; Shen et al., 2014).

Mechanisms of active demethylation

Despite decades of research, the exact mechanism of active demethylation remains elusive. However, it's generally accepted that active demethylation must involve a deamination event and/or series of oxidation events, leading to 5mC intermediates that can then be repaired through BER (Figure 2). Deamination of 5mC will produce thymine (T), which is subsequently recognized as a T:G mismatch by BER. The activation-induced cytidine deaminase (AID) and APOBEC family have been implicated as being able to catalyze this deamination event in vitro (Morgan et al., 2004). However, AID/APOBEC1 is only able to deaminate 5mC to T in the context of ssDNA (Morgan et al., 2004). The affinity of AID/APOBEC1 for 5mC is also considerably lower than unmodified cytosines, suggesting that AID/APOBEC1 would be more mutagenic than beneficial as a DNA 'demethylase' (Bransteitter et al., 2003; Larijani et al., 2005). Recently, AID has been reported to be required for DNA demethylation associated with heterokaryon-based reprogramming (Bhutani et al., 2010), but because reprogramming by cell fusion also requires DNA synthesis it is unclear how much demethylation occurs due to AID-mediated active demethylation vs. passive demethylation through DNA replication (Tsubouchi et al., 2013). Notably the TET family hydroxylases – which include TET1, TET2, and TET3 – have been shown to oxidize 5mC to 5-hydroxymethylcytosine (5hmC) and to further oxidized products 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) (He et al., 2011; Ito et al., 2010, 2011; Tahiliani et al., 2009). In vivo evidence further supports the role of TET proteins and oxidative demethylation in development, including ICM specification and ESC self-renewal (Gao et al., 2013; Gu et al., 2011; Ito et al., 2010; Koh et al., 2011; Shen et al., 2013). TET hydroxylases can also further oxidize T to 5-hydroxymethyluracil (5hmU) (Pfaffeneder et al., 2014).

The second step of active demethylation following deamination or oxidation requires BER enzymes to recognize and excise the 5mC intermediate product (Figure 2). Several BER enzymes have been implicated in this pathway, including MBD4, single-stranded monofunctional uracil DNA glycosylase 1 (SMUG1), and thymine DNA glycosylase (TDG). MBD4 was shown to cooperate with AID and GADD45a to remove T:G mismatches in zebrafish (Rai et al., 2008). Because the role of GADD45a is highly controversial and has not been reproduced (Barreto et al., 2007; Engel et al., 2009; Jin et al., 2008), it will not discussed here. SMUG1, on the other hand, can specifically remove 5hmU from both ssDNA and dsDNA (Boorstein et al., 2001).

Current prevailing models favor the repair of demethylation intermediates by thymine DNA glycosylase (TDG), which can recognize and act on T:G, 5hmU:G, 5fC:G, and 5caC:G substrates (list of substrates is not exhaustive, see Cortázar et al., 2007 for

comprehensive list) (Hashimoto et al., 2012a; He et al., 2011; Maiti and Drohat, 2011; Neddermann et al., 1996). Consistent with this, TDG is the only member of the uracil N-glycosylase (UNG) family whose knockout leads to embryonic lethality associated with developmental defects and aberrant methylation patterns (Cortázar et al., 2011; Cortellino et al., 2011). Furthermore, TDG-null ESCs accumulate 5fC and 5caC at promoters and gene regulatory elements genome-wide, which is consistent with TET/TDG-dependent active demethylation (Shen et al., 2013). In vitro data, however, suggests that TDG exhibits an exceedingly slow turnover rate due to its high affinity to the abasic site created following cleavage (Waters and Swann, 1998; Waters et al., 1999). Thus DNA demethylation through TDG must involve either other DNA demethylation intermediates – which has since then been shown – or other cofactors to explain the fast kinetics of active demethylation. One such cofactor capable of stimulating TDG activity will be discussed in Chapter 2. While there is clear evidence for the role of TDG in active demethylation, it's important to note that 5hmC, 5fC, and 5caC are not recognized by DNMT1 and thus can also result in passive demethylation (Hashimoto et al., 2012b; Song et al., 2012).

The emerging field of non-CpG methylation

Aside from canonical CpG methylation, there's increasing evidence for the role of non-CpG methylation (also termed CpH methylation, where H = A, T, or C) in ESCs and development (He and Ecker, 2015). Early observations noted the presence of CpH methylation in mammalian DNA through high performance liquid chromatography (HPLC) and thin layer chromatography (TLC) (Nyce et al., 1986; Salomon and Kaye, 1970), even accounting up to a little over half of all 5mC in human spleen DNA (Woodcock et al., 1987), but the function of CpH methylation has been elusive. It was not until the use of whole genome bisulfite sequencing, which can identify methylated cytosines (5mC and 5hmC) at base pair resolution, that the abundance of 5mC in the context of CpH could be assayed.

Mammalian CpH methylation is found in relatively high abundance in ESCs and brain tissue. Nearly 25% of all 5mC found in human ESCs (Lister et al., 2009; Ramsahoye et al., 2000) and murine neurons (Guo et al., 2014b) are within a CpH context, with the majority in post-mitotic neurons specifically being in CpHpH motifs, which lead to asymmetric methylation on only one DNA strand. CpH methylation can also be found in a wide range of differentiated cell types, but at a much lower frequency (Barrès et al., 2009; Haines et al., 2001; Lister et al., 2009, 2011, 2013; Ma et al., 2014; Schultz et al., 2015; Xie et al., 2013). Unlike CpG methylation, CpH methylation appears to be primarily maintained by DNMT3A and DNMT3B (Barrès et al., 2009; Guo et al., 2014b; Lister et al., 2009; Ramsahoye et al., 2000).

The biological function of CpH methylation has still yet to be established. Similar to CpG methylation, however, current evidence suggests CpH methylation may also be correlated with gene silencing. In type II diabetes mellitus patients, CpH methylation of the PGC-1a promoter is correlated with decreased PGC-1a expression and increased

obesity rates (Barrès et al., 2009). In neurons, CpH can be recognized by MeCP2 leading to the recruitment of HDACs and subsequent chromatin compaction (Guo et al., 2014b). Consistent with this, genome-wide distribution of CpH methylation in ESCs reveals that it is greatly enriched in repetitive elements, inactive enhancers, and gene bodies, but absent from promoters, active enhancers, and transcription factor binding sites. Interestingly, CpH enrichment in gene bodies correlates with highly active genes, but is present on the antisense strand, suggesting that CpH methylation may play a role in silencing antisense transcription (Lister et al., 2009). Because CpH methylation is found in regions deficient CpG nucleotides, CpH methylation has also been proposed as a compensating silencing mechanism when CpG methylation cannot occur (He and Ecker, 2015). CpH also appears more dynamic than CpG methylation and is rapidly removed upon downregulation of DNMT3A (Guo et al., 2014b).

Reprogramming and epigenetic memory

Transcription factor mediated cellular reprogramming is a largely inefficient, stochastic process (Hanna et al., 2009). Given the substantial changes in gene expression that must occur to silence somatic genes and reactivate an ESC-specific transcriptome, it is not surprising that DNA demethylation has been shown to be a significant roadblock in reprogramming in iPSCs and nuclear transfers (Bortvin et al., 2003; Dean et al., 2001; Mikkelsen et al., 2008; Simonsson and Gurdon, 2004). Underscoring the importance of DNA demethylation during reprogramming, DNMT1 knockdown or 5-azacytidine treatment increases reprogramming efficiency (Mikkelsen et al., 2008). TET1/2/3 and TDG are also essential for iPS reprogramming, in part through the demethylation of genes involved in the mesenchymal-to-epithelial transition and at the *Nanog* locus (Costa et al., 2013; Doege et al., 2012; Hu et al., 2014). TET1 has further been shown to act on the *Oct4* enhancer during reprogramming and is able to replace OCT4 as a reprogramming factor, in combination with SOX2, KLF4, and c-MYC (Gao et al., 2013).

Although iPSCs largely mimic ESCs, iPSCs still retain a distinct transcriptional signature that differs from ESCs (Chin et al., 2009). This may at least be in part due to DNA methylation that is not entirely reprogrammed, resulting in a somatic ‘epigenetic memory’ that is characteristic of the cell type of origin and biases differentiation potential (Kim et al., 2010; Ohi et al., 2011; Polo et al., 2010). Indeed, promoter methylation has been demonstrated to be strongly correlated with incomplete gene reactivation in iPSCs (Ma et al., 2014). Cells resulting from somatic cell nuclear transfer (SCNT) more accurately reflect the epigenetic landscape of ESCs, suggesting the persistence of epigenetic memory may be more specific to a partially reprogrammed state achieved through transcription factor-mediated reprogramming rather than a fundamental aspect of somatic reprogramming (Ma et al., 2014; Polo et al., 2010). Furthermore, the specific combination of factors used in iPS reprogramming also affects the extent of DNA demethylation, particularly at transcription factor-bound regions (Planello et al., 2014).

Concluding Remarks

Although the transcriptional and epigenetic regulations of pluripotency and self-renewal have been well studied over the past few decades, a number of exciting, critical questions remain open. Continued mechanistic exploration is also needed to understand the close, tight relationship between DNA methylation and transcription. Indeed, transcriptional activity may somehow mark methylation-free regions of the genome, as the removal of transcription factor (e.g. SP1) binding can result in locus-specific promoter methylation (Brandeis et al., 1994; Macleod et al., 1994), while its introduction near transgene and retroviral elements can promote demethylation (Hejnar et al., 2001; Siegfried et al., 1999). This raises the interesting possibility that there may be constant counteracting forces between transcription and methylation, such that methylation-free regions are not static, but must actively be maintained. Current ongoing research on the newly identified 5hmC, 5fC, and 5caC bases will undoubtedly uncover new avenues to how methylation can be dynamically altered and affect gene expression.

Figures

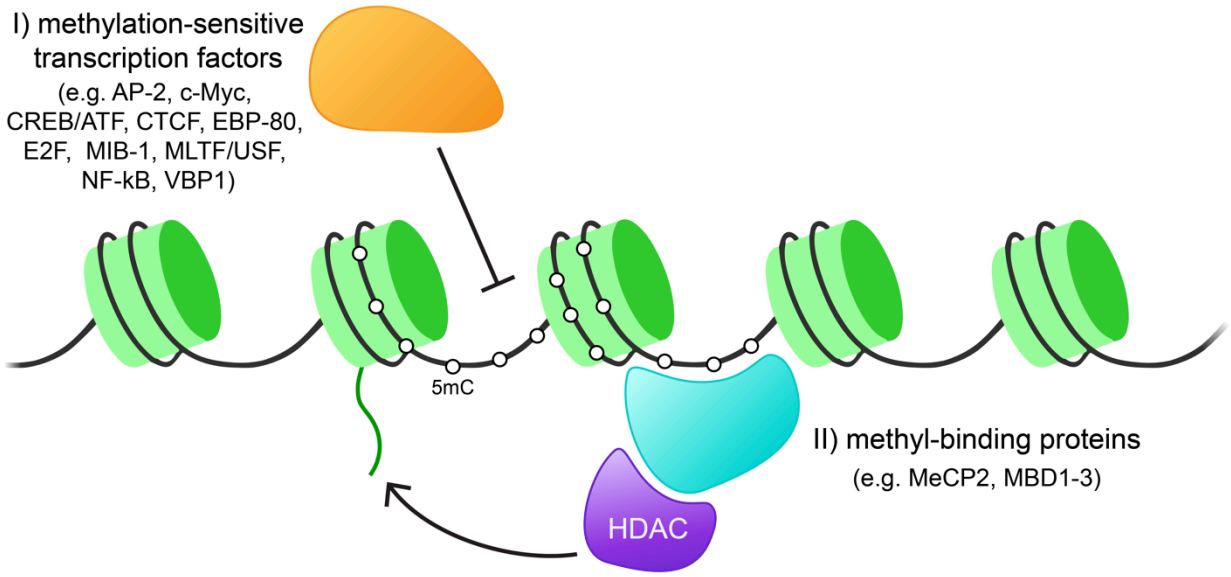

Figure 1. Mechanisms of gene silencing due to DNA methylation. Methylation-sensitive transcription factors can be excluded from binding sites due to presence of 5mC. Alternatively, methyl-binding proteins can bind to 5mC, hindering the binding of transcriptional activators while also recruiting HDACs and other chromatin remodelers.

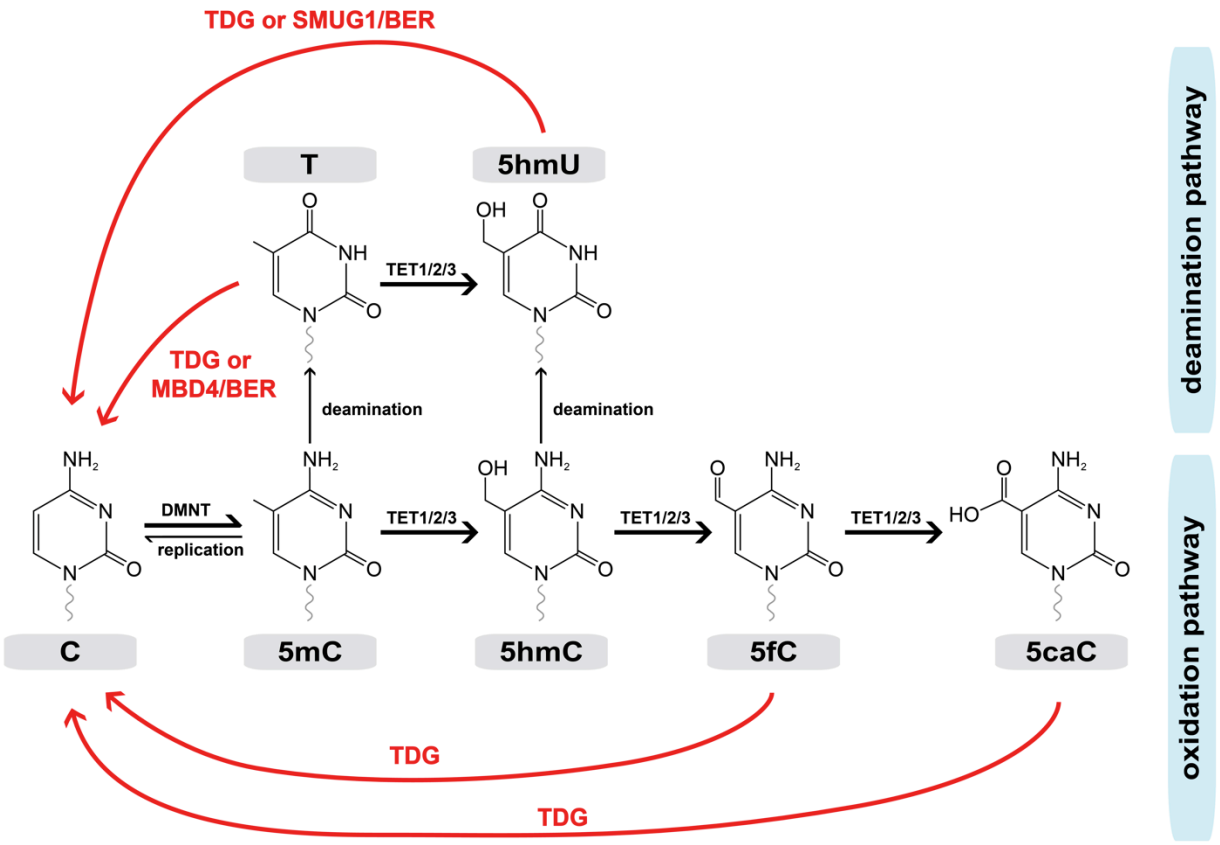

Figure 2. Overview of pathways involved in mammalian DNA demethylation. 5-methylcytosine (5mC) can be oxidized by TET hydroxylases to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), which can be recognized and repaired by thymine DNA glycosylase (TDG) followed by base excision repair (BER). Alternatively, 5mC or 5hmC can undergo a deamination event and be repaired by TDG/MBD4 or TDG/SMUG1, respectively. C, cytosine; DNMT, DNA methyltransferase;

CHAPTER TWO:

REGULATION OF DNA DEMETHYLATION BY THE XPC DNA REPAIR COMPLEX IN SOMATIC REPROGRAMMING

Abstract

Somatic cell reprogramming requires cells to undergo dramatic changes in gene expression, DNA methylation, and chromatin structure. While current reprogramming approaches give rise to induced pluripotent stem cells (iPSCs) that largely reflect embryonic stem cells (ESCs), persistent transcriptome differences and the perdurance of somatic 'epigenetic memory' hinder the safe and efficacious use of these cells. Therefore, understanding the molecular mechanisms by which a pluripotent-specific transcriptome and epigenome are established in iPSCs is essential for overcoming current barriers in reprogramming. Given that the XPC DNA repair complex has recently been shown to play an important role in ESC-specific transcription and a potential role in stimulating thymine DNA glycosylase (TDG) activity, we investigated how XPC influences the epigenetic landscape during reprogramming. Here, we manipulated the expression of XPC and demonstrated an inverse correlation between global 5-methylcytosine levels and XPC expression, independent of XPC's DNA repair activity. We confirmed XPC's ability to enhance the activity of TDG, a major player in active DNA demethylation, using complementary in vivo and in vitro techniques. ChIP-seq analyses further revealed that XPC and TDG colocalize extensively at gene promoters and regulatory elements in ESCs. Furthermore, overexpression of XPC during human somatic cell reprogramming resulted in the generation of more robust iPSCs, improving cell survival during single-cell dissociation challenges. These findings suggest that XPC may play a role in coupling efficient DNA demethylation and robust ESC-specific transcription through the recruitment of TDG activity to sites bound by XPC and core ESC transcription factors.

Introduction

Pluripotent stem cells (PSCs), such as embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs), have the unique ability to self-renew indefinitely and differentiate into specialized cell types. Thus, these cell types provide remarkable opportunities for regenerative medicine, disease-modeling, and drug discovery (Daley and Scadden, 2008; Yamanaka and Blau, 2010). The reacquisition of pluripotency during reprogramming requires cells to undergo a series of molecular events that involve extensive changes in gene expression, DNA methylation, and chromatin structure (Jaenisch and Young, 2008; Orkin and Hochedlinger, 2011; Sindhu et al., 2012; Stadtfeld and Hochedlinger, 2010). While current human reprogramming strategies give rise to iPSCs that largely reflect ESCs, persistent transcriptome differences and the perdurance of somatic 'epigenetic memory' hinder the safe and efficacious use of these cells in regenerative medicine (Daley and Scadden, 2008; Kim et al., 2010; Ohi et al., 2011; Polo et al., 2010).

Efficient DNA demethylation, specifically the removal of 5-methylcytosine (5mC) at gene regulatory elements, plays a critical role in the activation of the pluripotency network, suppression of somatic gene expression, and resetting epigenetic memory (Bird, 2002; De Carvalho et al., 2010). Thus, understanding the role of active DNA demethylation may be key to improving current iPSC induction methods. Although multiple DNA demethylation pathways have been proposed (Chapter 1, Figure 2), all pathways converge on a single enzyme, thymine DNA glycosylase (TDG), which can excise deaminated and oxidized products of 5mC, triggering base excision repair (BER) of the abasic site to an unmodified cytosine (Wu and Zhang, 2010). Consistent with this, TDG is required for the activation of de novo transcription of developmentally regulated genes by promoting demethylation (Cortázar et al., 2007). It is also the only enzyme of its class whose knock-out leads to embryonic lethality and aberrant methylation patterns (Cortázar et al., 2011; Cortellino et al., 2011). Given that a large number of genes must be silenced and reactivated during reprogramming, TDG is a likely candidate to mediate DNA demethylation and subsequent gene reactivation during this process.

Recently, the XPC DNA repair complex, composed of XPC-RAD23B-CETN2, has been described to be an essential component of the pluripotency regulatory network and a transcriptional co-activator for OCT4 and SOX2 – "core" stem cell specific transcription factors that regulate pluripotency (Boyer et al., 2005; Fong et al., 2011; Jaenisch and Young, 2008). Extensive biochemical characterization of the XPC complex and its genome-wide co-occupancy with OCT4/SOX2 both suggest that this complex is recruited to regulatory elements also bound by OCT4 and SOX2 (Fong et al., 2011). Interestingly, XPC has been previously shown to stimulate the repair of T:G mismatches by TDG in vitro (Shimizu et al., 2010). This raises the possibility that XPC could regulate transcription by coordinating with TDG in active DNA demethylation of regulatory regions bound by OCT4/SOX2. Unlike treatment with global DNA methyltransferase inhibitors, such as 5-azacytidine, which can lead to the improper demethylation of genes normally silenced in ESCs (Mikkelsen et al., 2008), XPC may

be a potential candidate for safely enhancing iPSC fidelity while also coupling robust ESC-specific transcription and genome surveillance.

Results

XPC expression affects global methylation, independent of DNA repair

To determine the relationship between the XPC complex and the DNA methylation landscape, we first manipulated the XPC expression in human dermal fibroblasts (HDFs) and ESCs and measured the amount of global 5mC. The loss-of-function of XPC through shRNA-mediated knockdown in HDFs (Figure 1A) and Cas9-mediated knockout of XPC in H9 hESCs (Figure 1B) led to an increase in global 5mC, as determined by a 5mC-specific ELISA. This global increase of 5mC is consistent with a previous report using siRNAs against XPC in HeLa cells (Le May et al., 2010). We also confirmed these results in mouse ESCs, suggesting that this is not a human-specific phenomenon and may be a conserved function of XPC across many mammalian cell types (Appendix A, Figure 1A).

Remarkably, the overexpression the XPC complex led to a dramatic decrease in global 5mC when assayed by ELISA, methylated DNA immunoprecipitation (MeDIP), and dot blot (Figure 1C-D). Given that the XPC complex has an established function in nucleotide excision repair (NER), we next examined whether XPC DNA binding or repair activity was required. Using the overexpression of a DNA repair-deficient mutant of XPC, W690S, that impairs DNA binding and abolishes DNA repair activity (Bunick et al., 2006; Maillard et al., 2007; Yasuda et al., 2007), we found that its overexpression was also capable of reducing global 5mC levels. Although the W690S mutant led to a subtler phenotype compared to the wildtype XPC complex, it still suggests that the DNA repair activity of XPC is not required for its affect on DNA methylation. Furthermore, it is possible that the milder effect is due to the levels at which we were able to overexpress the W690S mutant compared to the WT XPC (Figure 1F), consistent with the missense mutation resulting in destabilization of XPC (Yasuda et al., 2007). Similar results to the wildtype complex were also seen for the overexpression of the XPC P334H substitution – the only known patient-derived mutation leading to developmental defects (e.g. neurological symptoms) (Bernardes de Jesus et al., 2008; Hananian and Cleaver, 1980) – and internal deletion of residues 338-519 (Appendix A, Figure 1B), which indicates that neither of these regions are likely to be involved in XPC's influence on global methylation.

Interestingly, the overexpression of the XPC subunit alone is sufficient to induce global demethylation similar to that of the heterotrimeric complex (Appendix A, Figure 1C, E). This may be in part due to the stabilization of endogenous RAD23B and CETN, which occurs when XPC is upregulated (Appendix A, Figure 1D). RAD23B-CETN2 overexpression had no effect alone, further suggesting that this phenomenon is dependent on the XPC subunit (Appendix A, Figure 1C). CFSE proliferation assays (Appendix A, Figure 2A) and metabolic-based MTS proliferation assays (data not shown) revealed no significant changes in doubling time or growth rate of the HDFs upon XPC overexpression, indicating that passive demethylation through rapid cell division is unlikely a major contributing factor. However, there was a noticeable

increase in cell size accompanied with XPC overexpression (Appendix A, Figure 2B-C). The significance of this observation is currently unknown.

In order to more directly assess the mechanism of XPC-regulated DNA demethylation, we took advantage of a minimal in vitro glycosylase assay, which measures the cleavage activity of TDG. Recombinant human TDG purified from bacteria is able to cleave a 5' fluorescently labeled doubled-stranded oligonucleotide containing a T:G mismatch with high efficiency, which was then visualized on a denaturing PAGE (Figure 2A). As anticipated, electrophoretic mobility shift assays (EMSA) using recombinant TDG also confirmed that TDG binds as a monomer until high molar ratios of TDG:DNA where it then binds as a dimer (Figure 2B) (Morgan et al., 2011). Using limiting amount of TDG, this system thus allows us to directly measure changes in TDG activity in the presence of XPC.

Purified recombinant human XPC complex has a dose-dependent enhancement of TDG activity over the 10-90 minute period measured (Figure 2C). The XPC-mediated stimulation of TDG activity was similar across all 5mC demethylation intermediates (T, 5hmU, 5fC, 5caC) (Figure 2D). Furthermore, the addition of the DNA repair-deficient XPC mutant, W690S, to the TDG glycosylase assays also stimulated base excision by TDG, thus supporting our bulk in vivo findings that the DNA repair activity of XPC is not required for its role in DNA demethylation (Figure 2D). However, because the inherent instability of the W690S complex may result in protein aggregation during the glycosylase assay (data not shown), it is not possible to determine if DNA repair or binding is responsible for at least some part of XPC's stimulation of TDG activity.

Functional characterization of the XPC-TDG interaction

Since the DNA binding and repair activity of XPC is not essential for its stimulation of TDG, we hypothesized that this phenomenon might be a result of a protein-protein interaction between XPC and TDG. To test this, we purified a series of N- and C-terminal truncations of XPC, as well as an internal truncation which removes a highly-disordered, mammalian-specific domain (Δ338-519 aa) from Sf9 cells (Figure 3A-B) (Bunick et al., 2006). When added to the in vitro glycosylase assay the ΔN (residues 195-940) and Δ338-519 truncations of XPC showed similar stimulation of TDG activity when compared to wildtype (Figure 3D). However, the C-terminal truncation of XPC, which contains the CENT2 interaction domain (Nishi et al., 2005; Popescu et al., 2003), failed to enhance TDG activity (Figure 3D), indicating an important function of the C-terminus of XPC in regulating TDG activity. It is also important to note that all of these truncated XPC complexes are highly active in OCT4/SOX2-dependent transcription (Y. Fong, unpublished data), so the loss of XPC-mediated stimulation with the C-terminal truncation of XPC is specific to its role in DNA demethylation.

The C-terminus of XPC (residues 814-940) has also been shown to bind TFIIH (Uchida et al., 2002; Yokoi et al., 2000). However, because our purified recombinant

XPC complex lacks TFIIH (data not shown), it is unlikely that TFIIH contributes to the observed XPC-dependent stimulation of TDG activity in vitro. To address whether the defect is due to loss of CETN2 binding to the C-terminally truncated XPC complex, we purified a full-length XPC-RAD23B heterodimer in Sf9 cells in the absence of human CETN2 (Figure 3C). As shown in Figure 3E, the full-length XPC-RAD23B heterodimer was sufficient in stimulating TDG activity (Figure 3E). These observations are consistent with our hypothesis that the C-terminal region of XPC is required for TDG interaction.

We next set out to determine the protein domain of TDG required for XPC-TDG interactions. Serial truncations of TDG that removed the N-terminal domain, the C-terminal domain, the SUMOylation motif, and the catalytic core domain (residues 111-308) were purified from bacteria to near homogeneity (Figure 4A; Appendix A, Figure 4A). These truncated TDG proteins retain in vitro cleavage activity using a physiologically relevant 5fC:G (Figure 4B, top) or 5acC:G DNA substrates (Figure 4B, bottom). However, XPC failed to stimulate the glycosylase activity of TDG lacking the first 111 amino acids. Because stimulation was still observed using the 51-410aa TDG truncation mutant, taken together, our results suggest that amino acids 51-111 may be critical for XPC-TDG interaction (Figure 4B; Appendix A, Figure 3C).

The residues 51-111 of TDG has been previously been shown to be required for repair of T:G mismatches (Gallinari and Jirincny, 1996). So not surprisingly, the 111-410aa truncation exhibited no enzymatic activity on T:G substrates (Appendix A, Figure 3C). However, it has not be previously determined what regions are required for the remaining DNA demethylation intermediates – particularly 5hmU, 5fC, and 5caC. It appears the 51-111aa region is essential for glycosylase activity on 5hmU:G substrates, but not for 5fC or 5caC substrates, albeit the activity is diminished. It is important to note that approximately a third of this 51-111aa stretch is composed of basic residues, which may be post-translationally modified (Smet-Nocca et al., 2008) and/or exhibit HMGA-like properties (Steinacher and Schär, 2005). Thus it has been previously proposed that this region may be crucial in not only regulating its role in BER, but in regulating its interaction with transcriptional machinery, including the coactivator CBP/p300 (Tini et al., 2002).

Despite the well-established link between TDG enzymatic turnover and SUMOylation (Hardeland et al., 2002), we failed to observe a significant role of SUMOylation in the functional interaction between XPC and TDG. Using an ex vivo SUMOylation in lysate scheme we modified from Weber et al., 2014, we purified SUMO-TDG to near homogeneity. However, SUMOylation of TDG does not appear to enhance the basal activity of TDG, or promote the ability of XPC to stimulate the excision of 5fC and 5caC substrates by TDG (Appendix A, Figure 3C). However, because SUMOylation is proposed to occur following TDG binding and cleavage, other techniques may be required to definitively address the role of post-translational modifications in DNA demethylation.

Genome-wide analyses reveal XPC-TDG bound at regulatory elements and promoters in ESCs

TDG knockout in mouse ESCs leads to the accumulation of 5fC and 5caC at proximal and distal regulatory elements (Shen et al., 2013), which are enriched in motifs for HIF1A, ESRRB, OCT4, and SOX2 (Lu et al., 2015), suggesting that TDG acts with core ESC-specific transcription factors to regulate DNA demethylation at their target genes. Similarly, RAD23B has been previously shown to overlap significantly with OCT4/SOX2 at regulatory elements in mouse ESCs (Cattoglio et al., 2015; Fong et al., 2011). Taken together, these data raise the possibility that XPC and TDG may colocalize in the mouse ESC genome. To evaluate the extent at which XPC and TDG interact in vivo, we compared a recently published Biotin-TDG ChIP-seq dataset (Neri et al., 2015) to an endogenous RAD23B ChIP-seq dataset previously published by our lab (Cattoglio et al., 2015; Fong et al., 2011).

Our analysis revealed a striking ~93% overlap between TDG and RAD23B peaks identified by MACS2 (Figure 5A). As we predicted, TDG and RAD23B co-occupy enhancer and promoter elements of many pluripotency genes, including *Nanog*, *Oct4,* and *Tcf3* (Figure 5B), and of housekeeping genes, such as *Actb* (Figure 5C). Although there is overwhelming evidence to suggest a role of XPC and TDG preferentially act at promoters and regulatory regions, it is still intriguing that two DNA repair proteins with no known sequence specificity would bind in such a manner. We speculate that the specific recruitment of TDG and XPC to gene promoters and enhancers is likely mediated by their interactions with sequence-specific transcription factors such as OCT4 and SOX2 (Cattoglio et al., 2015; Fong et al., 2011).

To test our hypothesis that DNA demethylation mediated by TDG-XPC occurs at specific loci on a genome-wide scale, we performed MeDIP followed by high-throughput sequencing (MeDIP-seq). HDFs were transduced with control (mCherry) or XPC-expressing lentiviruses and were selected for integration (Figure 6A). Following selection and expansion of the HDFs, we induced reprogramming via episomal vectors containing cDNAs for OCT4, SOX2, KLF4, L-MYC, LIN28A, GFP, and a shRNA against p53 (Okita et al., 2011). Sequencing libraries were made from either uninduced HDFs or FACS-sorted partially reprogrammed HDFs (7 days post OKSML expression, cotransfected with GFP to allow for positive sorting of transfected cells) (Appendix A, Figure 5A). To rule out any effect XPC overexpression may have on iPS induction, we measured the mRNA levels of XPC and Oct4. XPC mRNA levels were elevated in the XPC gain-of-function cells as expected, regardless of whether they were induced (GFP+), indicating that we were not selecting for cells that have higher or lower XPC expression levels. Oct4 expression levels also remained high in both the control and XPC GFP+ populations (Appendix A, Figure 5B), indicating that XPC overexpression did not significantly affect exogenous gene expression.

In agreement with our initial bulk methylation experiments (Figure 1C-E), XPC overexpression results in a loss of DNA methylation in both HDFs and partially reprogrammed iPSCs. In XPC overexpressing cells, not only does DNA methylation occur in fewer regions compared to WT cells (~30,000 vs. ~205,000 in HDFs and

~85,500 vs. 230,000 in iPSCs) (Figure 6B, bottom panel), but also methylation levels are lower than in WT cells, as shown by reduced MeDIP enrichment over the background (Figure 6B, top panel). Moreover, when we analyzed the reads with relation to their distance to the closest TSS, we observed reduced MeDIP reads at nearly all regions +/- 5kb from the TSS, with exception of the proximal promoter (+/- 250bp from TSS) (Figure 6C). Intriguingly, the MeDIP-fragment coverage of uninduced, XPC-overexpressing HDFs is quite comparable to the coverage of both control and XPC-overexpressing pre-iPSCs, suggesting that the methylation levels have already been reduced to that of pre-iPSCs upon overexpression of XPC in HDFs.

Recent methylome studies have reported dynamic methylation changes occurring preferentially at promoters and regulatory elements of silent and poised genes (Lu et al., 2015; Shen et al., 2013).. Thus, we proposed that sequence-specific recruitment of TDG and XPC to gene promoters and enhancers (Figure 5) may drive local DNA demethylation at these elements, which are largely under the regulation of sequence-specific transcription factors. Because XPC has been previously shown to co-occupy promoters and enhancers with OCT4 and SOX2 (Cattoglio et al., 2015; Fong et al., 2011), we asked if methylation at OCT4/SOX2 binding motifs would be decreased in XPC gain-of-function pre-iPSCs compared to control pre-iPSCs. Contrary to what we expected, motif analysis did not identify any OCT4/SOX2 binding sites as enriched among regions with reduced or abolished methylation upon XPC overexpression. The result may be due to the fact we overexpress XPC prior to induction rather than simultaneously, contributing both potential OCT4/SOX2 recruited demethylation as well as non-OCT4/SOX2 driven demethylation. Motif analysis, however, did identify ZFX and SP1 binding sites.

Transcription factor binding, and subsequent transcriptional activity of the target gene, has been previously implicated to maintain promoter and CpG islands as methylation-free (Turker, 1999). ZFX is an X-linked transcription factor required for ESC self-renewal and regulates the expression of *Tbx3* and *Tcl1* (Galan-Caridad et al., 2007), both of which are important regulators of ESC maintenance (Ivanova et al., 2006). Similarly, SP1 is a transcription factor essential for embryonic development (Marin et al., 1997) and plays an important role in potentiating *Nanog* transcription in mouse embryonal carcinoma cell line F9 (Wu and Yao, 2006). It has been previously shown that the removal of SP1 binding can result in locus-specific promoter methylation (Brandeis et al., 1994; Macleod et al., 1994), while its introduction near transgene and retroviral elements can promote demethylation (Hejnar et al., 2001; Siegfried et al., 1999). Taken together, the data suggest the importance of transcription factor binding sites in areas of XPC-mediated demethylation during reprogramming, though further studies are needed to understand how and why these motifs are being preferentially demethylated upon XPC overexpression.

XPC enhances human iPSC generation

We next sought out to determine the role XPC may play in somatic cell reprogramming. Given the critical role of the XPC complex in pluripotent gene activation, it is not surprising that both mouse and human reprogramming is diminished upon XPC depletion (Appendix A, Figure 7A-C) (Fong et al., 2011). A modest but reproducible reduction in the number of human iPS colonies was obtained with XPC-knockdown HDFs, which is likely due to inefficient depletion of XPC mRNAs in HDFs, compared to the near 100% knockdown in MEFs (Appendix A, Figure 7D). Surprisingly, reprogramming of XPC loss-of-function cells halts prior to reactivation of endogenous *Nanog*, *Oct4*, and other pluripotent genes required to sustain reprogramming (Stadtfeld et al., 2008), suggesting a potential role of XPC in reprogramming initiation in addition to its established role in stem cell maintenance (Fong et al., 2011).

XPC expression is significantly lower in somatic cells compared to ESCs (data not shown; Fong et al., 2011). To address whether low endogenous levels of XPC in somatic cells may act as a barrier or limiting factor in reprogramming, we set out to examine the effect of XPC overexpression on iPSC conversion. We transduced HDFs with control (mCherry) or XPC-expressing lentiviruses. Following selection and expansion of the HDFs, we induced reprogramming via episomal vectors containing cDNAs for OCT4, SOX2, KLF4, LIN28A, L-MYC, and a shRNA against p53 (Figure 7A) (Okita et al., 2011). Flow cytometry analysis revealed that reprogrammed XPC gain-of-function cells had a higher proportion of iPSCs expressing TRA-1-60, a marker of late stage mature human iPSCs, 24 days post induction (Figure 7B).

It is worth noting that the number of iPSC-like colonies obtained from control and XPC overexpressing cells remains the same (Figure 7C). Taken together, these results suggest that XPC enhances TRA-1-60 expression not by increasing the number of iPSCs, but by facilitating the derivation of more robust iPSCs that more closely resemble bona fide ESCs. Human PSCs readily undergo apoptosis and differentiation upon single cell dissociation, in part due to the metastable state of 'primed' epiblast-like human PSCs (Ohgushi and Sasai, 2011). To examine the self-renewal capacity of both control and XPC gain-of-function iPSCs, iPSCs were dissociated to single cells and subjected to colony forming assays, scored by the number of iPSC colonies formed that stain positive for alkaline phosphatase (AP), an marker of undifferentiated PSCs. iPSCs derived from the XPC gain-of-function HDFs resulted in 2.7-fold more AP+ colonies following single cell dissociation compared to the control (Figure 7D). This was also seen whether or not the cells were pre-treated with a selective Rho-associated kinase inhibitor, Y-27632, which inhibits dissociation-induced apoptosis (Watanabe et al., 2007), suggesting that the increase in cell survival in XPC cells is unlikely to act in the same pathway (Appendix A, Figure 7B). RT-qPCR analyses on bulk iPSCs obtained 30 days post induction showed a mild but consistent increase in the expression of key pluripotent genes such as Nanog, Oct4, Sox2, and Rex1 in XPC gain-of-function versus control iPSCs suggesting that these XPC iPSCs may have a slight advantage at self-renewal when they are in a metastable state (Figure 7E), which may explain their increased colony forming potential as well (Figure 7D). Furthermore, we found that XPC overexpressing cells display higher levels of both early and late iPSC markers,

SSEA-4 (data not shown) and TRA-1-60 (Appendix A, Figure 7A) respectively, as early as 7 dpi, suggesting that its affect on reprogramming occurs relatively early. Overexpression of various mutant forms of XPC, including the DNA repair-deficient mutant, has varying consequences on reprogramming efficiency (Appendix A, Figure 7C), but further studies are required to understand how the molecular mechanism underlying the results.

Recently, a privileged group of ultrafast cycling cells has been described to drive nonstochastic reprogramming from hematopoietic progenitors and MEFs (Guo et al., 2014c). Indeed, these ultrafast cycling somatic cells were shown to give rise to the bulk of iPSCs, though the precise mechanism that underpins this remains unknown. Characterization of these cells reveals that these cells cycle much faster than their counterparts and express elevated levels of genes associated with DNA repair, RNA processing, and cell cycle control (Guo et al., 2014c), presumably to provide the cellular components necessary for rapid proliferation to overcome some major bottlenecks in reprogramming (Banito et al., 2009; Hanna et al., 2009; Hong et al., 2009; Ruiz et al., 2011; Utikal et al., 2009). Given the importance of XPC in reprogramming, we asked if XPC, and possibly TDG, may be elevated in and/or contribute to this 'privileged' somatic cell state. We first labeled induced MEFs with CFSE, a fluorescent dye that is cell permeable but covalently bonds with intracellular macromolecules, and allowed them to proliferate for an additional 48 hours. CFSE intensity is diluted with each cell division, allowing separation of rapid and slow dividing cell populations (Lyons and Parish, 1994; Weston and Parish, 1990). Slower dividing populations will have gone through less cell divisions and thus retain much higher concentrations of CFSE. From the variegated dye concentrations, we were able to separate four distinct cell populations by FACS that had undergone various numbers of cell divisions (Appendix A, Figure 8A). Strikingly, ultrafast cycling cells (Lo) expressed the highest levels of *Xpc* and *Tdg*, reaching levels similar to that of ES cells for the latter (Appendix A, Figure 8B). These cells are likely pre-iPSCs as they have lost their mesenchymal identity (*Slug*, *Snail*, *Zeb1*) and initiated the transition into cells of epithelial origin (*Ecad*, *Epcam*) (data not shown). Taken together, these results suggest that high levels of XPC and TDG may be required to induce this reprogramming-permissive state in somatic cells. Interestingly, *Tet1*, *Dnmt1*, and particularly *Dnmt3b*, which all regulate DNA methylation and are enriched in ESCs, were all upregulated in the fastest dividing population (Appendix A, Figure 8C). By contrast, Tet2 expression remained relatively similar among all the sorted populations (Lo-Hi).

Given that both XPC and TDG are elevated in pre-iPSCs and interact with one another, as shown above, we next asked whether XPC enhances reprogramming through its interaction with TDG. Underscoring the importance of TDG in reprogramming, the shRNA-mediated knockdown of TDG abolishes reprogramming in MEFs (Appendix A, Figure 9A), which is consistent with recently published data (Hu et al., 2014) and similar to phenotype we observe with XPC loss-of-function. To address if TDG could further enhance the XPC gain-of-function phenotype in reprogramming, we overexpressed TDG alone or concomitantly with XPC in HDFs. Unlike XPC gain-of-function, TDG gain-of-function does not significantly change the percentage of iPSCs that expresses early and late stage iPSC markers (Figure 8A), but rather increases the

number of resulting AP+ human iPSC colonies (Figure 8B). This would be consistent with TDG accelerating the rate of the mesenchymal-epithelial transition (MET) during reprogramming (Hu et al., 2013), allowing more cells to initiate the first stage(s) of reprogramming. So it appears that the reprogramming dynamics of XPC gain-of-function HDFs differ dramatically from that of TDG gain-of-function HDFs, even though XPC may at least in part be acting through TDG. In essence, while XPC overexpression results in increased "stemness" of iPSCs, TDG overexpression alone significantly increases reprogramming efficiency.

Given the two separate phenotypes the overexpression of XPC and TDG has on reprogramming, it is not surprising that the simultaneous overexpression of both leads to an increase in both reprogramming efficiency and fidelity. We observed a subtle, but consistent increase in the number of AP+ colonies as compared to the control or XPC alone populations (Figure 8C). When XPC and TDG are simultaneously overexpressed in MEFs, we observe not only the same increase in AP+ colonies formed (data not shown), but we indeed observe a striking increase in the Thy1.2-/SSEA-1+/EpCAM+ population (Appendix A, Figure 9B), suggesting a functional synergy between XPC and TDG. Interestingly, when we immunoblot for XPC and TDG in the HDFs, we can detect both modified and unmodified TDG, but at varying ratios in each sample (Figure 8D). However, the significance remains unknown and further studies would be required to test if this could affect reprogramming as well. Taken together, these data suggest that XPC and TDG may be playing complementary roles in reprogramming, but further studies are required to determine the molecular mechanisms of how each XPC and TDG together and alone can enhance reprogramming.

Discussion

Our results reveal an unexpected role of the XPC DNA repair complex in influencing the methylation landscape of somatic and pluripotent cells. Here, we've reported the ability of XPC to enhance demethylation both at a global level in vivo and through a functional biochemical interaction with TDG in vitro. ChIP-seq analyses further reveal XPC and TDG co-occupy promoters and regulatory elements extensively in ESCs and may play a role in loci-specific demethylation in pre-iPSCs. Lastly, overexpression of XPC during human somatic cell reprogramming results in the generation of more robust iPSCs, improving cell survival during single-cell dissociation challenges.

The XPC complex adds to a growing list of DNA repair proteins that play multi-disciplinary roles in stem cells and are obligate for efficient reprogramming. This includes proteins involved in homologous recombination (González et al., 2013), non-homologous end joining (Molina-Estevez et al., 2013), intrastrand crosslink repair (Muller et al., 2012), nucleotide excision repair (Fong et al., 2011), and base excision repair (Doege et al., 2012; Hu et al., 2014). The relationship between DNA repair and reprogramming is not entirely surprising given that genome integrity is especially crucial in ESCs, as they give rise to all tissues of the adult organism. Underscoring this importance, ESCs express higher levels of DNA repair proteins than somatic tissues and readily undergo apoptosis upon DNA damage (Aladjem et al., 1998; Fan et al., 2011; Tichy and Stambrook, 2008). DNA repair proteins may also be critical in reprogramming to account for the DNA damage induced by reprogramming factors via replication stress associated with increased proliferation and transcriptional load (Helmrich et al., 2013). However, as tightly coupled pluripotency and DNA repair may be, it would be interesting to tease apart the divergent roles of XPC in DNA demethylation from DNA repair and transcription. This may be possible now that our findings reveal the C-terminus of XPC is required for XPC-mediated stimulation of TDG activity. Interestingly, the C-terminus of XPC is dispensable for transcription and is still highly active in potentiating OCT4/SOX2-dependent transcription in vitro (Fong et al., 2011). Therefore, future studies using the C-terminal truncation of XPC in reprogramming may be able to elucidate the extent DNA demethylation impacts reprogramming through XPC. The C-terminus of XPC is also conserved in the yeast Rad4 homolog. Rad4 is not active in OCT4/SOX2-dependent transcription (Zhang et al., 2015), but is capable of stimulating TDG glycosylase activity in vitro to a similar degree as human XPC (data not shown). Thus it is tempting to speculate that the role of XPC in demethylation may be evolutionary conserved, perhaps originally acting as an added measure of genome surveillance against DNA deamination or oxidation events.

Our findings provide strong evidence that XPC and TDG interact biochemically and genome-wide. Given the involvement of the XPC complex in transcription as a stem cell coactivator, XPC-mediated stimulation of TDG may also function to couple DNA demethylation and active transcription in ESCs. Indeed, transcription factor binding has been previously associated with the resistance to de novo methylation. For instance, SP1 binding is required upstream the mouse *Aprt* gene in order to maintain

the unmethylated state of the CpG island; removal of these binding sites leads its methylation (Brandeis et al., 1994; Macleod et al., 1994). NF-κB binding has also been implicated in B-cell specific demethylation of the Igκ locus (Kirillov et al., 1996). Conversely, the introduction of SP1 binding sites near transgene and retroviral elements can promote demethylation of those elements, which are normally silenced (Hejnar et al., 2001; Siegfried et al., 1999). It is unclear whether transcription factors can directly recruit DNA 'demethylases', sterically block DNMTs from maintaining methylation at a given region, or modulate methylation through other mechanisms. However, this suggests DNA methylation must be actively and dynamically maintained even near transcriptionally active genes and may be tightly co-regulated with transcriptional activity. Further characterization of the motifs preferentially demethylated upon XPC overexpression may provide insight as to whether or not XPC and TDG may be coordinating DNA demethylation and transcriptional activity in a loci-specific manner.

In the context of reprogramming, the roles of XPC as a transcriptional coactivator, modulator of DNA methylation, and DNA damage sensor may work hand-in-hand to coordinate robust gene expression and the maintenance of transcriptional and epigenetic signatures that unique to ESCs. ESCs in particular must maintain high genomic integrity during development in order for proper cell fate specification to occur. It is possible that DNA damage could recruit the XPC complex to sites of damage, limiting the amount of XPC and effectively halting pluripotent gene transcription (Figure 9). Regions of high transcriptional activity may also recruit higher levels of TDG activity, which in turn prevents spurious de novo methylation. Upon differentiation or downregulation of the XPC complex, pluripotent gene transcription is silenced and de novo methylation can occur. The XPC complex may therefore act as a molecular bridge to couple efficient DNA demethylation and robust ESC-specific transcription through the recruitment of TDG activity to sites bound by XPC and core ESC transcription factors, which is abolished in the presence of genotoxic stress.

Figures

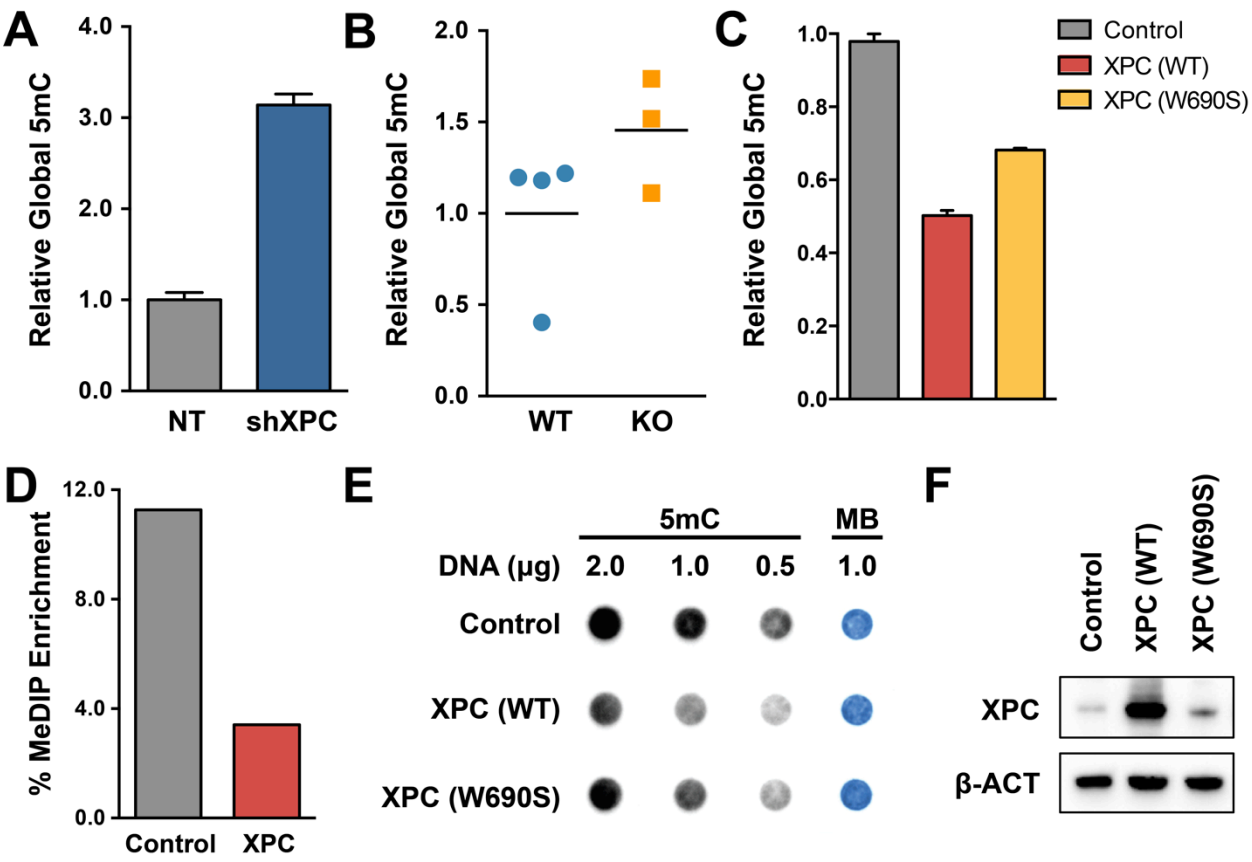

Figure 1. Global methylation is inversely correlated with XPC expression, independent of DNA repair activity. Relative global methylation was assayed by 5mC-specific ELISA using genomic DNA from (**A**) XPC knockdown (shXPC) in human dermal fibroblasts (HDFs), (**B**) Cas9-mediated knockout of XPC in H9 hESCs, and (**C**) HDFs overexpressing wildtype (WT) or DNA repair-deficient (W690S) human XPC. Relative global methylation was also assayed by (**D**) methylated DNA immunoprecipitation (MeDIP) enrichment and (**E**) 5mC dot blot. Methylene blue (MB) staining was used to control for total DNA transferred to membrane. (**F**) Immunoblot analysis for HDF samples depicted in E. Error bars depict the standard deviation.

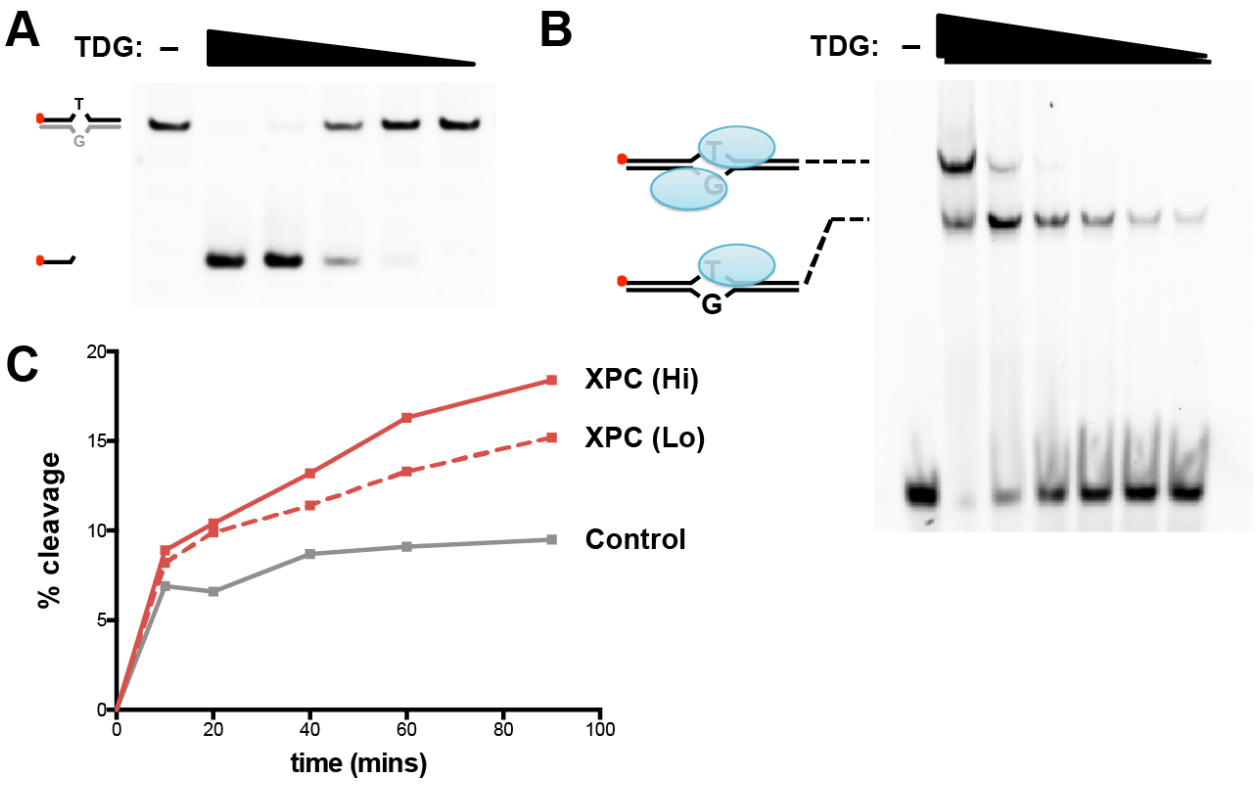
A
TDG: –
T
G
B
TDG: –
T
G
T
G
C
% cleavage
time (mins)
XPC (Hi)
XPC (Lo)
Control
D
XPC: –
WT
W690S
% cleavage
T:G
5hmU:G
5fC:G
5caC:G

Figure 2. The XPC DNA repair complex stimulates the activity of TDG. (**A**) TDG activity was determined by in vitro glycosylase assays. Decreasing amounts of purified, recombinant TDG was added to a fluorescently labeled 37mer dsDNA substrate containing a T:G mismatch. Efficient cleavage is detected by the presence of a 20mer band, corresponding to the fragment upstream of the mismatch, when run on a denaturing PAGE. The complementary strand is marked in grey to represent the substrate used, but is not visualized on the gel. (**B**) DNA-binding of TDG analyzed by EMSA on labeled 37mer dsDNA containing a T:G mismatch. Products were run on a 5% native PAGE. (**C**) Percentage of cleaved 5'-labeled oligonucleotide by TDG is graphed over a 10-90 minute time course. Solid red and dashed red lines denote reactions containing XPC in high (4.0 pmol) or low (2.0 pmol) concentrations, respectively. (**D**) Relative TDG activity was determined by in vitro glycosylase assays in the presence or absence of decreasing amounts of wildtype (WT) or repair-deficient (W690S) XPC. Graphs indicate amount of cleavage as total percentage of labeled substrate. The 37mer dsDNA TDG substrate tested is designated under the graph. Error bars represent the standard deviation, n = 3. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, calculated by 2-way ANOVA.

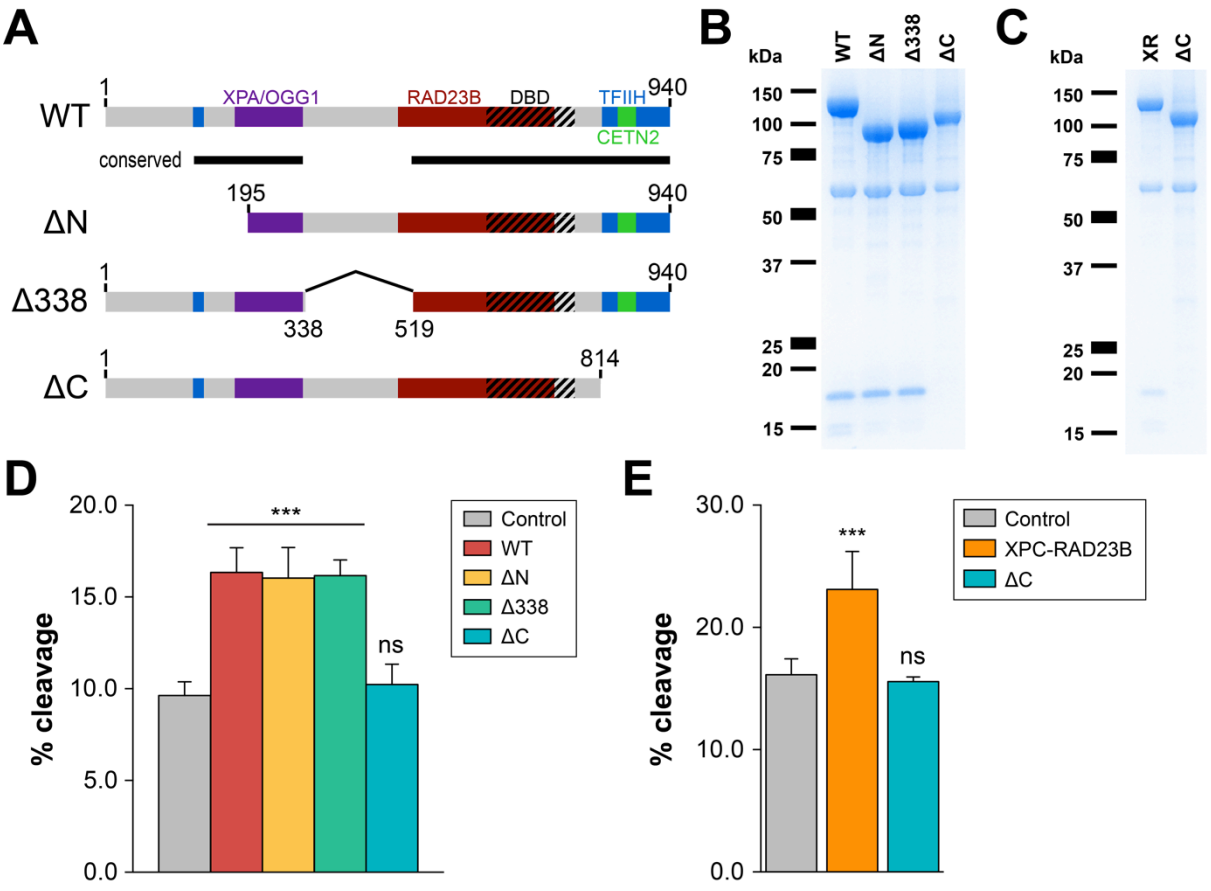

Figure 3. C-terminal deletion of XPC abolishes XPC-mediated stimulation of TDG activity. (**A**) Schematic representation of human XPC (WT) and relevant DNA binding (DBD) and protein-interaction domains. XPC truncations used in this study – ΔN (195-940aa), Δ338 (del338-519aa), and ΔC (1-814aa) – are also shown. (**B**) Recombinant human XPC complexes were expressed in Sf9 cells, purified to near homogeneity, and visualized by PageBlue. The wildtype XPC subunit runs at ~120 kDa, RAD23B at ~60 kDa, and CETN2 at ~17 kDa. (**C**) The XPC-RAD23B (XR) heterodimer was expressed in Sf9 cells, purified to near homogeneity, and visualized by PageBlue. (**D**) Cleavage of 37mer dsDNA T:G by recombinant TDG with or without the addition of wildtype XPC (WT) or truncations (ΔN, Δ338, and ΔC). (**E**) TDG activity was assayed with the addition of XR heterodimer or ΔC. Error bars represent the standard deviation, n = 3. *** $p < 0.001$, n.s. = non-significant, calculated by 2-way ANOVA.

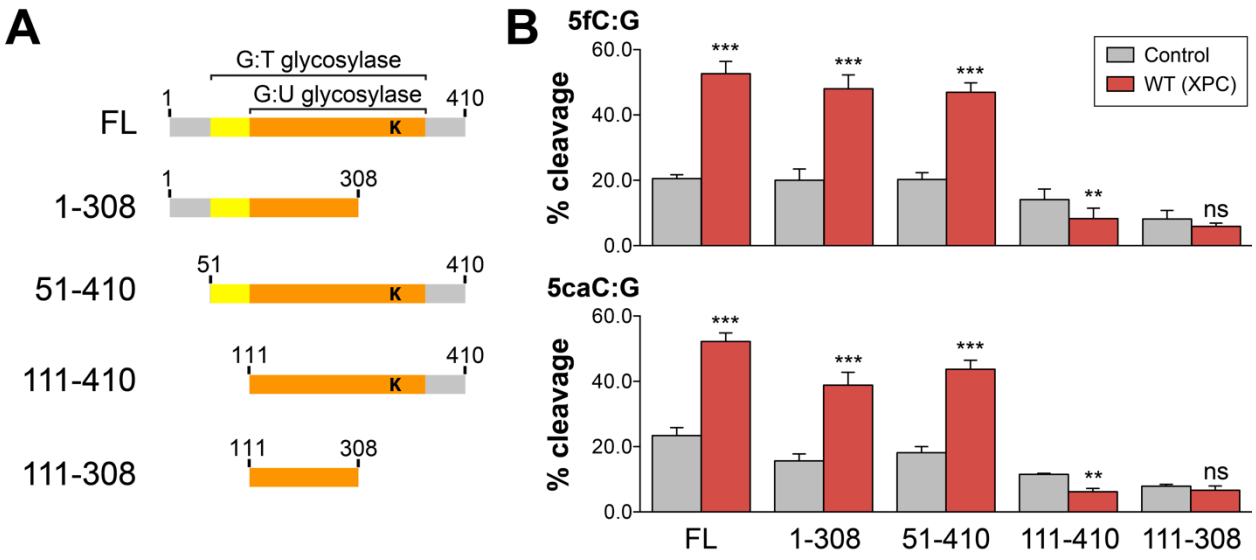

Figure 4. N-terminal deletion of TDG abolishes XPC-mediated stimulation of TDG activity. (**A**) Schematic representation of full length human TDG (FL) and truncations. The domains previously shown to be essential for G:U and G:T mismatch repair are noted (Gallinari and Jirincny, 1996); 'K' denotes the SUMOylation modification site (Hardeland et al., 2002). (**B**) Relative glycosylase activity of TDG truncations in the presence or absence of XPC (WT) was assayed on the 5fC:G (top) and 5caC:G (bottom) substrates. Error bars represent the standard deviation, n = 3. ** $p < 0.01$, n.s. = non-significant, calculated by 2-way ANOVA.

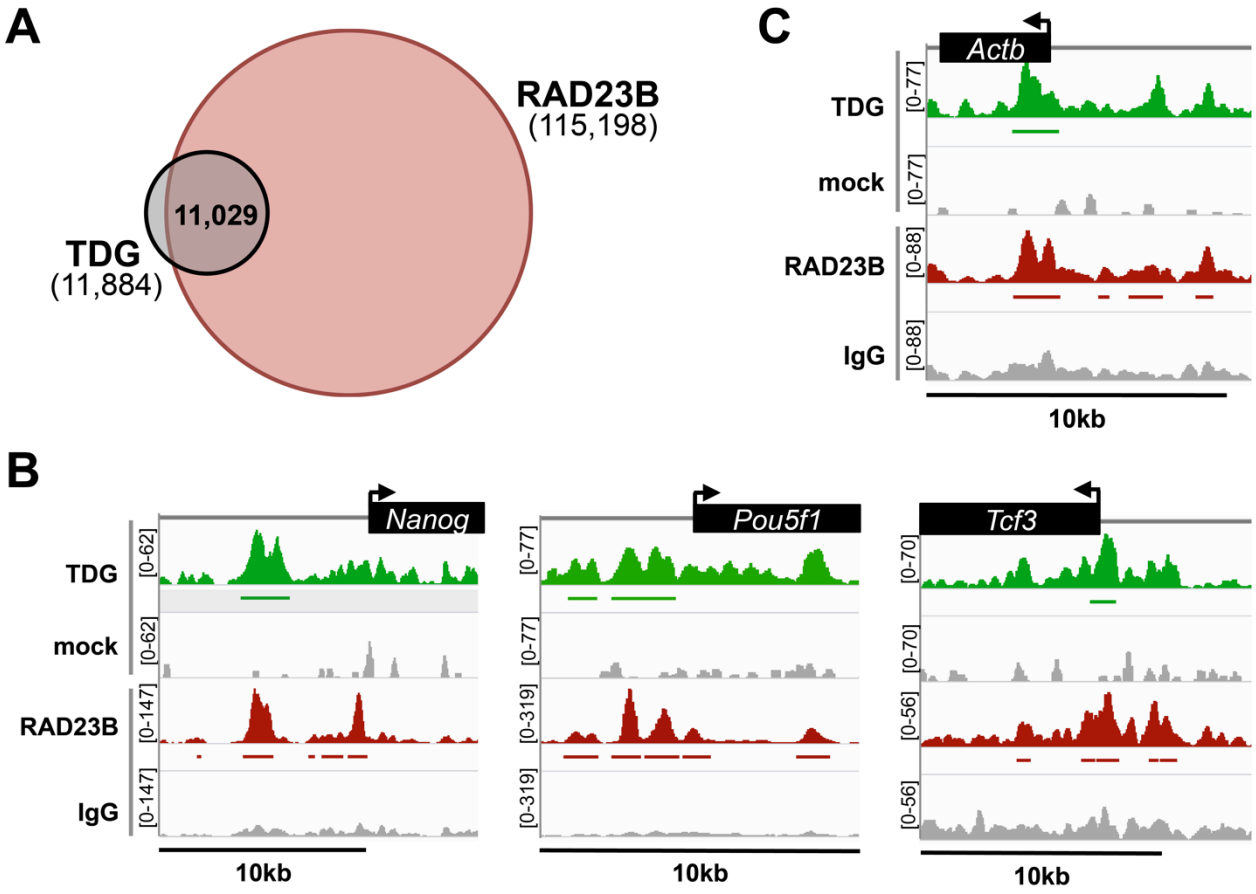

Figure 5. TDG is enriched at enhancers and promoters genome-wide and colocalizes with RAD23B. (**A**) Approx. 93% Biotin-TDG (Neri et al., 2015) ChIP-seq peaks overlap with RAD23B in mouse ESCs. Parentheses indicate number of peaks identified using MACS2 for each dataset. (**B-C**): IGV-computed ChIP-seq tracks are plotted as (number of reads) × [1,000,000/(total read count)] for pluripotency-related genes *Nanog*, *Pou5f1*, and *Tcf3,* and housekeeping gene *Actb.* Mock treated and normal IgG were used as specificity controls for the TDG and RAD23B ChIP, respectively.

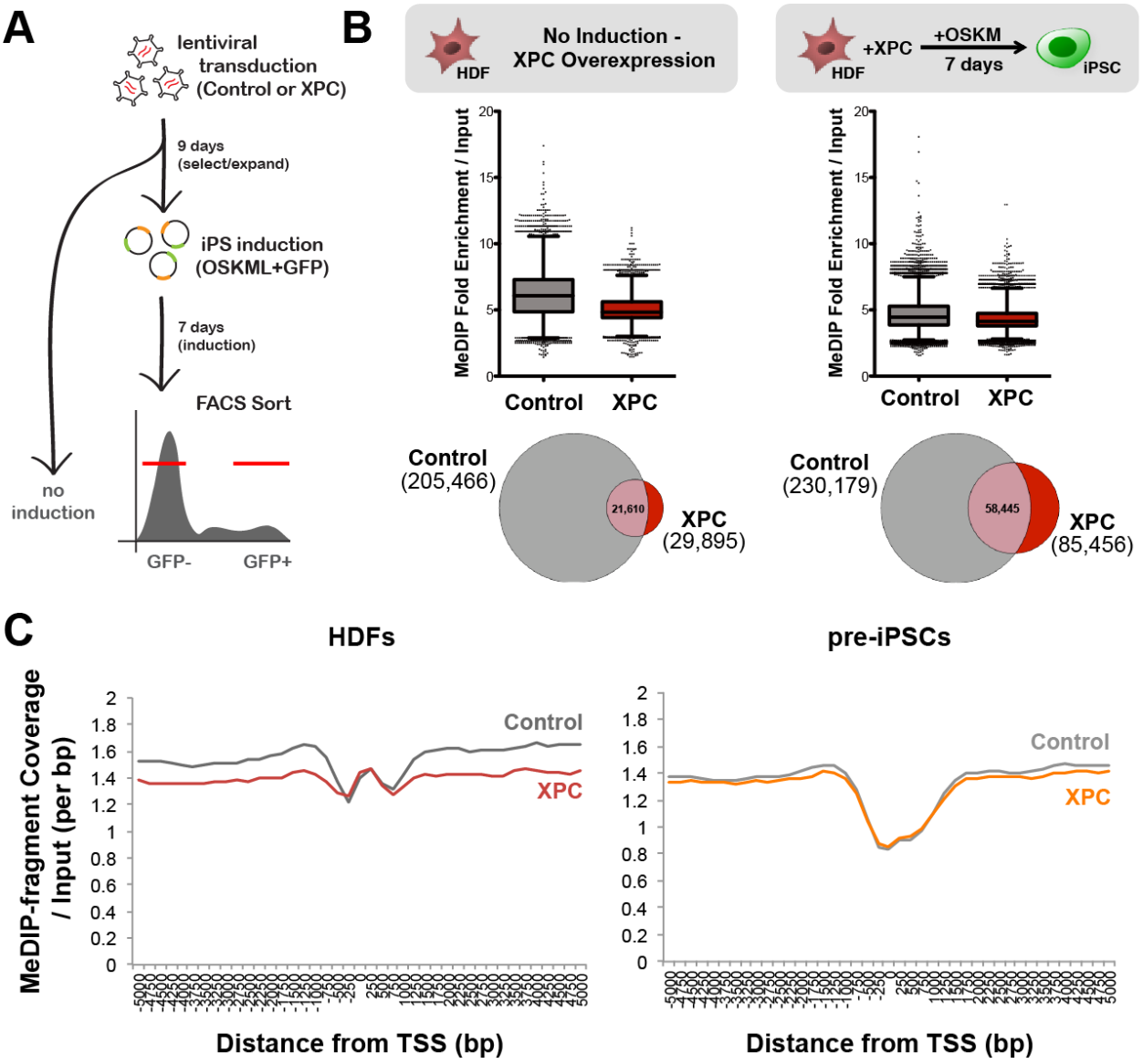
A
lentiviral transduction (Control or XPC)
9 days (select/expand)
iPS induction (OSKML+GFP)
7 days (induction)
FACS Sort
no induction
GFP-
GFP+
B
HDF
No Induction - XPC Overexpression
HDF
+XPC
+OSKM
7 days
iPSC
MeDIP Fold Enrichment / Input
Control
XPC
Control (205,466)
21,610
XPC (29,895)
Control (230,179)
58,445
XPC (85,456)
C
HDFs
pre-iPSCs
MeDIP-fragment Coverage / Input (per bp)
Control
XPC
Distance from TSS (bp)
D
ZFX
p-value 0.0003856
E-value 0.0563108
q-value 0.112025
SP1
p-value 0.001603
E-value 0.234038
q-value 0.468076

Figure 6. XPC overexpression results in lower MeDIP-seq enrichment in both HDFs and pre-iPSCs. (**A**) Design of MeDIP-seq experiments. Reprogramming factors are cotransfected with a GFP-expressing plasmid under the same promoter to allow for sorting of transfected cells. (**B**) XPC overexpression results in lower enrichment over background and fewer peaks called in MeDIP-seq analyses for both uninduced HDFs (left) and pre-iPSCs (right, 7 days post induction). Parentheses indicate number of peaks identified using MACS2. (**C**) Distribution of MeDIP-seq reads by their distance 5kb +/- from the TSS of RefSeq genes, normalized by input. (**D**) Motif discovery of pre-iPSCs MeDIP peaks present only in control and not in the XPC dataset.

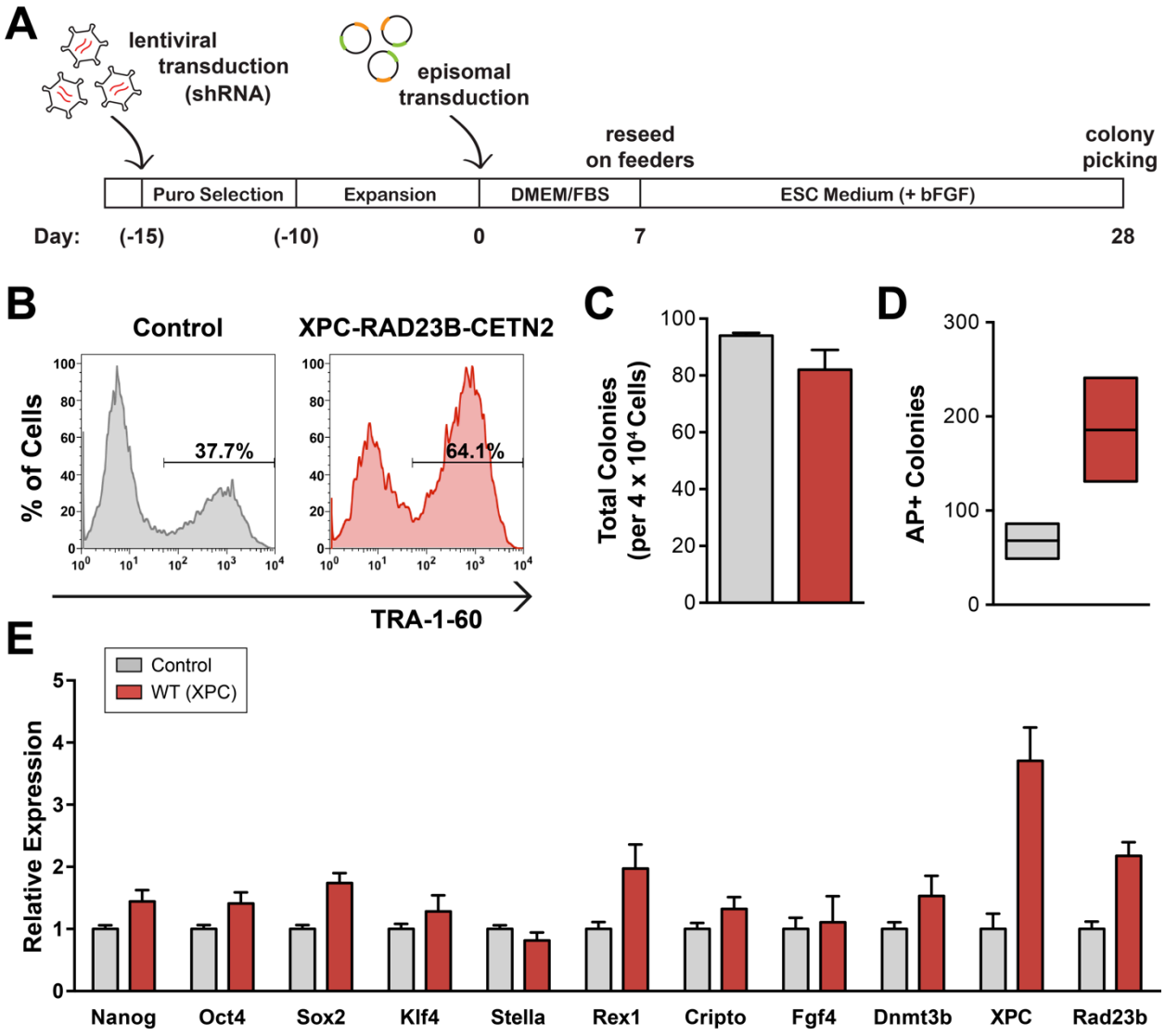

Figure 7. XPC increases reprogramming fidelity and survival from single-cell passaging. (**A**) Graphical scheme for human iPS reprogramming modified from Okita *et al*., 2011. (**B**) Flow cytometry analysis of the late stage human iPS marker, TRA-1-60, 24 days post induction (dpi) reveals a greater population of TRA-1-60+ cells with overexpression of XPC-RAD23B-CETN2 during reprogramming, compared to control. (**C**) Average number of colonies obtained in reprogramming experiment depicted in A. Error bars represent the standard deviation, n = 3. (**D**) iPSCs derived from control or XPC-RAD23B-CETN2 overexpression were challenged with single cell passaging and allowed to recover for 3-4 days before staining with alkaline phosphatase (AP). 2.5 x 10^4 cells were plated per 24-well, n = 3. (**E**) RT-qPCR analyses of pluripotent gene expression in bulk 30 dpi iPSCs resulting from the experiment shown in A. Error bars depict the standard deviation.

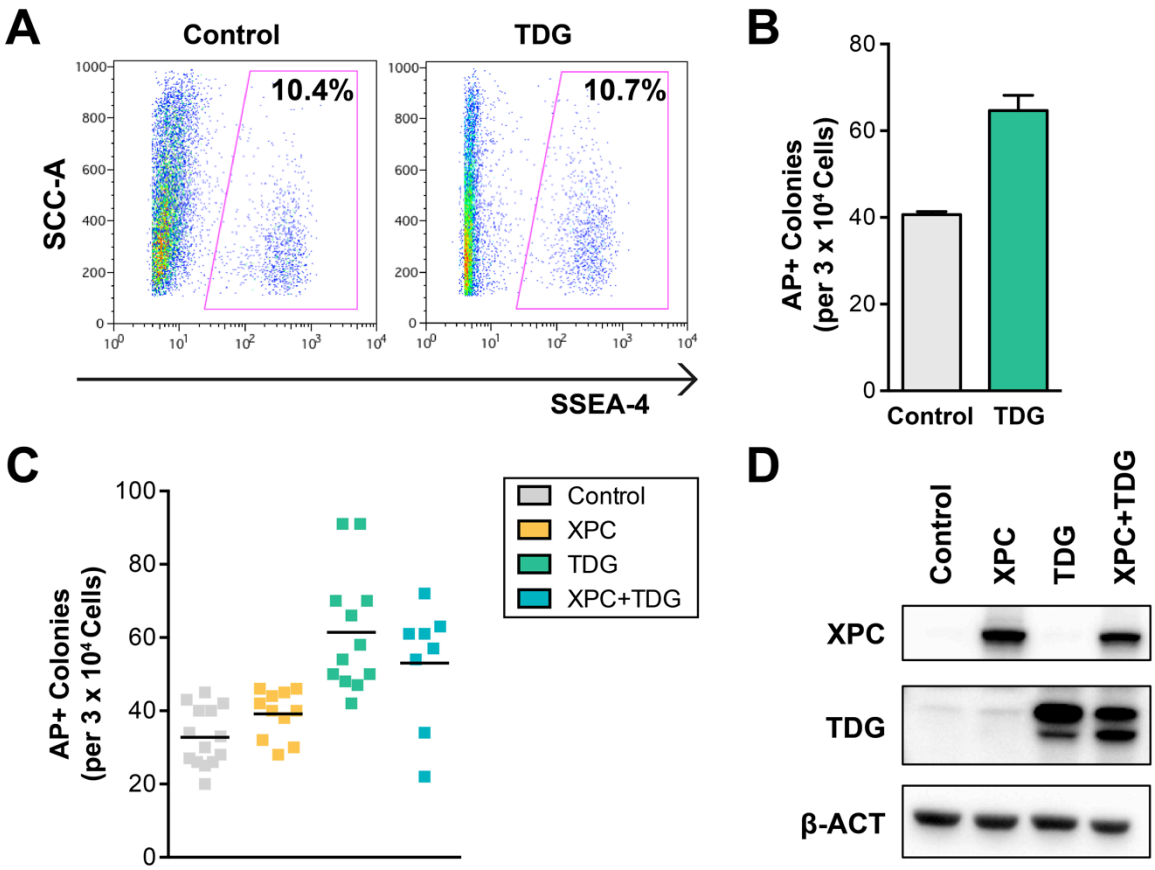

Figure 8. TDG overexpression increases human iPS reprogramming efficiency. (**A**) Flow cytometry analysis of the early stage human iPS marker, SSEA-4, 14 days post induction (dpi) with TDG overexpression, compared to control. (**B**) Average number of colonies obtained in reprogramming experiment depicted in A, 28 dpi. Error bars represent the standard deviation, n = 3. (**C**) Average number of colonies obtained with HDFs overexpressing XPC, TDG, or both XPC and TDG simultaneously. Error bars represent the standard deviations, n = 8 - 12. (**D**) Immunoblot analysis of HDFs for XPC and TDG. ACTB is used as a loading control.

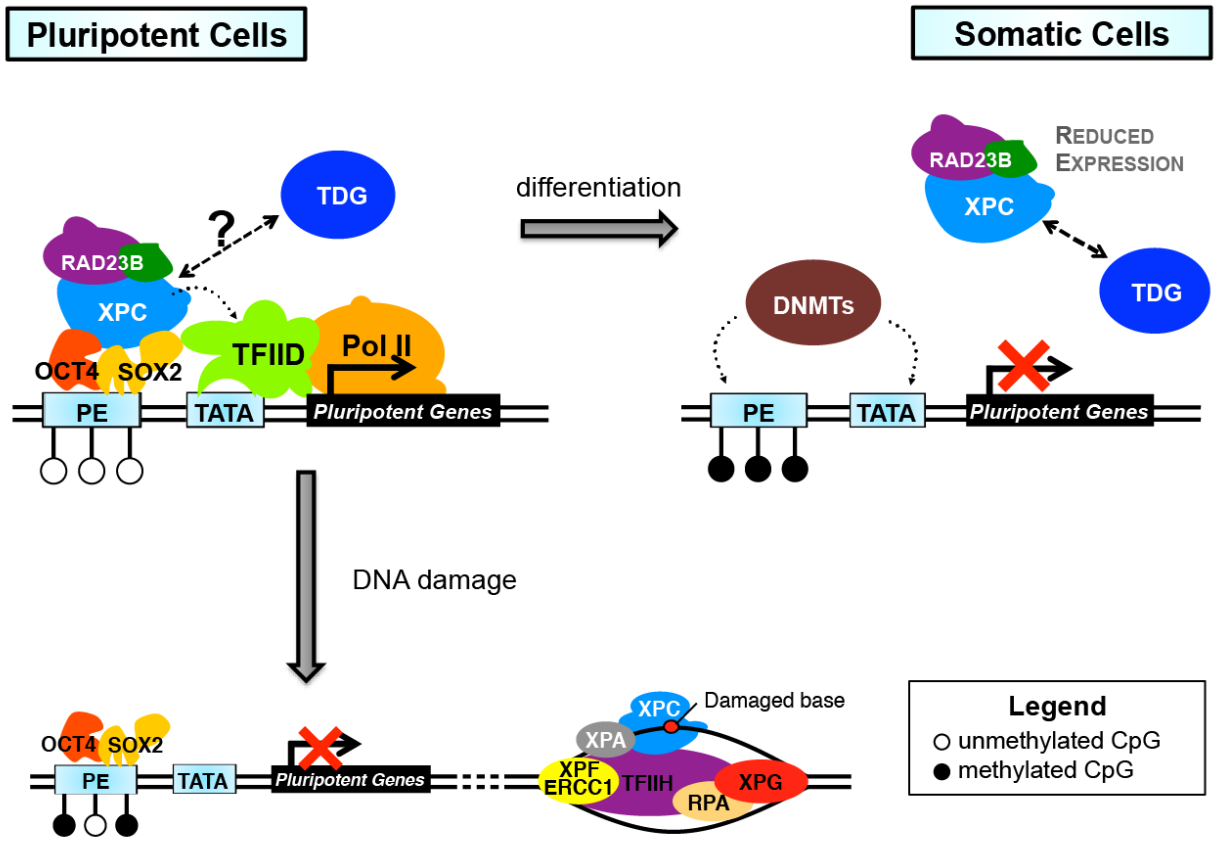

Figure 9. Working model: pluripotency, methylation, and genome integrity – a fine balance. The XPC complex potentiates OCT4/SOX2-dependent transcription of pluripotent genes in PSCs, which may recruit higher levels of TDG activity, preventing spurious de novo methylation. Pluripotency is compromised upon differentiation or downregulation of XPC. Upon DNA damage, XPC is recruited to damage sites, potentially compromising pluripotent gene transcription and allowing for de novo methylation.

Materials and Methods

DNA constructs and antibodies

cDNAs for human XPC and TDG were generated from total RNAs isolated from human NTERA-2 (NT2) cells. Mammalian expression plasmids were cloned using the pHAGE-EF1α-STEMCCA construct (Sommer et al., 2009), wherein OCT4, KLF4, SOX2, and c-MYC were replaced with cDNAs for XPC, RAD23B and CETN2, TDG, or mCherry (pHAGE-EF1α-XPC, pHAGE-EF1α-RAD23B-CETN2, pHAGE-EF1α-TDG, and pHAGE-EF1α-mCherry, respectively). For expressing full length and truncated human TDG in *E. coli*, N-terminal His_6-tagged TDG was cloned into a pST44 polycistronic expression plasmid (Tan et al., 2005). Constructs for XPC expression in Sf9s were previously described (Fong et al., 2011). Antibody to immunoprecipitate mouse RAD23 was generated in guinea pigs (Cattoglio et al., 2015; Fong et al., 2011). Commercial antibodies were as follows: anti-ACTB (A2228) and anti-FLAG M2 (F3165) from Sigma-Aldrich; anti-XPC (A301-122A) and anti-RAD23B (A302-305A) from Bethyl Laboratories; anti-CETN2 (15977-1-AP) and anti-TDG (13370-1-AP) from ProteinTech; anti-COXIV (4850P) from Cell Signaling Technologies; anti-5mC (33D3) from Diagenode; anti-SSEA4 (clone MC-813-70), anti-TRA-1-60 (clone TRA-1-60-R), anti-SSEA1 (clone MC-480), and anti-CD90/Thy1.2 (clone 53-2.1) from Biolegends.

Cell culture

The human embryonal carcinoma NTERA-2 (NT2) cell line and mouse D3 ESC line was obtained from ATCC. Human dermal fibroblasts (HDF) was obtained from Lonza. NT2, HDFs, and 293T cells were cultured in DMEM high glucose with GlutaMAX (Life Technologies) supplemented with 10% fetal bovine serum (FBS; HyClone). Mouse D3 ESCs were cultured in knockout DMEM (Life Technologies) supplemented with 15% FBS (Hyclone), 2mM GlutaMAX (Life Technologies), non-essential amino acids (Life Technologies), 0.1mM 2-mercaptoethanol (Sigma Aldrich), and 1000 units of LIF (Millipore) on 0.1% gelatin in the absence of feeders. Human ES cell line H9 (WiCell, Madison, WI) was maintained in feeder-independent conditions, using Synthemax SC-II Substrate (Corning) and grown in mTeSR1 (Stemcell Technologies). Media was changed daily and cell cultures were passaged using Dispase (Stemcell Technologies), according to the manufacturer's protocol.

Mouse embryonic fibroblasts (MEFs) were prepared from E13.5 CF-1 embryos (Charles River) and cultured in DMEM high glucose with GlutaMAX (Life Technologies), supplemented with 10% FBS (Hyclone) non-essential amino acids (Life Technologies), and penicillin/streptomycin (Life Technologies). Inactivation of MEFs was accomplished using mitomycin C (Sigma Aldrich).

Overexpression and shRNA-mediated knockdown of XPC by lentiviral infection

For lentivirus production, mammalian expression plasmids [control (pHAGE-EF1α-mCherry), TDG (pHAGE-EF1α-TDG), and XPC complex (pHAGE-EF1α-XPC, pHAGE-EF1α-RAD23B-CETN2)] and pLKO plasmids targeting human XPC were co-transfected with packaging vectors into 293T cells using lipofectamine 2000 (Invitrogen). Supernatants were collected at 48 hrs and 72 hrs post transduction and viruses were pelleted by ultracentrifugation. Lentivirus titers were determined using the Lentivirus-Associated p24 ELISA Kit (Cell Biolabs). HDFs were infected at a MOI of 5 in the presence of 4 μg/ml polybrene (Millipore), and replaced with fresh medium without polybrene after 24 hr.

Quantification of global methylation

Genomic DNA was obtained was using the DNeasy Blood & Tissue Kit (Qiagen). Relative global methylation by 5mC-specific ELISA was determined using the MethylFlash Methylated DNA 5-mC Quantification Kit (Epigentek), according to the manufacturers' protocols. For 5mC dot blots, genomic DNA was denatured at 95°C for 5 min then quickly spun down, placed on ice, and neutralized with 0.1vol of 6.6M ammonium acetate. Denature DNA was spotted onto a Hybond-N+ memberane (GE Healthcare) and air-dried. The membrane was then UV cross-linked and blocked in 10% milk, 1% BSA, 0.1% Tween-20 in PBS. The membrane was incubated with anti-5mC antibody (clone 33D3, Diagenode) overnight at 4°C overnight. After extensive washing, the membrane was incubated with a HRP-conjugated anti-mouse IgG secondary antibody (Pierce) for 30 minutes at room temperature. The membrane was then washed three times and treated with the Western Lightning ECL+ detection system (Perkin Elmer). Methylene blue was used as a control for total DNA crosslinked to membrane.

Methylated DNA immunoprecipitation

Genomic DNA from HDFs (10 μg/sample) or pre-iPSCs (1.5 μg/sample) was sheared using the Covaris S2 Focused Ultrasonicator (10.0 duty power, 175 peak power, 200 cycles/burst, 430s in 130 μl AE Buffer). Size selection was performed using the Agencourt AMPure XP beads (Beckman Coulter) to obtain DNA fragments averaging 150-250bp in length. End repair was performed using the NEBNext End Repair Module (NEB) followed by A-tailing with Klenow (NEB). Annealed TruSeq adapters were ligated to the sample DNA and subjected to another round of size selection to remove adapter dimers (note: some batches of Illumina TruSeq adapters may contain methylated sequences that will interfere with the MeDIP). DNA was denatured at 95°C for 10 min then quickly spun down and placed on ice. An adapter blocking oligo (5'-AGATCGGAAGAGCGTC-3') was added to the sample to prevent adapters from annealing to each other and increasing background in the subsequent

immunoprecipitation (IP). Denatured sample DNA and anti-5mC antibody (clone 33D3, Diagenode) or mouse IgG (015-000-003, Jackson) were incubated together at 4°C, overnight in MeDIP buffer (1mM EDTA, 0.05% TritonX-100, in PBS). Input DNA was kept as a control and processed along side the samples. The IP was added to pre-cleared, equilibrated M-280 anti-mouse IgG Dynabeads (Life Technologies) for 2 hr at 4°C. The IPs were washes 6 times in MeDIP wash buffer and eluted from the beads with proteinase K. Following the IP, recovered ssDNA was converted to dsDNA with KAPA HiFi HotStart Polymerase (Kapa Biosystems) for 4 cycles. No sample DNA was recovered for IgG IPs, indicating very little background binding was occurring. Sample DNA was subsequently subjected to another round of size selection to obtain fragment sizes averaging 200-400bp in length suitable for sequencing. Libraries were amplified using KAPA HiFi HotStart Polymerase (Kapa Biosystems) for a final 10 cycles. Size, purity, and concentration of the libraries were checked by Agilent Technologies 2100 Bioanalyzer. Sequencing was performed at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant.

Expression and purification of recombinant XPC complexes

Recombinant baculovirus for the infection of Sf9 cells was generated using the Bac-to-Bac Baculovirus Expression System (Invitrogen). Baculoviruses were amplified twice in Sf9 cells. Sf9 suspension cultures (1 L, 10^6 cells/mL) were infected with wildtype, W690S, ΔN (aa 195-94), Δ338-519, or ΔC (aa 1-814) His_6-XPC baculoviruses along with baculoviruses for FLAG-RAD23B and untagged CETN2. Cultures were collected at 48 hrs post infection, washed three times with ice-cold PBS and lysed in 6 packed cell volume of high salt lysis buffer A (HSLB-A; 50 mM HEPES pH 7.9, 0.5 M KCl, 0.1% NP-40, 10% glycerol, 10 mM 2-mercaptoethanol, and protease inhibitors), and sonicated. Cleared lysates were supplemented with 10 mM imidazole and incubated with pre-equilibrated Ni-NTA resin (Qiagen) for 16 hr. Resin slurries were poured into gravity columns, washed, and bound XPC and associated proteins were eluted with HSLB-A supplemented with 0.25M imidazole. Eluted fractions were analyzed by SDS-PAGE followed by PageBlue staining. Peak XPC fractions were pooled and incubated with anti-FLAG (M2) agarose (Sigma Aldrich) for 3 hr, washed with HSLB-A, and re-equilibrated with elution buffer (0.3 M KCl, 50 mM HEPES, pH 7.9, 10% glycerol, 0.05% NP-40, 1 mM DTT, 1 mM benzamidine, and 0.5 mM PMSF). Bound XPC complexes were eluted with FLAG peptide (Sigma Aldrich) at 0.4 mg/ml.

Expression and purification of recombinant TDG and SUMO-TDG

For bacterial purification of recombinant TDG, pST44 expression plasmids were transformed into BL21-Codon Plus RIPL competent cells (Agilent). Cultures were induced at 18°C overnight with 0.5 mM IPTG. Cell pellets were lysed in high salt lysis buffer B (HSLB-B; 25 mM HEPES pH 7.9, 0.6 M NaCl, 0.6% TritonX-100, 0.05% NP-40,

10% glycerol, 10 mM 2-mercaptoethanol, and protease inhibitors) and lysozyme (0.5 mg/ml). Sonicated lysates were cleared by ultracentrifugation, supplemented with 10 mM imidazole, and incubated with Ni-NTA resin for 16 hr. Bound proteins were washed extensively with HSLB-B with 20 mM imidazole, equilibrated with 0.2 M NaCl HGN (25 mM HEPES, pH 7.9, 10% glycerol, 0.01% NP-40) with 20 mM imidazole, and eluted with 00.2 M NaCl HGN supplemented with 0.25 M imidazole. Eluted fractions were analyzed by SDS-PAGE followed by PageBlue staining. Peak TDG fractions were pooled and dialyzed to 0.1 M NaCl HGN. Dialyzed peak Ni-NTA fractions were applied to a Poros 20 HQ column (Applied Biosystems) and subjected to a linear gradient from 0.1 M to 0.6 M NaCl. Full length TDG eluted from the column at ~0.15M NaCl. Peak fractions from the Poros-HQ were pooled and dialyzed to 0.2 M NaCl HGN.

In lysate SUMOylation of TDG was performed as described (Weber et al., 2014). Following incubation of lysates containing pSUMO1, Ubc9, SAE1, SAE2, and His_6-TDG at 30°C for 3 hours, the bacterial lysate were treated as described above. SUMO-TDG eluted from the Poros-HQ at ~0.2M NaCl, allowing for separation from unmodified TDG.

In vitro glycosylase assay

5'-labeled Cy3 oligonucleotides were obtained containing a T, 5hmU, 5fC, or 5caC internal modification (Trilink Biotechnologies). Cy3-labeled oligonucleotides were annealed overnight with unlabeled oligonucleotides that were complementary to the upper strand except at the position of the modified nucleotide. The annealed substrates were then purified twice by sequential native PAGE. Glycosylase reactions were performed in 10 µl reactions containing 20mM HEPES, pH 7.9, 100 mM NaCl, 1 mM EDTA, 0.1% BSA, 0.01% NP-40, and 2 pmol DNA substrate. Reactions were incubated with purified TDG (0.5 pmol) at 30°C for 45 mins unless otherwise specified. For reactions containing XPC, purified XPC complexes (2.0-4.0 pmol) were added simultaneously with TDG to the reactions. Reactions were stopped 0.1vol of 1M NaOH and heated at 95°C for 5 min. Reactions were then neutralized with 0.1vol of 1M glacial acetic acid. 20 µl of loading buffer (95% deionized formamide, 1 mM EDTA, and bromophenol blue) were added and the reactions were heated at 95°C for an additional 3 min. Samples were immediately run on 15% denaturing PAGE containing 7M urea at 200V for 45-60 min. Cleavage products were visualized using the PharosFX Plus (Bio-Rad). Sequences of DNA substrates used in glycosylase assays are provided in the appendix.

RNA isolation, reverse transcription, and real time PCR analysis

Total RNA was extracted and purified using TRIzol reagent (Life Technologies), according to manufacturers' protocol. cDNA synthesis was performed with 1 µg of total RNA using iScript cDNA Synthesis Kit (Bio-Rad) and diluted 10-fold. Real time PCR analysis was carried out with SYBR Select Master Mix for CFX (Life Technologies)

using the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Gene specific primer sequences are provided in the appendix.

Human somatic cell reprogramming and flow cytometry

Human iPS reprogramming was induced by the transfection of episomal vectors containing cDNAs for OCT4, SOX2, KLF4, L-MYC, LIN28A, and a shRNA against p53 (Okita et al., 2011) into HDFs using nucleofection (Lonza), according to the manufacturers' protocol. Following nucleofection, cells were plated in 60mm^2 dishes and cultured for 7 days in HDF media. Cells were then trypsinized and plated (2-4 x 10^3 cells/6-well) on inactivated MEFs (2.5 x 10^5 cells/6-well) in human iPSC media containing Knockout DMEM/F12 (Life Technologies), 20% Knockout Serum Replacement (Life Technologies), 2mM GlutaMAX (Life Technologies), non-essential amino acids (Life Technologies), 0.1mM 2-mercaptoethanol (Sigma Aldrich), and 4 ng/mL bFGF (Life Technologies). Reprogramming was assayed by alkaline phosphatase staining (Millipore) or by flow cytometry analysis using anti-SSEA4 (Biolegends) and anti-TRA-1-60 (Biolegends) on a BD LSRFortessa, performed according to the manufacturers' protocols.

Mouse somatic cell reprogramming and flow cytometry

CF-1 MEFs (Charles River) were transduced with inducible STEMCCA and rtTA lentivirus-containing supernatants overnight in 8 µg/ml polybrene (Sigma Aldrich). Alternatively, MEFs isolated from mice carrying an integrated dox-inducible transgene expressing OCT4, KLF4, SOX2, and c-MYC (Jackson Laboratories) were also used. Doxycycline (Sigma Aldrich; 2 µg/ml) was supplemented to complete mouse ES cell media to induce expression of OKSM. Reprogramming was assayed by alkaline phosphatase staining (Millipore) or by flow cytometry analysis using anti-CD90.2/Thy1.2 (Biolegends) and anti-SSEA1 (Biolegends) on a BD LSRFortessa, performed according to the manufacturers' protocols.

CFSE labeling

To determine the doubling time, HDFs were labeled with CFSE-Violet (Life Technologies) at a working concentration of 5.0 µM, according to the manufacturers' protocol. Cells were analyzed for remaining fluorescence on a BD LSRFortessa every day for 1-5 days.

Induced MEFs were labeled with CFSE-Violet (Life Technologies) at a working concentration of 7.5 µM, as described in Guo et al., 2014. CFSE-labeled MEFs were sorted into distinct fast to slow dividing populations at the UC Berkeley Li Ka Shing Flow Cytometry Facility. MEFs cultured in the absence of doxycycline were used as controls.

5mC detection using thin layer chromatography (TLC)

Genomic DNA (2 µg/sample) was digested overnight at 37°C with 100U of MspI (NEB), followed by dephosphorylation with CIP (NEB) for 1 hr. DNA was purified with the QiaQuick Nucleotide Removal Kit (Qiagen). Samples were denatured by heating at 95°C for 10 min followed by radioactive end labeling with [γ-32P]-ATP (10 µCi, Perkin Elmer) and T4 PNK (NEB) for 2 hr at 37°C. Samples were heat inactivated and treated with 100U S1 Nuclease (Promega) for 3 hr at 37°C. Samples (2 µl) were spotted onto 20 x 20cm PEI cellulose F coated TLC plates (Millipore) and developed with isobutryic acid:water:ammonium hydroxide (66:20:2 v/v/v) until the front reached 1cm from the top, approximately 8 hours. Plates were dried, exposed to a phosphorimager screen (Kodak) for 2 hr, and subsequently on the PharosFX Plus (Bio-Rad).

CHAPTER THREE:

THE IDENTIFICATION AND CHARACTERIZATION OF NOVEL COACTIVATORS INVOLVED IN OCT4/SOX2-DEPENDENT TRANSCRIPTION IN EMBRYONIC STEM CELLS

Abstract

Acquisition of pluripotency is driven largely at the transcriptional level by activators OCT4, SOX2, and NANOG that must in turn cooperate with diverse coactivators to execute stem cell-specific gene expression programs. Using a biochemically defined in vitro transcription system that mediates OCT4/SOX2 and coactivator-dependent transcription of the *Nanog* gene, we report the purification and identification of the dyskerin (DKC1) ribonucleoprotein complex and ABCF1 as OCT4/SOX2-dependent stem cell coactivators. The DKC1 complex and ABCF1 occupy enhancers and regulate the expression of key pluripotency genes critical for self-renewal in embryonic stem cells. Depletion of DKC1 or ABCF1 in fibroblasts significantly decreased the efficiency of induced pluripotent stem (iPS) cell generation. This study thus reveals an unanticipated transcriptional role of both the DKC1 complex and ABCF1 in stem cell maintenance and somatic cell reprogramming.

Introduction

The acquisition of pluripotency in the epiblast, a transient population of cells with unrestricted developmental potential during early embryogenesis, is controlled by a core set of transcription factors that include OCT4, SOX2 and NANOG (Avilion et al., 2003; Chambers et al., 2003; Mitsui et al., 2003; Nichols et al., 1998; Silva et al., 2009). This undifferentiated, pristine stem state can be captured as embryonic stem (ES) cells (Brook and Gardner, 1997; Evans and Kaufman, 1981; Martin, 1981), regenerated from somatic cells by cell fusion and nuclear transfer (Yamanaka and Blau, 2010), or by the ectopic expression of defined transcription factors (Takahashi and Yamanaka, 2006; Yu et al., 2007). These reprogrammed pluripotent cells display a transcriptome that is highly similar to ES cells. Not surprisingly, OCT4, SOX2, and NANOG also play key roles in the maintenance of pluripotency in ES cells and its reacquisition in induced pluripotent stem (iPS) cells by targeting a common set of genes that underpin the pluripotent state (Boyer et al., 2005; Chen et al., 2008; Loh et al., 2006). However, execution of these complex stem cell-specific gene expression programs also require a growing list of co-regulators including enhancer binding transcription factors (KLF4 (Jiang et al., 2008), SALL4 (Wu et al., 2006; Zhang et al., 2006), ESRRB (Festuccia et al., 2012; Zhang et al., 2008)), coactivators (Mediator (Chia et al., 2010; Kagey et al., 2010), YAP (Lian et al., 2010), TAFs/TFIID (Fong et al., 2011; Liu et al., 2011; Pijnappel et al., 2013)), chromatin remodelers (esBAF (Ho et al., 2009)), and histone modifiers (p300/CBP (Chen et al., 2008), the trithorax histone methyltransferase (Ang et al., 2011)). Perhaps the involvement of this rather elaborate collection of cofactors arose from the need for ES cells to significantly expand their transcriptional repertoire in order to accommodate the wide range of gene expression responses governing self-renewal and the transition into diverse differentiated cell-types (Fong et al., 2012). Intriguingly, recent studies have implicated additional cofactors that have not been traditionally associated with transcriptional regulation, such as the XPC DNA repair complex (Fong et al., 2011), as well as microRNAs and long non-coding RNAs as part of the pluripotency regulatory network (Jia et al., 2013; Orkin and Hochedlinger, 2011; Wilusz et al., 2009).

Unconventional transcriptional coactivators like the XPC complex and YAP are often found to be multifunctional. For example, the XPC complex safeguards genome integrity of self-renewing stem cells as well as their differentiated progeny by scanning the genome for DNA damage and initiating excision repair (de Laat et al., 1999; Riedl et al., 2003; Sugasawa, 2011), while YAP controls the expansion of stem cells by sensing diffusible signals and external cues in the niche (Dupont et al., 2011; Lian et al., 2010; Mori et al., 2014; Schlegelmilch et al., 2011). It therefore seems reasonable to speculate that co-opting these protein complexes into performing gene regulatory functions may represent a prevalent evolutionary strategy that allows rapidly dividing stem cells to expand and enhance the pluripotency network while coping with the enormous pressure to maintain genome stability and cellular homeostasis. Indeed, coactivators like the XPC complex and YAP are highly enriched in ES and iPS cells, perhaps because they are performing double duty (Fong et al., 2011; Lian et al., 2010; Ramalho-Santos et al., 2003). Not surprisingly, depletion of these multifaceted complexes compromises

pluripotency gene expression, stem cell maintenance, and somatic cell reprogramming (Fong et al., 2011; Lian et al., 2010). Therefore, it appears that a critical threshold level of these coactivators may be required for a stem cell to maintain its pluripotency. Likewise, high levels of these cofactors may be necessary to establish an appropriate gene regulatory environment for a somatic cell to re-enter the cell cycle and become responsive to transcription factor-mediated reprogramming.

Somatic cell reprogramming by a small cadre of specific transcription factors is thought to be a stochastic and inefficient process where only a small fraction (0.1–1%) of somatic cells become iPS cells (Buganim et al., 2013). However, recent data suggests that the iPS induction is not entirely a random event, but may depend in part on cell-intrinsic determinants that are somehow restricted to a privileged subpopulation (Guo et al., 2014c). These privileged somatic cells exhibit ultrafast cell duplication and express higher levels of proteins involved in DNA repair, RNA processing, and cell cycle control (Guo et al., 2014c). It is thought that these enrichments are required to fuel the rapid cellular proliferation necessary to overcome some major bottlenecks in reprogramming (Banito et al., 2009; Hanna et al., 2009; Hong et al., 2009; Ruiz et al., 2011; Utikal et al., 2009). Another roadblock to cellular reprogramming is the requisite early reactivation of a robust transcriptional circuitry governed by OCT4 and SOX2 (Buganim et al., 2012). Although this process can be enhanced by a number of transcription factors, reprogramming efficiency remains stubbornly low. It seems likely, therefore, that some key components of reprogramming remain undiscovered and there is a need to better define the molecular mechanisms by which OCT4 and SOX2 activate a stem cell-specific transcriptional program in ES and iPS cells.

To directly screen in an unbiased manner for cofactor requirements that support OCT4 and SOX2 mediated activation, we developed an in vitro transcription assay that faithfully recapitulates OCT4/SOX2 and coactivator-dependent gene activation observed in ES cells using purified components to reconstitute the human transcriptional apparatus (Fong et al., 2011). Deploying this sensitive biochemical complementation assay, we recently detected two additional stem cell coactivators (SCC-A and -B) that, in concert with the XPC coactivator complex, co-dependently stimulate the transcriptional activation of the *Nanog* gene by OCT4 and SOX2. Here we report that SCC-A activity is delivered by a subset of the dyskerin ribonucleoprotein complexes (DKC1 RNPs), while the SCC-B activity is delivered by the ATP-binding cassette (ABC) subfamily F member 1, ABCF1. Furthermore, we combined promoter occupancy data with pluripotency gene expression profiles from loss-of-function studies to directly link the DKC1 complex and ABCF1 to transcriptional coactivator function in ES cells. Our studies unveil a previously unanticipated role of DKC1 and ABCF1 in regulating transcription.

Results

Purification and identification of Q0.3

We previously have shown an activity present in a partially purified protein fraction, Q0.3, that is required for the XPC coactivator complex to stimulate a full, synergistic activation of the human Nanog proximal promoter by OCT4 and SOX2 but is dispensable for basal or Sp1-activated transcription (Fong et al., 2011; Rodda et al., 2005). Q0.3 separated from the XPC complex at the Poros-HQ anion exchange chromatographic step (Figure 1A,B). Although Q0.3 appeared to migrate as a single activity on a size exclusion column with an apparent molecular mass (Mr) of ~500 kDa (Figure 1C), this coactivator activity splits again into two distinct chromatographic fractions on a Poros-Heparin (Poros-HE) cation exchanger. One cofactor, SCC-B, eluted at ~0.4 M KCl whereas the second activity, SCC-A, eluted at ~0.6 M KCl (Figure 1A,D). Taken together, it appears that at least three distinct stem cell coactivators (one being the XPC complex) are required to generate a full, OCT4/SOX2-dependent transcriptional response. Starting with nuclear extracts prepared from 400 L of a pluripotent embryonal carcinoma (EC) cell line NTERA-2 (NT2), we used the reconstituted transcription system supplemented with recombinant XPC complex, purified OCT4, SOX2, and a modified human Nanog template, to purify SCC-A over six successive chromatographic columns resulting in >30,000-fold increase in specific activity (Figure 1A).

To identify the polypeptides comprising the SCC-A and SCC-B activities, peak Poros-HE fractions were pooled, concentrated, and separated by SDS-PAGE. Tryptic digestion of the four excised gel bands by mass spectrometry analysis revealed SCC-A to be the dyskerin (DKC1) complex comprised of DKC1, GAR1, NHP2, and NOP10 subunits (Figure 2A) (Meier, 2005). Similarly, mass spectrometry analysis also revealed the SCC-B peak activity to be the single polypeptide ATP-binding cassette protein, ABCF1 (Figure 2B).

Reconstitution and mechanism of coactivation by the dyskerin complex

The DKC1 complex is an evolutionarily conserved, four-subunit protein complex that interacts with a large heterogeneous class of small non-coding RNAs called H/ACA small nucleolar RNAs (snoRNAs) (Meier, 2005; Terns et al., 2006). The assembly of a DKC1 RNP in vivo follows an elaborate, multi-step process mediated by the protein chaperones SHQ1 and NAF1 (Darzacq et al., 2006; Grozdanov et al., 2009). The GAR1 subunit subsequently replaces NAF1 in the intermediate complex containing NAF1, DKC1, NHP2, and NOP10 to form the mature RNP only after snoRNAs are incorporated and properly processed (Kiss et al., 2010). These H/ACA snoRNAs guide sequence-specific pseudouridylation of ribosomal RNAs (rRNAs) and spliceosomal small nuclear RNAs (snRNAs) by the catalytic subunit DKC1 (Liang and Li, 2011). The DKC1 complex also plays a key role in the biogenesis of telomerase by binding and promoting the

processing and intranuclear trafficking of telomerase RNA (TERC) (Egan and Collins, 2012). Given the intimate association of the DKC1 complex with numerous RNAs and the multiple factors required to assemble the RNP in vivo, it is remarkable that an RNA-free, ternary 'apo-complex' can be generated in vitro. Indeed, several crystal structures of the archeal and yeast partial and holo-complexes of DKC1 revealed direct protein–protein contacts among the four subunits independent of RNA (Li and Ye, 2006; Li et al., 2011).

To firmly establish that the DKC1 complex was indeed responsible for the SCC-A coactivator activity, we set out to reconstitute the human DKC1 complex from recombinant expression in insect (Sf9) and bacterial cells for use in our in vitro transcription assays. Recombinant hetero-dimeric (DKC1-NOP10), -trimeric (DKC1-NHP2-NOP10), ternary (NAF1-DKC1-NHP2-NOP10), and holo-DKC1 complexes were tested for their ability to potentiate OCT4/SOX2-dependent transcriptional activation of Nanog in vitro, in place of the SCC-A fraction. Remarkably, all partial and complete recombinant complexes whether produced in E. coli or Sf9 cells exhibited similar specific activities for coactivation, but were reproducibly less active than the purified native endogenous DKC1 complex from NT2 cells (Figure 2C). It was not clear whether the reduced specific activities of the recombinant purified complexes resulted from poorly folded or assembled subunits, presence of inhibitory RNAs, or both. Nevertheless, these results using purified recombinant subunits confirm that at least the protein components of the DKC1 complex represent a major contributor of the SCC-A coactivator function. Indeed, it appears that the largest subunit DKC1 and the smallest protein NOP10 are sufficient to provide the bulk of the transcriptional coactivator function and that an RNA component may not be strictly required for this moonlighting activity of the DKC1 complex, since bacterially expressed DKC1 lacks detectable associated RNAs. Although snoRNAs may not be essential for conferring coactivator competence to the recombinant DKC1 complexes, we note that the endogenous DKC1 complexes are twofold to threefold more active than their recombinant counterparts, suggesting that some mammalian snoRNAs may play a role in enhancing the transcriptional activity of the DKC1 complex.

Some snoRNAs may modulate DKC1 coactivator function

The DKC1 RNPs in mammalian cells are highly heterogeneous—with more than 100 known H/ACA snoRNAs that form an equally large number of distinct RNPs by associating with the same four core protein subunits, some of which with unknown functions (i.e. orphan snoRNAs that lack base complementarity to rRNAs or snRNAs) (Kiss et al., 2010). New classes of snoRNAs have also been identified and shown to directly participate in disparate cellular processes from pre-mRNA splicing to chromatin decondensation (Jády et al., 2012; Schubert et al., 2012; Yin et al., 2012). Furthermore, as much as 60% of snoRNAs can be processed into microRNAs (miRNAs), most of which have unknown targets (Ender et al., 2008; Taft et al., 2009). Thus, our understanding of the full repertoire of H/ACA snoRNAs and their 'non-canonical' functions remains limited. It is also unclear if the binding of different snoRNAs to the

human DKC1 complex induces structural changes or masks protein surfaces that may positively or negatively impact coactivator function. Given that most, if not all, GAR1-containing DKC1 complexes are mature RNPs in vivo (Kiss et al., 2006), it seemed prudent for us to examine the range and specific activity of these native but heterogeneous mixtures of human DKC1 RNPs.

Even though these 100 or more DKC1 RNPs have highly similar if not identical core protein composition and architecture, we reasoned that these RNPs are likely to display distinct chromatographic properties due to their unique snoRNAs and/or associated factors. In an attempt to biochemically fractionate this heterogeneous population of DKC1 RNPs, a partially purified fraction prepared from 200 L of NT2 cells that contains >95% of the total population of human DKC1 RNPs was applied to a Poros-HQ anion exchange column and fractionated using a salt gradient (Figure 3A). As expected, DKC1 RNPs were found to elute in a broad profile from 0.3 to 0.9 M KCl with the majority of the complexes eluting at ~ 0.5 M (Figure 3B), consistent with extensive heterogeneity of the DKC1 RNPs in NT2 cells. We next immuno-affinity purified the various DKC1 RNPs from different salt eluted Poros-HQ fractions using a monoclonal antibody against human DKC1 followed by peptide elution. The various affinity-purified DKC1 RNP pools all contain stoichiometric amounts of the four core protein subunits, indicating that they are likely mature RNPs (data not shown). However, we failed to detect any other major associated polypeptides in these purified samples. Therefore, differences in protein composition alone are unlikely to fully account for the observed chromatographic heterogeneity of the DKC1 RNPs separated by the salt gradient on a Poros-HQ column. Instead, we strongly suspect the differential chromatographic behavior of the endogenous human DKC1 RNP complexes to derive from association with distinct RNA species. Indeed, 5′ end radiolabeling of the purified RNA species from the various affinity-purified DKC1 RNP preparations revealed distinct patterns of associated RNAs (Figure 3C). The DKC1 RNPs purified from high salt eluted fractions (# 22, 26, and 30) were enriched for longer RNAs (>180 nucleotides) and some select shorter RNAs (<100 nucleotides). The 130–140 nucleotide-long snoRNA clusters were recovered from DKC1 immunoprecipitates from multiple fractions spanning a wide spectrum of the salt gradient. Thus, it appeared that parameters in addition to RNA length may contribute to the observed differential chromatographic properties of different DKC1 RNPs. Of note, the DKC1 RNPs purified from fraction 9 did not appear to contain significant amounts of RNA (Figure 3C). This is unexpected because the presence of GAR1 usually signifies that some RNA species should have been loaded into the complex in the normal course of DKC1 RNP assembly. However, we cannot exclude the possibilities that, although unlikely, RNAs were present but somehow refractory to labeling at both 5′ (Figure 3C) and 3′ ends (data not shown). It remains possible that some snoRNAs were degraded or had dissociated from a small fraction of the DKC1 RNPs during purification.

These highly purified pools of DKC1 RNPs were assayed over a threefold dose–response range in our fully reconstituted in vitro transcription reactions containing OCT4, SOX2, recombinant XPC complex and SCC-B. Remarkably, DKC1 RNPs purified from higher salt eluted Poros-HQ fractions (fractions 26 and 30) displayed

significantly higher specific activity than those from lower salt fractions (fractions 9 and 14) (Figure 3D). We estimated a ~sixfold enhancement in the specific activities of DKC1 RNPs purified from fraction 30 compared to fraction 9, which, as we had shown in Figure 3C, contained no detectable RNAs (Figure 3D, compare lanes 3 and 13). It is unclear if this endogenous apo-complex lacking any detectable RNA component is physiologically relevant or an experimental artifact. However, the fact that this apo-complex activated transcription with reduced specific activity (Figure 3D, compare lanes 1 and 3) is consistent with our previous observation that the bacterial apo-complex is less active than DKC1 RNPs purified from NT2 cells in supporting transcription (Figure 2C). Paradoxically, recombinant DKC1 RNPs purified from Sf9 cells contained insect snoRNAs, but exhibited low specific activities similar to the bacterial and apo-complexes suggesting that some RNAs may be inhibitory. Taken together, these results uncover a previously unrecognized gene regulatory role of the DKC1 RNP complex wherein a subset of mammalian non-coding snoRNAs may enhance the DKC1 coactivator function while other RNAs may inhibit its transcription activity.

Mechanisms of DKC1 coactivation in vitro and in vivo

Identification of the DKC1 RNP and the XPC DNA repair complexes – and subsequently ABCF1, as will be discussed later in the chapter – as co-dependent coactivators for OCT4/SOX2 was unexpected on two fronts. These two multi-subunit protein assemblies had not been previously implicated in directing stem cell-specific transcription nor had they been functionally linked to each other in any cellular processes. We therefore set out to determine the functional relationship between these newly identified stem cell coactivators and their mechanisms of coactivation in vitro and in vivo. Our ability to recombinantly express and purify these coactivators allowed us to systematically test the contribution of each coactivator alone in supporting OCT4/SOX2-activated transcription in vitro. Addition of individual coactivator complexes only marginally activated Nanog transcription (Figure 2E). However, when the DKC1 complex was supplemented with the XPC complex, we observed a noticeable increase in transcriptional output that was substantially further enhanced by adding the third coactivator, SCC-B (Figure 2E). These results confirmed the co-dependent nature of these three coactivators in supporting an optimal, synergistic activation of the Nanog gene by OCT4 and SOX2. To further explore the mechanism by which the DKC1 complex cooperates with the XPC complex in OCT4/SOX2 activated transcription, we co-expressed both complexes along with (or without) OCT4 and SOX2 in 293T cells and performed co-immunoprecipitation assays to probe for a potential interaction between these two coactivators. Immunoprecipitation of the XPC complex pulled down the DKC1 complex both in the presence and absence of the activators. This finding suggests that the DKC1 complex may function as an OCT4/SOX2 coactivator in part through a direct physical interaction with the XPC complex, which in turn binds OCT4 and SOX2 (Figure 4A). In support of this observation, a recent global proteomic study using large scale biochemical fractionation of human cell extracts to isolate stable protein complexes identified WDR79, a known accessory protein of the mature DKC1 RNP (Jády et al., 2012; Tycowski et al., 2009), as a candidate XPC-interacting protein

(Havugimana et al., 2012). Whether the DKC1 complex also forms direct contacts with OCT4 and SOX2 in the absence of XPC is unclear. Our attempt to address this was hampered by the fact that we could not express any of the four subunits of the DKC1 complex to a significant level in 293T or several other cell lines (data not shown). However, the fact that co-expression of OCT4/SOX2 did not increase the amount of DKC1 pulled down by the XPC complex argues against a stable tripartite complex wherein the coactivators interact with each other and form independent contacts with the activators.

Mutations in the Dkc1, Nhp2, and Nop10 genes have been linked to dyskeratosis congenita (DC), a rare but fatal human genetic disorder that impairs stem cell function and proliferation generally attributed to defects in telomerase or ribosome biogenesis (Mason and Bessler, 2011; Mitchell et al., 1999). Our discovery of a stem cell-specific transcriptional role of the DKC1 complex adds a potentially important alternative mechanism for interpreting the molecular basis of DC disease phenotypes. However, it was unclear if amino acid residues critical for telomerase and ribosome biogenesis impinge on distinct or overlapping domains with respect to our newly uncovered DKC1 transcription coactivator function. To begin to address this potentially important link to disease, we focused on DC mutations in the large DKC1 subunit and the small NOP10 protein because a partial complex of these two subunits was sufficient to activate Nanog transcription in vitro (Figure 2C). We recombinantly expressed and purified a panel of mutant DKC1 complexes in Sf9 cells that are representative of both position (L37del (Heiss et al., 1998), A353V (Knight et al., 1999), Δ22C (He et al., 2002), and frequency (A353V) at which DC mutations occur in Dkc1 (Marrone et al., 2005). We also generated an artificial, pseudouridine synthase inactive mutant DKC1 (D125A (Gu et al., 2013)) as well as a mutant NOP10 containing (R34W (Walne et al., 2007)) complex (Figure 4C). Remarkably, all mutant DKC1 RNPs were consistently more active than the WT holo-complex in potentiating OCT4/SOX2-mediated transcription (Figure 4D). Therefore, it appeared that neither the enzymatic activity nor amino acids mutated in DC are essential for coactivator activity although the enhanced coactivator phenotype could lead to changes in gene expression and altered stem cell function. The transcriptional phenotypes of these DKC1 mutations are highly reminiscent of our findings with the XPC complex in that disease-relevant amino acids and domains critical for DNA repair functions were also dispensable for OCT4/SOX2-activated transcription (Fong et al., 2011).

To further probe the molecular mechanisms by which the DKC1 complex might function as a transcriptional coactivator for OCT4 and SOX2 in ES cells, we performed chromatin immunoprecipitation (ChIP) assays to investigate whether the DKC1 complex is directly recruited to regulatory regions of key OCT4/SOX2-target genes. ChIP-qPCR analysis revealed that sites of DKC1 occupancy at the Oct4, Nanog, Sox2 genes indeed coincide with those of OCT4 (Boyer et al., 2005; Loh et al., 2006) and SOX2 binding to enhancer and promoter DNA sequences in the mouse ES cell line D3 (Figure 5A,B). Importantly, DKC1 binding is also enriched at the enhancers of Oct4 and Nanog in human ES cell line H9 (Figure 5C) and EC cell line NT2 (Figure 5D), thus confirming the generality of a co-recruitment mechanism to transcriptional regulatory elements in pluripotent stem cells. Curiously, we failed to detect a significant enrichment of DKC1 at

some OCT4/SOX2-target genes such as Fgf4 in D3 cells (Figure 5A,B). This suggests that the DKC1 complex may be differentially employed by OCT4 and SOX2 to regulate a subset of their target genes. Additional experiments such as genome-wide analyses of DKC1 occupancy will be required to ascertain the extent to which DKC1 associates with OCT4/SOX2 target genes in mouse ES cells. Since over 90% of snoRNAs are embedded in the introns of coding and non-coding genes (Filipowicz and Pogacic, 2002), the DKC1 complex has also been found to localize at gene bodies where it is thought to co-transcriptionally process nascent snoRNAs (Ballarino et al., 2005; Darzacq et al., 2002; Yang et al., 2005). Now our finding of the DKC1 complex co-occupying pluripotent gene promoters and enhancer elements with sequence-specific activators OCT4 and SOX2 in ES cells strongly suggests a classical coactivator function of the DKC1 complex rather than acting purely as a snoRNP maturation factor.

The DKC1 function in stem cell maintenance and somatic cell reprogramming

Many transcriptional activators (OCT4, SOX2, NANOG) and coactivators (Mediator, TAFs/TFIID, the XPC complex) critical for stem cell pluripotency are often highly enriched in ES cells, but become rapidly down-regulated upon differentiation. Dynamic regulation of these transcription factors in ES cells is thought to confer not only stability to the transcriptional circuitry governing self-renewal but also the flexibility to exit the pluripotent state and switch between competing developmental programs during differentiation (Fong et al., 2012; Jaenisch and Young, 2008; Liu et al., 2011). Consistent with the notion that the DKC1 complex is performing as a stem cell-specific coactivator in ES cells, the DKC1, GAR1, and NOP10 subunits are highly enriched in pluripotent D3 cells (Figure 6A). Their levels in ES cells decreased rapidly upon retinoic acid (RA)-induced differentiation, while general transcription factor TFIIB and loading control β-Actin remained unchanged. Importantly, the selective decrease of DKC1 levels was not simply a reflection of a reduced proliferative state or protein translational activity in differentiating ES cells because components of the C/D snoRNP (FBL and NOP58), another major machinery involved in the ribosome biogenesis pathway, stayed largely constant (Su et al., 2014). Indeed, it has been shown that transcription of the Dkc1 gene is regulated by OCT4 and NANOG in ES and iPS cells (Agarwal et al., 2010), thus providing a transcriptional mechanism whereby Dkc1 expression levels are tightly coupled to the pluripotent state.

To gain additional in vivo evidence that the DKC1 complex is required for the proper expression of genes critical for stem cell self-renewal, we performed loss-of-function studies using lentiviruses expressing two independent short hairpin RNAs (shRNAs) specifically targeting DKC1 in mouse D3 ES cells (Figure 6B). We also depleted XPC in D3 cells using a previously characterized shRNA (Fong et al., 2011) to investigate potential functional interactions between these two coactivator complexes. Interestingly, knockdown of DKC1 resulted in co-depletion of the small NOP10 subunit indicating that the stability of individual subunits likely depends on the integrity of the DKC1 complex (Figure 6B). This may also explain why a decrease in protein levels of GAR1 and NOP10 during RA-induced ES cell differentiation follows the same kinetics

as DKC1 even though Gar1 and Nop10 do not appear to be direct targets of OCT4 and SOX2 (Figure 6A). Compared to control knockdown D3 cells, shRNA-mediated silencing of XPC (shXPC) or DKC1 (shDKC1-1 and shDKC1-2) resulted in pronounced morphological abnormalities including rapid collapse of the tightly packed ES cell colonies and appearance of large, flattened cells with concurrent dramatic reductions in alkaline phosphatase activity, all indicative of enhanced spontaneous differentiation of ES cells (Figure 6C). At this point, we cannot rule out the possibility that the severe phenotype observed in DKC1 knockdown ES cells is at least partially contributed by disruption of other well documented DKC1-dependent cellular processes (telomerase function and ribosome biogenesis) in addition to its transcription coactivator function. However, mouse ES cells lacking telomerase activity (Terc −/− (Niida et al., 1998)) or carrying a pathogenic mutation in Dkc1 (A353V (Mochizuki et al., 2004)) can be maintained in culture for over 300 population doublings with no observable impact on growth rate and only a very mild effect on ribosome biogenesis. Since the self-renewal defects observed in DKC1 knockdown ES cells became apparent by 3 days post lentiviral infection (<9 population doublings), cellular senescence or a gross defect in rRNA processing are unlikely to be major contributors to the DKC1 knockdown phenotypes we observed in ES cells.

Consistent with the evident morphological changes associated with compromised stem cell identity, single knockdown of XPC or DKC1 in D3 cells resulted in a significant reduction in mRNA levels of core pluripotency genes including Nanog, Oct4, Sox2, Klf4, as well as stem cell marker Fgf4 (Figure 6D). Interestingly, simultaneous knockdown of XPC and DKC1 did not further reduce their expression. This is consistent with the co-dependent nature of the DKC1 and XPC complexes in gene activation wherein the absence of one coactivator severely limited the ability of the other two stem cell coactivators to stimulate Nanog transcription in vitro (Figure 2E). To further explore the spontaneous differentiation phenotype in DKC1 and XPC-deficient ES cells, we performed qPCR analyses to monitor the expression level of lineage-specific markers representing the three germ layers and the trophectoderm. Depletion of DKC1 or XPC upregulated the expression of neuroectodermal maker Fgf5 and trophoblast-specifier Cdx2 at the expense of Gata6, a primitive endoderm marker, while mesodermal marker T remained unchanged (Figure 6E). Double knockdown of DKC1 and XPC appeared to further augment the expression of Cdx2, but not Fgf5. The observed differentiation bias in DKC1 and XPC knockdown ES cells may be in part due to the reduced levels of OCT4 and NANOG, both of which have well-documented functions in antagonizing differentiation of extraembryonic lineages including the trophectoderm (Hay et al., 2004; Hyslop et al., 2005; Niwa et al., 2000; Silva et al., 2009).

The essential role of the DKC1 complex in establishing an OCT4/SOX2-dependent gene expression program in ES cells led us to hypothesize that DKC1 may be required for the reacquisition of pluripotency during cellular reprogramming by ectopic expression of OCT4, SOX2, KLF4, and c-MYC (Takahashi and Yamanaka, 2006). Of note, recent studies showed that primary human adult fibroblasts (HFs) carrying pathogenic mutations in Dkc1 are refractory to cellular reprogramming (Agarwal et al., 2010; Batista et al., 2011). However, it is important to point out several key differences between using MEFs and HFs derived from DC patients to study DKC1

function in somatic cell reprogramming. Unlike MEFs which display high levels of telomerase activity and long telomeres (>50 kb (Blasco et al., 1997)), HFs lack measurable TERT activity and have relatively short telomeres (10–15 kb (Harley et al., 1990)). In fact, telomerase null MEFs can be propagated in culture for more than 200 cell divisions without loss of viability (Blasco et al., 1997), which make MEFs a potentially better cell culture system for studying telomerase-independent functions of DKC1 in reprogramming. By contrast, DC patient-specific fibroblasts have shorter telomeres and could also accumulate secondary mutations due to genome instability, which are both detrimental to the reprogramming process (Fong et al., 2013). Consistent with this notion, it was shown that ectopic expression of wild type DKC1 (or TERT) in a DC mutant fibroblast line (Dkc1 L37del) failed to rescue the reprogramming defect phenotype (Agarwal et al., 2010). Therefore, it remains unclear what impact, if any, acute DKC1 depletion in MEFs will have on iPS cell generation.

To address this question, we infected MEFs with lentiviruses expressing non-targeting control shRNA or two independent shRNAs specific for DKC1 and initiated reprogramming by doxycycline (dox)-induced expression of OCT4, KLF4, SOX2 and c-MYC (OKSM) (Sommer et al., 2009). We observed a marked decrease in the number of AP-positive iPS cell colonies (~20–50-fold reduction) whether or not we plated the induced DKC1 knockdown MEFs directly onto gelatin coated plates (where the surrounding DKC1 knockdown MEFs refractory to reprogramming acted as feeder cells) or onto mitomycin-treated feeder cells (Figure 7A). This suggests that failure of DKC1-deficient MEFs to acquire pluripotency is likely a cell autonomous phenomenon. Flow cytometry analysis showed that the majority of both control and DKC1 knockdown cells down-regulated fibroblast-associated cell surface marker THY1 indicating a loss of MEF identity (Figure 7B). However, unlike control cells where many of them became SSEA1+ and ultimately gave rise to AP and NANOG-positive iPS cell colonies, most DKC1 knockdown cells do not (Figure 7B,C). Because of the observed early arrest in reprogramming associated with DKC1-depleted MEFs, we next asked whether these cells were able to undergo the mesenchymal-to-epithelial transition (MET), a requisite initiating event prior to expression of SSEA1 antigen (Golipour et al., 2012; Li et al., 2010; Polo et al., 2012; Samavarchi-Tehrani et al., 2010). By day 14 post dox-induction, control knockdown MEFs showed reduced expression of fibroblast-enriched, pro-mesenchymal genes Slug and Snail, but their levels remained noticeably higher than those in ES cells (Figure 7D). This is likely due to contaminating partially reprogrammed iPS cells and residual fibroblasts present in the induced cell culture (Figure 7B). These non-target knockdown cells also acquired epithelial characteristics indicated by elevated levels of Ecad and Epcam (Figure 7D), as expected, given that THY1-/SSEA1+ partially and fully reprogrammed iPS cells represent the bulk of these control cells (Figure 7B). By contrast, depletion of DKC1 in MEFs blocked the reactivation of epithelial genes (Ecad and Epcam) without significantly perturbing the silencing of mesenchymal genes (Figure 7D), thus effectively uncoupling the otherwise tightly coordinated MET induced by OKSM (Liu et al., 2013). These data taken together suggest that DKC1 could be required for reprogrammed MEFs to acquire an epithelial identity during the critically important mesenchymal-to-epithelial transition.

To address whether the early reprogramming arrest observed in DKC1-depleted MEFs can be attributed to a gross defect in cellular proliferation, we labeled control and DKC1 knockdown MEFs with a stable dye (CFSE). The doubling time of these cells was measured by monitoring the decrease in dye intensity resulting from cell division over a 4 day period (Figure 8A). Although a lengthening of doubling time of DKC1-depleted MEFs by both targeting shRNA hairpin was observed, compared to shDKC1-1, knockdown by shDKC1-2 has a significantly more pronounced effect on iPS cell generation while having a minimal impact on cellular proliferation (Figure 7—figure supplement 1; Figure 8A). This suggests that reprogramming efficiency does not strictly correlate with changes in proliferation rates caused by DKC1 depletion. However, we cannot exclude the possibility that differences in doubling rates could be due to differential off-target effects of the two hairpins. We also note that DKC1 depletion does not cause abrupt growth arrest of all MEFs but appears to selectively impair the proliferation of the faster cycling subpopulation without affecting the rest of the slower-dividing MEFs (Figure 8A). Therefore, factors in addition to growth impairment are at least contributing to the observed defect in somatic cell reprogramming. These observations are also consistent with recent findings suggesting that essentially all of the reprogramming potential in OKSM-induced MEF cultures is confined to a small fraction (1–8%) of cells characterized by an ultrafast cell cycle (Guo et al., 2014c; Smith et al., 2010). Given the multiple functions of the DKC1 complex in regulating cellular proliferation (Alawi et al., 2011), MET (Figure 7D) and pluripotency gene expression (Figure 6D), we asked whether DKC1 might also be involved in overcoming barriers to deterministic cellular reprogramming. After 4 days of dox-induced expression of OKSM in MEFs, we labeled cells with CFSE and continued dox treatment for another 2 days before subjecting them to FACS (Guo et al., 2014c). We identified and characterized four distinct cell populations bearing variegated dye concentrations (Figure 8B). The fastest dividing population (CFSE-Lo) had undergone at least 4 more cell divisions than the bulk MEFs and gave rise to substantially more AP-positive iPS cell colonies than the slower-dividing populations (Figure 8B). Strikingly, CFSE-Lo cells also expressed the highest levels of Dkc1 reaching that of ES cells (Figure 8C). These cells have lost their mesenchymal identity and initiated the transition into cells of epithelial origin (Figure 8D). By contrast, the slower dividing populations expressed significantly lower levels of Dkc1 and failed to fully silence mesenchymal genes or robustly reactivate epithelial markers indicating a delayed or abortive MET. Importantly, using MEFs carrying an integrated dox-inducible OKSM expression cassette (Carey et al., 2010), we observed a similar preferential enrichment of Dkc1 in the fastest-dividing population (CFSE-Lo) despite uniform Oct4 expression levels among CFSE-Lo, Med, and Hi cells (data not shown). Therefore, an early onset of MET appears to be a defining property of these ultrafast cycling cells wherein appropriately high levels of DKC1 are necessary and likely serve as an important gene-specific transcriptional coactivator.

Reconstitution and mechanism of coactivation by ABCF1

ABCF1 (also termed ABC50) is a member of the ABC protein family, defined by highly conserved nucleotide-binding motifs (Walker A and Walker B motifs) and a highly

conserved ABC 'signature motif'(LSGGQ) (Klein et al., 1999; Schneider and Hunke, 1998). Unlike nearly all identified ABC proteins, which contain transmembrane domains and function as transporters, ABCF1 is the only mammalian ABC to lack a transmembrane domain, containing three nuclear localization signals instead (Klein et al., 1999). ABCF1 was first identified as a tumor necrosis factor alpha (TNFα)-inducible gene in synoviocytes (Richard et al., 1998) and subsequently has been shown to play roles in translation initiation (Paytubi et al., 2009; Tyzack et al., 2000), innate immunity and cytosolic dsDNA sensing (Lee et al., 2013), as well as a mediator of phagocytosis (Guo et al., 2015). Interestingly, genetic ablation of *Abcf1* in mice results in embryonic lethality at E3.5, though the exact mechanism for lethality remains unknown (Wilcox, 2010).

To further confirm that ABCF1 was indeed responsible for the SCC-B coactivator activity, we recombinantly expressed and purified ABCF1 from bacterial cells and tested its ability to replace SCC-B peak fractions in OCT4/SOX2-dependent transcription. Recombinant ABCF1 was able to stimulate transcription just as well as the endogenously expressed ABCF1 present in the SCC-B peak fractions, if not to a slightly higher extent (Figure 2D). The addition of all three stem cell coactivators – the XPC complex, the DKC RNP, and ABCF1 – remain required for synergistic, co-dependent activation of *Nanog* transcription (Figure 2E).

Mechanisms of ABCF1 coactivation in vitro and in vivo

Given that little is known about the potential role or roles that ABCF1 plays in the cell, we attempted to narrow down the protein domain requirements required for its coactivator activity through mutational analysis. Surprisingly, the ATP-binding activity of ABCF1 was not required for its transcriptional activity (data not shown). Furthermore, the deletion of the N-terminal region markedly decreased transcriptional activity, but the N-terminus was not sufficient in potentiating transcriptional activity when expressed alone (data not shown). Given the likely involvement of the N-terminal domain, we tested the transcriptional activity of the yeast homolog, Gcn20. Gcn20 similarly contains 2 ABC units and shares 28.9% identity with the human ABCF1 and has been previously linked with translation initiation regulation (ref).

Because no antibodies against ABCF1 capable of ChIP are commercially available, we are currently employing a CRISPR/Cas9 genome editing strategy to knock-in a HA tag adjacent to ABCF1 in D3 mESCs.

The ABCF1 function in stem cell maintenance and somatic cell reprogramming

A number of surprising transcription factors and cell modulators have been recently found to serve multifunctional roles, including stem cell pluripotency. As expected, ABCF1 is more highly enriched in the pluripotent cells and is critical stem cell maintenance. We performed ABCF1 loss-of-function studies in mES cells. shRNA-

mediated knockdown of ABCF1 (shABCF1-2 and shABCF1-4) results in rapid, spontaneous differentiation, as seen by the collapse of colony morphology and reduced AP activity (Figure 9A). Consistent with this, there is a marked reduction in pluripotency genes *Nanog*, *Klf4,* and *Fgf4* in ABCF1 knockdown cells, but not *Oct4* and *Sox2,* which are both reduced in XPC and DKC1 knockdown cells (Figure 9B). Additionally, neuroectodermal marker *Fgf5* and mesodermal marker *T* increased upon knockdown of ABCF1, while primitive endoderm marker, *Gata6*, decreased (Figure 9B). These changes are consistent with mES cells exiting the naïve pluripotent state and progressing toward lineage specification.

To further address the role of ABCF1 in pluripotency, we transduced MEFs with lentiviruses expressing a non-targeting or ABCF1-specific shRNAs and initiated reprogramming by dox-induced expression of OKSM (Sommer et al., 2009). Similar to the results we observed with the two other stem cell coactivators (XPC and DKC1 complexes), ABCF1 knockdown led to a striking reduction in the number of AP-positive iPS cell colonies formed (Figure 10A-B). Flow cytometry analysis further demonstrated that loss of ABCF1 led to a decrease in SSEA-1+ cells, suggesting that a reduced number of cells are capable of advancing through the early stage of reprogramming (Figure 10C) (Polo et al., 2012; Stadtfeld et al., 2008).

Because of the observed early arrest in reprogramming associated with ABCF1 depletion, we next asked whether these cells were similar to DKC1-depleted MEFs, which failed to undergo MET. By 14 dpi, control knockdown MEFs showed reduced expression of mesenchymal genes *Slug* and *Snail*, compared to uninduced MEFs; epithelial genes *Ecad* and *Epcam* were also elevated (Figure 10D). By contrast, ABCF1 knockdown MEFs failed to activate *Ecad* and *Epcam,* even though the negative regulator *Slug* is downregulated (Figure 10D). While *Snail* remains high in the ABCF1 loss-of-function cells, it is important to note that *Snail* has been recently reported to play a paradoxical role in the positive regulation of pluripotency, thus making this result difficult to interpret without further information (Lin et al., 2014; Unternaehrer et al., 2014). Overall, these data indicate that ABCF1 knockdown cells arrest early in reprogramming, failing to obtain an epithelial-like identity.

One possible explanation for an early roadblock in reprogramming is reduced cell proliferation. Cells that are unable to cycle quickly have much lower reprogramming efficiency, and in some cases are even refractory to the reprogramming process (Banito et al., 2009; Guo et al., 2014c; Hong et al., 2009; Kawamura et al., 2009; Ruiz et al., 2011). To address whether ABCF1-depletion led to a gross defect in MEF proliferation, we labeled control and knockdown cells with CFSE. The doubling time of these cells was measured by monitoring the reduction in fluorescent intensity resulting over a 3 day period by flow cytometry. Indeed, a noticeable lengthening of doubling time of ABCF1-depleted MEFs was observed and appears to correlate not only with knockdown efficiency, but also reprogramming efficiency (Figure 10E).

We next asked whether or not *Abcf1* expression would correlate with deterministic reprogramming, as it did with DKC1 (Figure 8B-C) and XPC (data not shown). Given that the loss of ABCF1 led to a decrease in cell cycling (Figure 10E), it

was not surprising that *Abcf1* expression would not strictly correlate with doubling speed. This was observed in MEFs that were induced through lentiviral-delivered OKSM (Figure 8B, 11A-B) and MEFs with an integrated OSKM cassette (Figure 11C-D). Thought *Abcf1* levels were generally elevated in the fastest cycling populations, the levels more closely reflect the *Oct4* expression (Figure 11A,D), rather than cycling speed or the expression of MET genes (Figure 8E).

Discussion

Our de novo identification of the DKC1 complex and ABCF1 as transcriptional coactivators for OCT4/SOX2 underscores the expanding repertoire of unconventional, multifunctional coactivators involved in stem cell regulation. Beyond its well-documented role in ribosome and telomerase biogenesis, the DKC1 complex has been shown to effect diverse cellular processes including internal ribosome entry site (IRES)-dependent translation (Yoon et al., 2006) and base excision of 5-hydroxymethyluridines in rRNA by uracil-DNA glycosylase 1 (SMUG1) (Jobert et al., 2013). Interestingly, the telomerase complex itself has been implicated in the regulation of MYC and WNT/β-catenin associated gene expression programs critical for stem cell function (Choi et al., 2008; Park et al., 2009). However, the reverse transcriptase TERT, but curiously not its catalytic activity, was reported to be required for gene activation (Choi et al., 2008), a finding that remains somewhat controversial in the telomerase field (Listerman et al., 2014). In light of our findings that the core DKC1 complex possesses transcriptional coactivator activity, it is tempting to speculate that in the context of WNT-responsive genes, TERT may function to tether the DKC1 complex (as part of the telomerase RNP) to gene promoters and activate transcription by binding to a β-catenin-TCF3 activating complex (Park et al., 2009), an integral component of the core stem cell-specific regulatory circuitry (Cole et al., 2008). It is, however, unlikely that the coactivator activity detected in our assay is dependent on TERT because we did not detect TERT or any other known components of the telomerase complex (Fu and Collins, 2007) in our purified fractions by mass spectrometry (data not shown). This is further supported by our observation that recombinant DKC1-NOP10 heterodimers purified from bacteria was active in transcription. Instead, we favor the model whereby the DKC1 complex (free of any accessory factors) can be recruited to key pluripotency genes via a direct interaction with the XPC complex as we observed in co-immunoprecipitation experiments. This DKC1-XPC assembly is, in turn, recruited to target gene promoters via activator–coactivator interactions with OCT4 and SOX2 (Fong et al., 2011; Gao et al., 2012). The mechanism by which different activators recruit the same coactivator by targeting distinct subunits or protein surfaces within the DKC1 complex may represent a common strategy that is frequently observed with other transcriptional coactivators such as Mediator (Taatjes et al., 2002) and TAFs/TFIID (Liu et al., 2009). Therefore, the DKC1 complex may coordinate diverse transcriptional outputs contributing to stemness by cooperating with both stem cell-specific and cell-ubiquitous activators and coactivators (Mediator, the XPC complex, SCC-B). Interestingly, the *Dkc1* gene itself is also a target of OCT4 and NANOG (Agarwal et al., 2010). Integrating *Dkc1* into the core regulatory circuitry could further stabilize the autoregulatory loops established by OCT4, SOX2 and NANOG that are postulated to confer stability to self-renewing ES cells without sacrificing their responsiveness to developmental cues during differentiation (Boyer et al., 2005; Loh et al., 2006).

Considering that many of the disease-causing mutations found in *Dkc1* have been shown to disrupt the binding and/or stability of a select subset of mammalian snoRNAs which we have shown could play a critical role in conferring coactivator competence to the DKC1 complex, it was somewhat surprising that neither the

enzymatic activity nor amino acids mutated in dyskeratosis congenita (DC) patients negatively impacted coactivator activity. Although targeted disruption of *Dkc1* in mice is lethal (He et al., 2002), many DC patients carrying mutations in the DKC1 complex live well into their teens and beyond, indicating that these mutations are hypomorphic and can be tolerated during embryogenesis. These findings suggest that amino acid residues in DKC1 critical for transcription are likely largely distinct from those mutated in DC patients. In cases where such mutations do overlap, their effect, if any, on transcription are expected to be subtle because mutations that significantly compromise coactivator function would likely be severely detrimental to the tightly regulated process of mammalian development.

Acute depletion of DKC1 in mouse ES cells rapidly down-regulated key pluripotency genes well in advance of telomere attrition and the ensuing cellular senescence that one would expect to occur (after >300 population doublings) due to compromised telomerase function (Mochizuki et al., 2004; Niida et al., 1998). Taken together with our ChIP results showing specific recruitment of the DKC1 complex to OCT4/SOX2 enhancers of core pluripotency genes in both mouse and human ES cells, as well as the strong dependence of OCT4/SOX2-activated transcription on the DKC1 complex in vitro, we suggest that defects in pluripotency gene expression and stem cell self-renewal upon DKC1 knockdown are at least in part due to compromised transcriptional activation rather than a sole consequence of telomerase deficiency. Given the importance of establishing a robust OCT4/SOX2-depedent transcriptional circuitry during iPS cell induction, it is perhaps not surprising that DKC1 knockdown also severely limits reprogramming capacity of MEFs. Our data, however, indicate that reprogramming of DKC1-deficient MEFs by OKSM aborted at a rather early stage—during the mesenchymal-to-epithelial transition (MET). Specifically, DKC1 appears to be required for the proper induction of epithelial markers like *Ecad* (also known as *Cdh1*) critical for iPS cell generation (Chen et al., 2010) even though key negative regulators of epithelial gene expression, *Snail* and *Slug*, were already repressed in DKC1 knockdown MEFs (Thiery et al., 2009). Because SOX2 (or KLF4) alone is sufficient to induce *Ecad* expression in MEFs (Liu et al., 2013), and OKSM cotarget many MET genes early in the reprogramming process (Soufi et al., 2012), the DKC1 complex could cooperate with SOX2 and other reprogramming factors to activate an epithelial gene expression program. It has also been reported that restoring telomerase activity in DC patient-specific fibroblasts carrying a loss-of-function mutation in *Dkc1* (L37del) by overexpressing TERT (Wong and Collins, 2006) is insufficient to overcome the reprogramming defect associated with these cells (Agarwal et al., 2010). Given that L37del fibroblasts show normal rRNA pseudouridine content as well as rRNA processing kinetics (Wong and Collins, 2006), these data taken together strongly suggests a telomerase and ribosome independent mechanism by which the DKC1 complex participates in somatic cell reprogramming. We propose that the DKC1 complex may function to promote the requisite MET during iPS cell generation by activating pro-epithelial genes consistent with its transcriptional coactivator function.

Direct reprogramming of fibroblasts into iPS cells is a slow and presumably stochastic process (Yamanaka, 2009). However, accumulating evidence suggests that it is nonetheless amenable to acceleration (and thereby enhanced efficiency) by

manipulating pathways that promote cell division (Banito et al., 2009; Hanna et al., 2009; Hong et al., 2009; Utikal et al., 2009). These highly proliferative cells competent for reprogramming are also found to exist naturally or become primed by OKSM in a small subset of so called 'privileged' somatic cells within a largely homogeneous population that can proceed through reprogramming in a non-stochastic manner with shorter latency (Guo et al., 2014c). However, the cell-intrinsic determinants conducive to this privileged state remain unclear, but are likely distinct from factors previously implicated in deterministic cellular reprogramming including SOX2 (Buganim et al., 2012) and MBD3 (Liu et al., 2013; Rais et al., 2013). Here we show that a subpopulation of MEFs proliferating at a significantly faster rate is especially sensitive to DKC1 depletion. Similarly, a small fraction of ultrafast cycling MEFs enriched for *Dkc1* and depleted of mesenchymal signatures also emerged upon OKSM expression. Although the relationship between these two highly proliferative MEF populations is unclear, we surmise that they could represent a similar privileged somatic cell state. Therefore, the intrinsic variable levels of DKC1 in regular MEFs, or the ability of some MEFs to upregulate *Dkc1* to a critically high threshold level in response to ectopic expression of OKSM, could be a limiting factor in the acquisition of this rare somatic cell state possibly by facilitating an early MET. The precise role of DKC1 in establishing this privileged state in MEFs is unclear but may involve regulating both cellular proliferation and gene expression critical for the early phase of iPS cell formation.

Relatively little is known about the function of ABCF1 in vivo. In one study, ABCF1 was identified as a cytosolic dsDNA sensor with the ability to activate CXCL10, an IFN-γ responsive gene. ABCF1 depletion also leads to a decrease in phospho-TBK1 and phospho-IRF3 (Lee et al., 2013), downstream effectors of TLR3/TRIF signaling in innate immunity responses (Kawai and Akira, 2010). Innate immunity responses, particularly the TLR3/TRIF pathway, have been previously implicated in being required for efficient iPS reprogramming (Lee et al., 2012). Thus it is interesting to speculate that the decrease in reprogramming efficiency observed in TLR3/TRIF loss-of-function cells may in part be synonymous with ABCF1 knockdown by affecting the same target genes. It has also become increasingly evident that genome safeguarding is an important process during somatic reprogramming in order to generate bona fide iPS cells (Fong et al., 2013). This raises the possibility that ABCF1 could function – similar to XPC – as both a transcriptional coactivator and a DNA repair/genome stability protein.

In summary, using an unbiased biochemical approach to probe the transcriptional regulation of the *Nanog* gene by OCT4 and SOX2, we uncovered an unanticipated transcriptional coactivator role of the DKC1 complex and ABCF1 in ES and iPS cells. We surmise that the DKC1 complex could be one of the cell-intrinsic determinants that impinge on somatic cells during reprogramming by coupling cellular proliferation to stem cell-specific transcription. Continued mechanistic characterization is required to understand the connection, if any, between the role ABCF1 has in ESC transcriptional regulation and innate immunity.

Figures

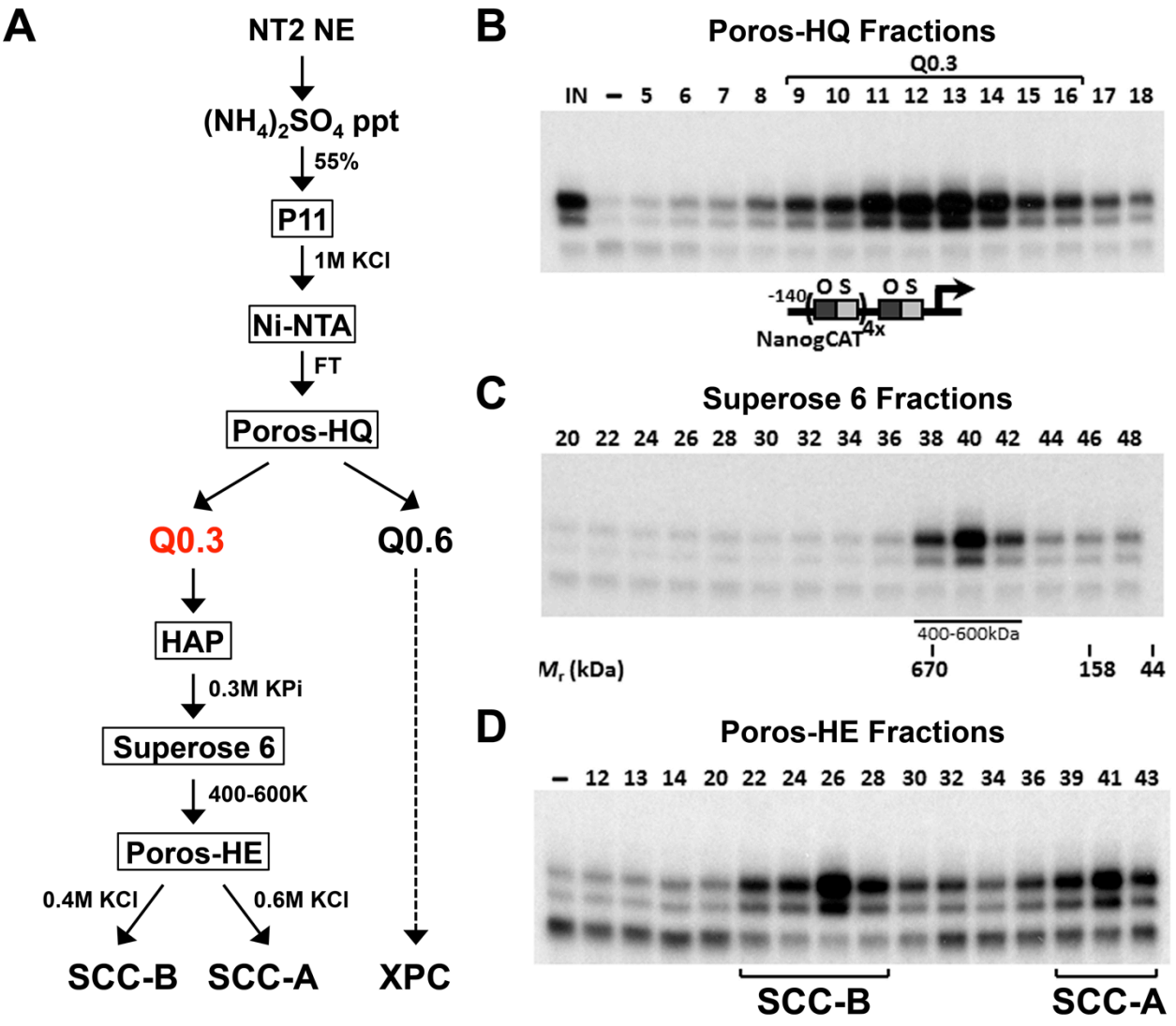

Figure 1. Purification of Stem Cell Coactivator-A and -B required for OCT4/SOX2-dependent activation of the *Nanog* gene. (**A**) Chromatography scheme for purification of Q0.3 from NT2 nuclear extracts (NT2 NE). NT2 NE is first subjected to ammonium sulfate precipitation (55% saturation) followed by a series of chromatographic columns including cation exchangers phosphocellulose (P11), heparin (Poros-HE), the anion exchanger Poros-HQ, hydroxyapatite (HAP), and gel filtration medium Superose 6. (**B**) Input (IN, Ni-NTA flowthrough), buffer control (−), and fractions containing Q0.3 eluted from a Poros-HQ anion exchanger (fraction number indicated) are assayed in the presence of OCT4, SOX2, and recombinant XPC complex in in vitro transcription assays. (**C**) Q0.3 appears to migrate as a single activity. Superose 6 fractions are assayed as in (**B**). Mobilities of peak activity (400–600 K) and gel filtration protein standards are shown at bottom. (**D**) Q0.3 is composed of two distinct coactivator activities, SCC-A and SCC-B. Transcription reactions contain buffer control (−), Poros-HE fractions and are assayed as in (**B**).

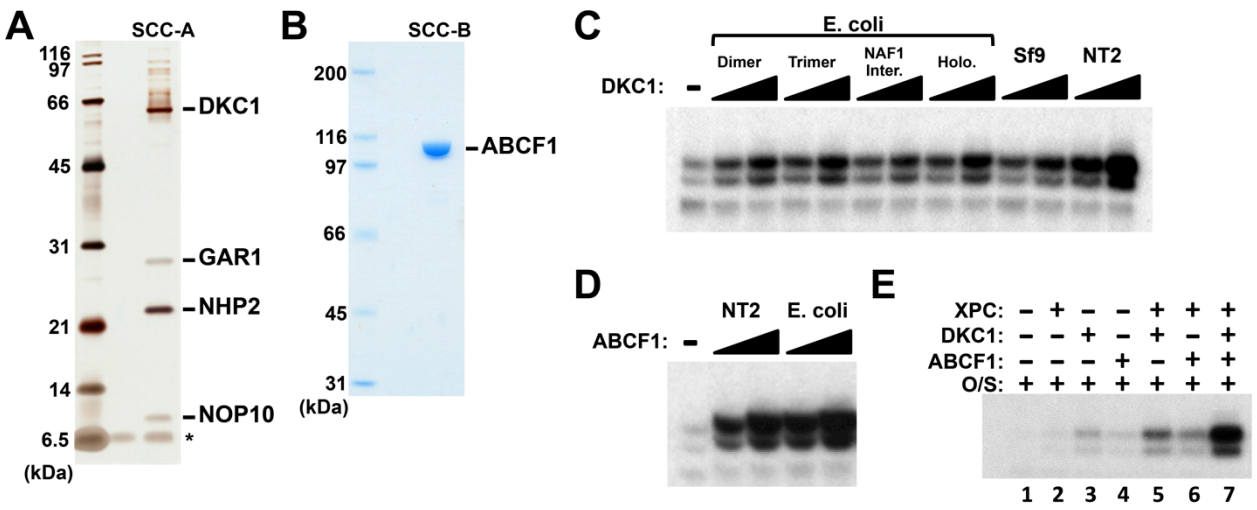

Figure 2. Mass spectrometry analyses reveal the identities of SCC-A and SCC-B to be the dyskerin (DKC1) complex and ABCF1. (**A**) Silver stained SDS-PAGE gel of Poros-HE peak activity corresponding to the SCC-A activity. Peptide identities were determined by mass spectrometry analyses. The asterisk indicates insulin added to Poros-HE fractions as a protein stabilizer. (**B**) SDS-PAGE and Coomassie staining of the peak activity corresponding to the SCC-B activity. Mass spectrometry analysis reveals the single polypeptide species to be the ATP-binding cassette protein, ABCF1. (**C**) Recombinant DKC1 complexes can functionally replace the SCC-A peak activity and enhance OCT4/SOX2-activated transcription of *Nanog*. Buffer control (−), bacterially expressed DKC1-NOP10 heterodimer (dimer), DKC1-NHP2-NOP10 heterotrimer (trimer), NAF1-DKC1-NHP2-NOP10 intermediate (NAF1 Inter.), and holocomplex DKC1-GAR1-NHP2-NOP10 (holo.), recombinant Sf9-expressed DKC1 complex (Sf9), and endogenous holo-complex purified from NT2 (NT2) are assayed over a twofold concentration range. Transcription reactions contain OCT4, SOX2, recombinant XPC complex, and a Poros-HE fraction containing SCC-B. (**D**) Purified recombinant ABCF1 can functionally replace the SCC-B peak activity and enhance OCT4/SOX2-activated transcription of *Nanog*. Buffer control (−), purified endogenous ABCF1 from NT2 (NT2) and bacterially expressed recombinant ABCF1 (E. coli) are assayed over a twofold concentration range. Transcription reactions contain OCT4, SOX2, recombinant XPC complex, and a Poros-HE fraction containing SCC-A. (**E**) Co-dependent activation of *Nanog* transcription requires the XPC complex, the DKC1 complex, and ABCF1. Recombinant purified XPC complex, DKC1 complex, and ABCF1 were added individually (lanes 2-4), or in various combinations (lanes 5-7) to in vitro transcription reactions containing OCT4 and SOX2. Strongest synergistic activation is observed when all three coactivators are present in the transcription reaction (lane 7).

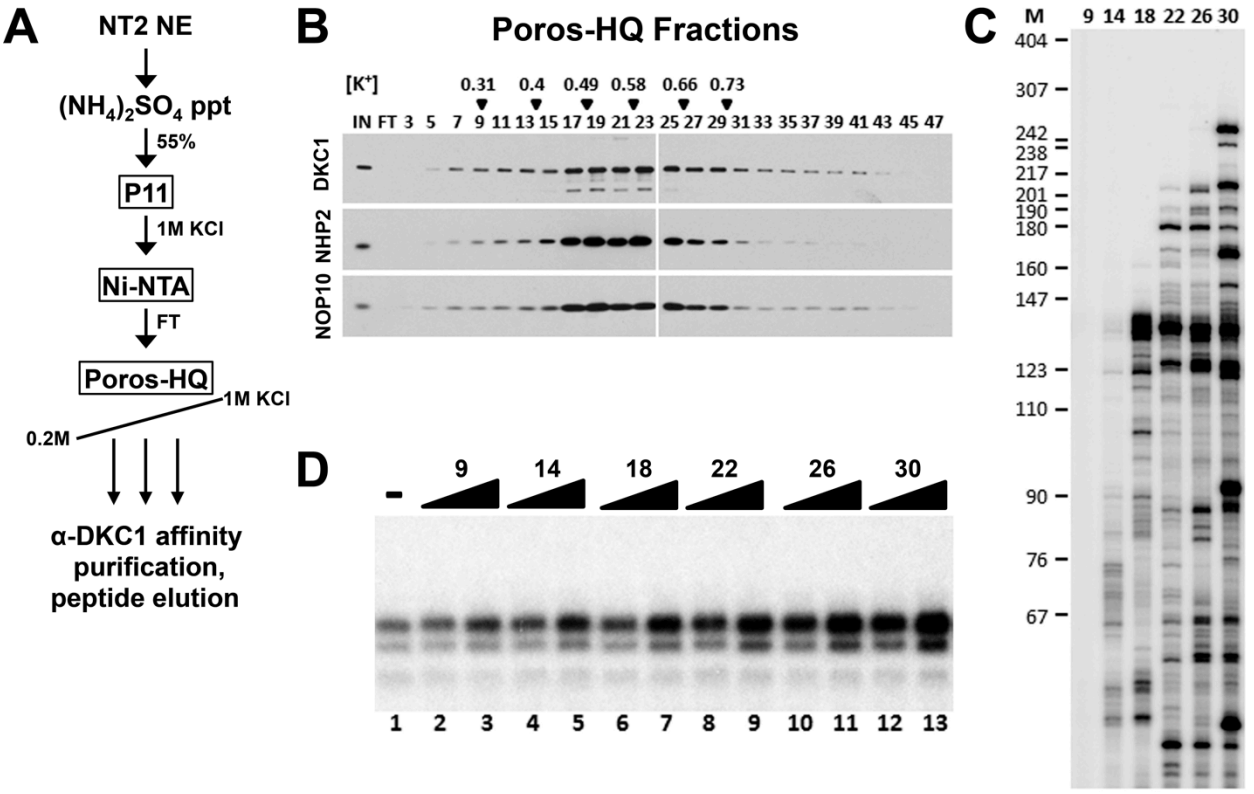

Figure 3. DKC1-associated small RNAs modulate transcriptional coactivator activity. (**A**) Purification scheme of endogenous DKC1 ribonucleoprotein complexes (RNPs). A partially purified fraction (Ni-NTA FT) containing the bulk of the DKC1 RNPs in NT2 cells is fractionated over a Poros-HQ anion exchange column followed by affinity purification using a monoclonal antibody against DKC1 and peptide elution. (**B**) Extensive heterogeneity of the endogenous DKC1 RNPs from NT2 cells. Western blotting of input (IN), flowthrough (FT), and various salt-eluted Poros-HQ fractions using antibodies against DKC1, NHP2, and NOP10. Filled inverted triangles indicate fraction numbers used for affinity purification. Salt concentrations ($[K^+]$ in M) of selected fractions are indicated. (**C**) 5′ end labeling of RNAs isolated from affinity-purified DKC1 RNPs from indicated Poros-HQ fractions. Radiolabeled RNAs were separated on a 6% denaturing urea-polyacrylamide gel. Size markers are in nucleotides. (**D**) Buffer control (−) or affinity-purified DKC1 RNPs from salt-eluted Poros-HQ fractions over a threefold concentration range are assayed using in vitro transcription. Reactions contain OCT4, SOX2, recombinant XPC complex, a Poros-HE fraction containing SCC-B.

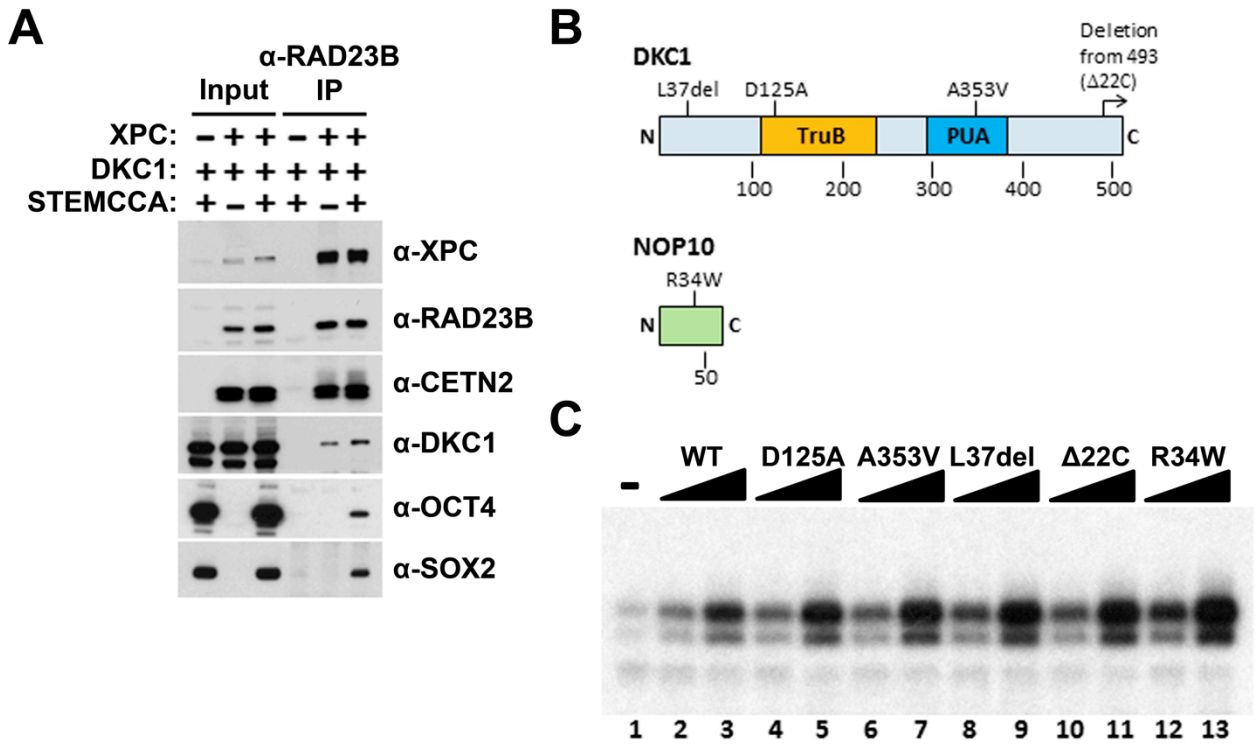

Figure 4. Mechanism of Coactivation by the DKC1 Complex. (**A**) DKC1 interacts with the XPC complex independent of OCT4 and SOX2 in vivo. Control (−), plasmids expressing mouse XPC complex (XPC), mouse DKC1 complex (DKC1), and STEMCCA were co-transfected into 293T cells. Cell lysates are immunoprecipitated with anti-RAD23B antibody. Input extracts (2%) and RAD23B-bound proteins were analyzed by western blotting. (**B**) Schematic diagrams showing the two structural domains in DKC1 (TruB and PUA) and mutations in DKC1 and NOP10 selected for functional analyses in (**C**). All mutations except D125A are identified in patients with dyskeratosis congenita (DC). (**C**) Wild type, pseudouridine synthase inactive (D125A), and DC mutant DKC1 complexes (*Dkc1* A353V, *Dkc1* L37del, *Dkc1* Δ22C, and *Nop10* R34W) are reconstituted in Sf9 cells and assayed over a threefold concentration range in in vitro transcription reactions containing OCT4, SOX2, recombinant XPC complex, and SCC-B.

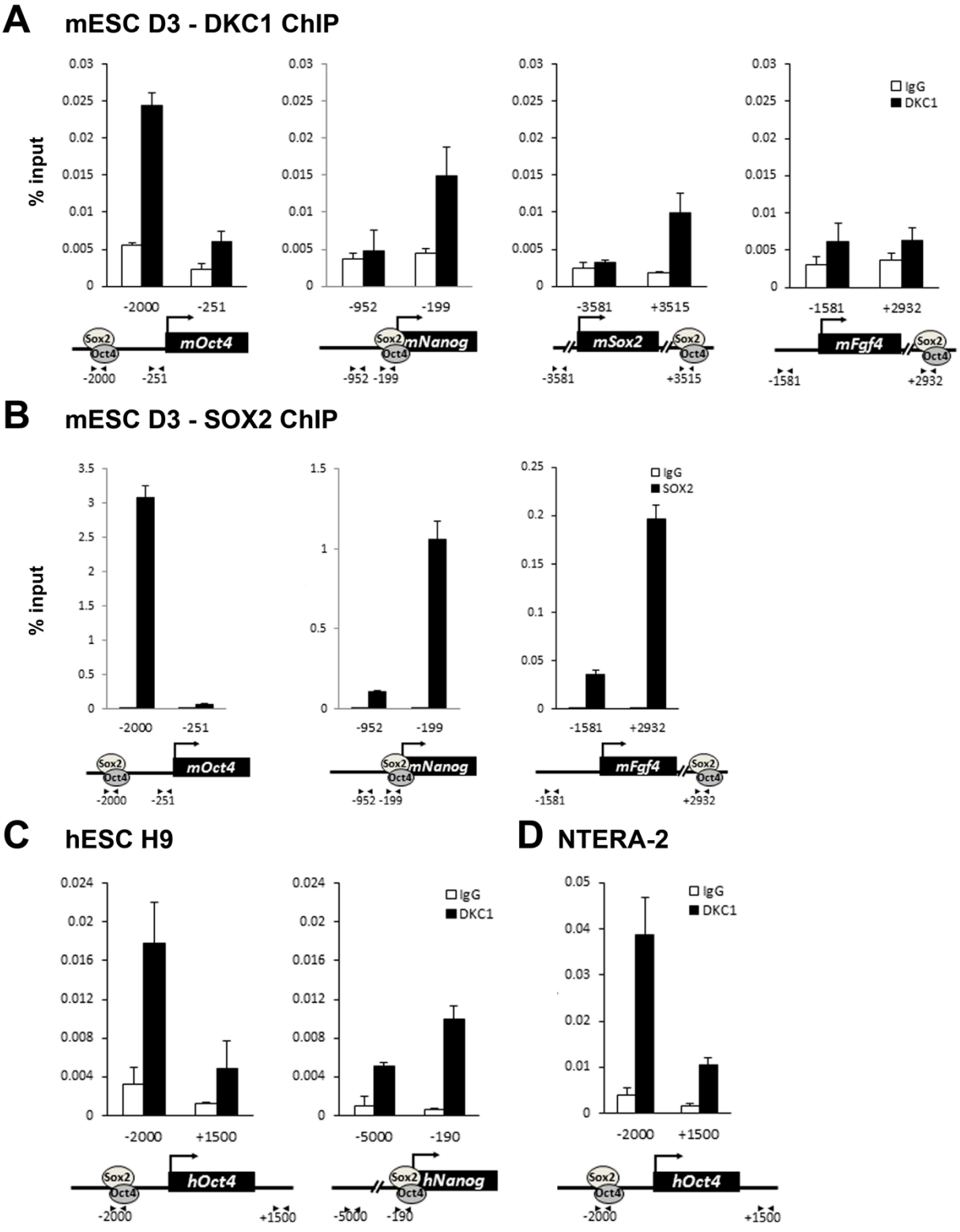
A
mESC D3 - DKC1 ChIP
% input
IgG
DKC1
-2000
-251
mOct4
-952
-199
mNanog
-3581
+3515
mSox2
-1581
+2932
mFgf4
B
mESC D3 - SOX2 ChIP
SOX2
C
hESC H9
+1500
hOct4
-5000
-190
hNanog
D
NTERA-2
Sox2
Oct4

Figure 5. The DKC1 complex is recruited to regulatory regions of key pluripotency genes in mouse and human ES cells. (**A**) Co-occupancy of DKC1, OCT4, and SOX2 on enhancers of *Oct4*, *Nanog*, *Sox2*, but not *Fgf4*, in mouse ES cell line D3. Chromatin immunoprecipitation (ChIP) analysis of DKC1 occupancy on control and enhancer regions of the *Oct4*, *Nanog*, *Sox2*, and *Fgf4* gene loci. Enrichment of DKC1 (black bars) compared to control IgGs (white bars) is analyzed by qPCR and expressed as percentage of input chromatin. Schematic diagrams of OCT4/SOX2 binding sites of each gene and the relative positions of the amplicons used to detect enriched ChIP fragments are shown at the bottom. (**B**) SOX2 is enriched on the regulatory regions of Oct4, Nanog, and Fgf4 in mouse ES cells. Representative ChIP data (n = 3) are analyzed as described in (**A**). (**C**) DKC1 is recruited to regulatory regions of *Oct4* and *Nanog* in human ES cell line H9. Representative ChIP data (n = 3) are analyzed as described in (**A**). Error bars represent standard deviation, n = 3. (**D**) DKC1 is enriched on *Oct4* promoter in human embryonal carcinoma cell line NT2. Representative ChIP data (n = 3) are analyzed as described in (**A**). Error bars represent standard deviation, n = 3.

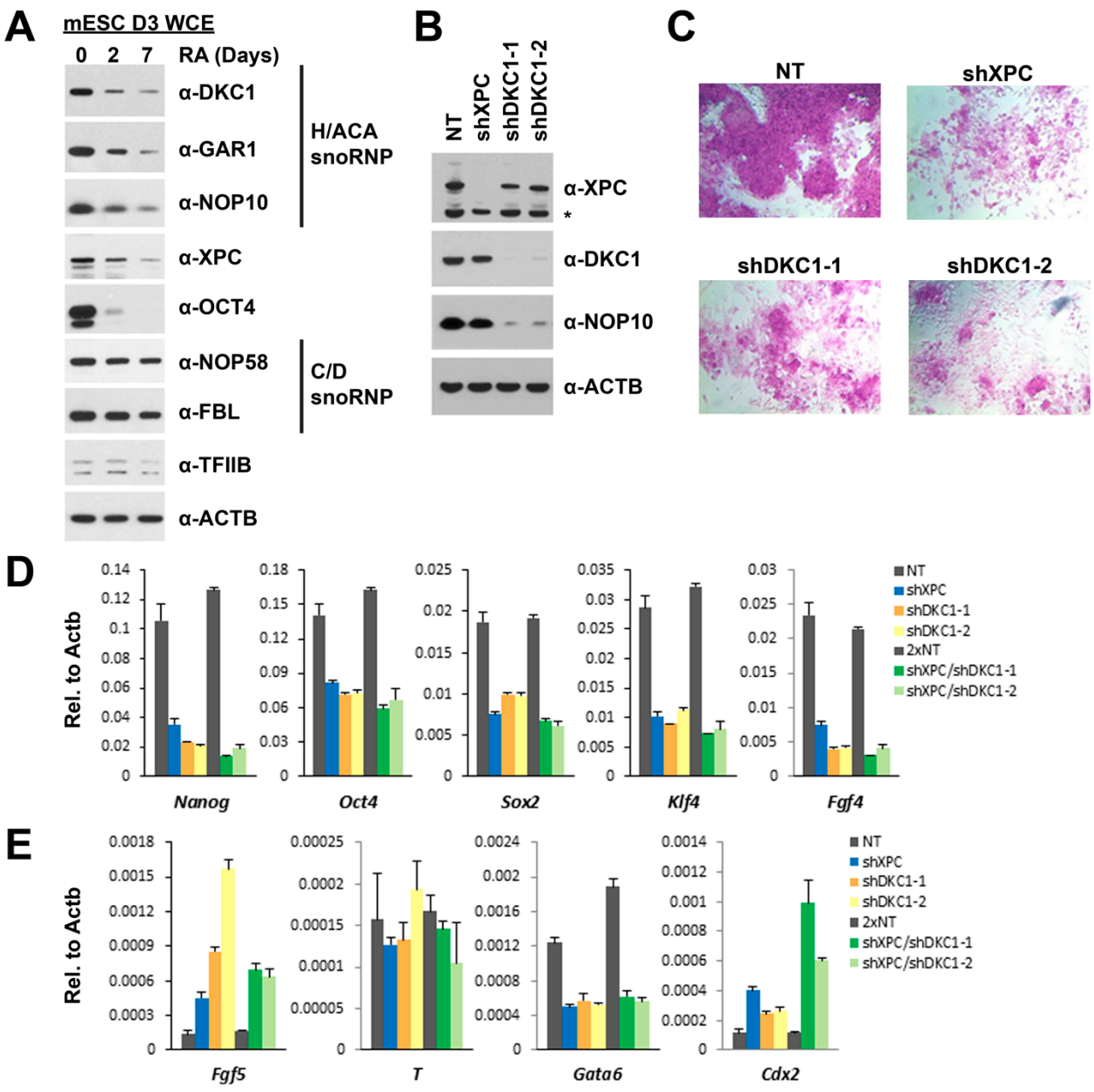
A
mESC D3 WCE
0 2 7 RA (Days)
α-DKC1
α-GAR1
α-NOP10
H/ACA snoRNP
α-XPC
α-OCT4
α-NOP58
α-FBL
C/D snoRNP
α-TFIIB
α-ACTB
B
NT
shXPC
shDKC1-1
shDKC1-2
α-XPC
*
α-DKC1
α-NOP10
α-ACTB
C
NT
shXPC
shDKC1-1
shDKC1-2
D
Rel. to Actb
Nanog
Oct4
Sox2
Klf4
Fgf4
NT
shXPC
shDKC1-1
shDKC1-2
2xNT
shXPC/shDKC1-1
shXPC/shDKC1-2
E
Rel. to Actb
Fgf5
T
Gata6
Cdx2
NT
shXPC
shDKC1-1
shDKC1-2
2xNT
shXPC/shDKC1-1
shXPC/shDKC1-2

Figure 6. The DKC1 complex is required for stem cell maintenance. (**A**) Downregulation of the DKC1 complex upon retinoic acid (RA)-induced differentiation of mouse ES cell line D3. Western blot analyses of whole cell extracts prepared from D3 cells (mESC D3 WCE) collected at indicated days post LIF withdrawal and RA treatment using antibodies against the DKC1 complex (DKC1, GAR1, and NOP10), XPC, OCT4, the NOP58/fibrillarin (FBL) complex, TFIIB, and β-actin as loading control. (**B**) shRNA-mediated knockdown of the DKC1 complex in mouse ES cells. Whole cell extracts of mouse D3 cells infected with control non-target (NT) lentiviruses or with lentiviruses targeting XPC (shXPC) or DKC1 (shDKC1-1 and shDKC1-2) are analyzed by western blotting. MOI = 25. Asterisk denotes non-specific signals. (**C**) ES cell colony morphology and alkaline phosphatase activity are maintained in control non-target shRNA infected D3 cells (NT), but are compromised in XPC (shXPC) and DKC1 depleted cells using two independent shRNAs (shDKC1-1 and shDKC1-2). (**D**) DKC1 and/or XPC depletion in ES cells compromised pluripotency gene expression. Quantification of *Nanog*, *Oct4*, *Sox2*, *Klf4*, and *Fgf4* mRNA levels in single and double knockdown of XPC and DKC1 in D3 cells are analyzed by qPCR and normalized to β-actin (*Actb*). For double knockdown experiments, a cumulative MOI = 50 is used. Data from representative experiments are shown. Error bars represent standard deviation, n = 3. (**E**) DKC1 and/or XPC depletion in ES cells induces spontaneous differentiation towards primitive ectoderm and trophectoderm. Quantification of mRNA levels of primitive ectoderm marker *Fgf5*, mesoderm marker *T*, primitive endoderm marker *Gata6*, and extraembryonic trophectoderm marker *Cdx2* in single and double knockdown of XPC and DKC1 in D3 cells are analyzed as in (**D**).

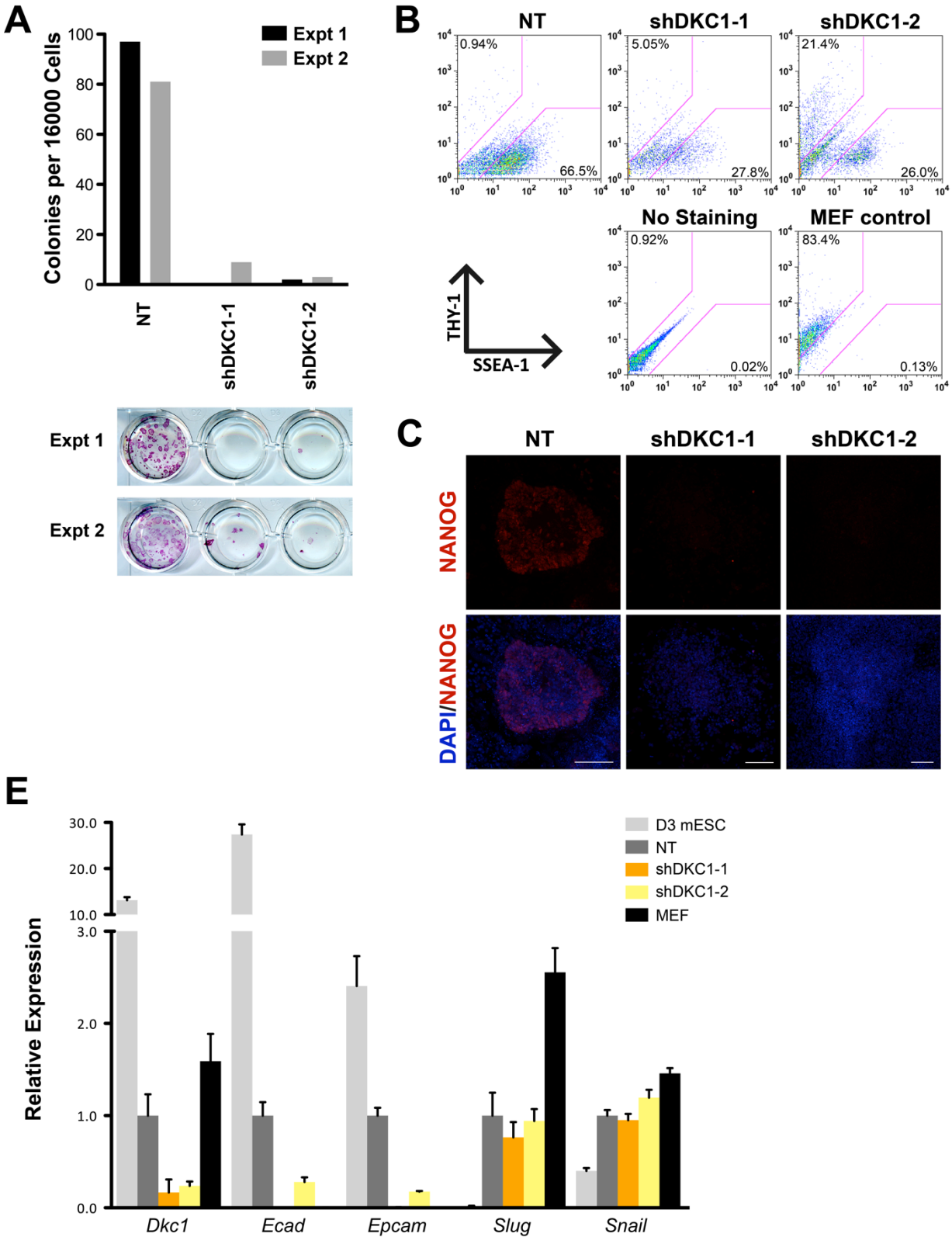
A
Colonies per 16000 Cells
100
80
60
40
20
0
Expt 1
Expt 2
NT
shDKC1-1
shDKC1-2
Expt 1
Expt 2
B
NT
shDKC1-1
shDKC1-2
0.94%
66.5%
5.05%
27.8%
21.4%
26.0%
No Staining
MEF control
0.92%
0.02%
83.4%
0.13%
THY-1
SSEA-1
C
NT
shDKC1-1
shDKC1-2
NANOG
DAPI/NANOG
E
Relative Expression
30.0
20.0
10.0
3.0
2.0
1.0
0.0
D3 mESC
NT
shDKC1-1
shDKC1-2
MEF
Dkc1
Ecad
Epcam
Slug
Snail

Figure 7. The DKC1 complex is required for mesenchymal-to-epithelial transition (MET) during somatic cell reprogramming. (**A**) Depletion of DKC1 blocks somatic cell reprogramming. CF-1 mouse embryonic fibroblasts (MEFs) are infected with lentiviruses expressing OCT4, KLF4, SOX2, and c-MYC (STEMCCA) and reverse tetracycline-controlled transactivator (rtTA) together with control non-target shRNA (NT) or two independent shRNAs targeting DKC1 (shDKC1-1 and shDKC1-2). Infected MEFs are plated onto gelatin coated 24-well plates (Experiment 1) or 24-well plates containing mitomycin-treated feeder MEFs (Experiment 2); cellular reprogramming is initiated by the addition of doxycycline (dox). Cells are stained for AP activity and counted after 14 days (11 days with dox followed by 3 days without dox) post induction (dpi). (**B**) Single cell suspensions of 14 dpi CF-1 MEFs as described in (**A**) are stained with anti-mouse SSEA-1 and anti-THY-1 antibodies and analyzed by flow cytometry. (**C**) Representative confocal images of NANOG stained colonies 17 dpi (14 days with dox followed by 3 days without dox) as described in (**A**). The same acquisition settings—excitation laser intensity, gain, and exposure time—were used for all NANOG images. Scale bar, 100 μm. (**D**) DKC1-depleted MEFs are arrested at the MET during iPS cell generation. Somatic cell reprogramming of CF-1 MEFs is performed as in (**A**). Cells were collected at 14 dpi (11 days with dox followed by 3 days without dox). mRNA levels of *Dkc1*, epithelial markers *Ecad* (also known as *Cdh1*) and *Epcam*, and mesenchymal markers *Slug* and *Snail* are compared with that of uninduced MEFs and mouse ES cell line D3 by qPCR. Values are normalized to expression levels in control non-target knockdown samples. Error bars represent standard deviation, n = 3.

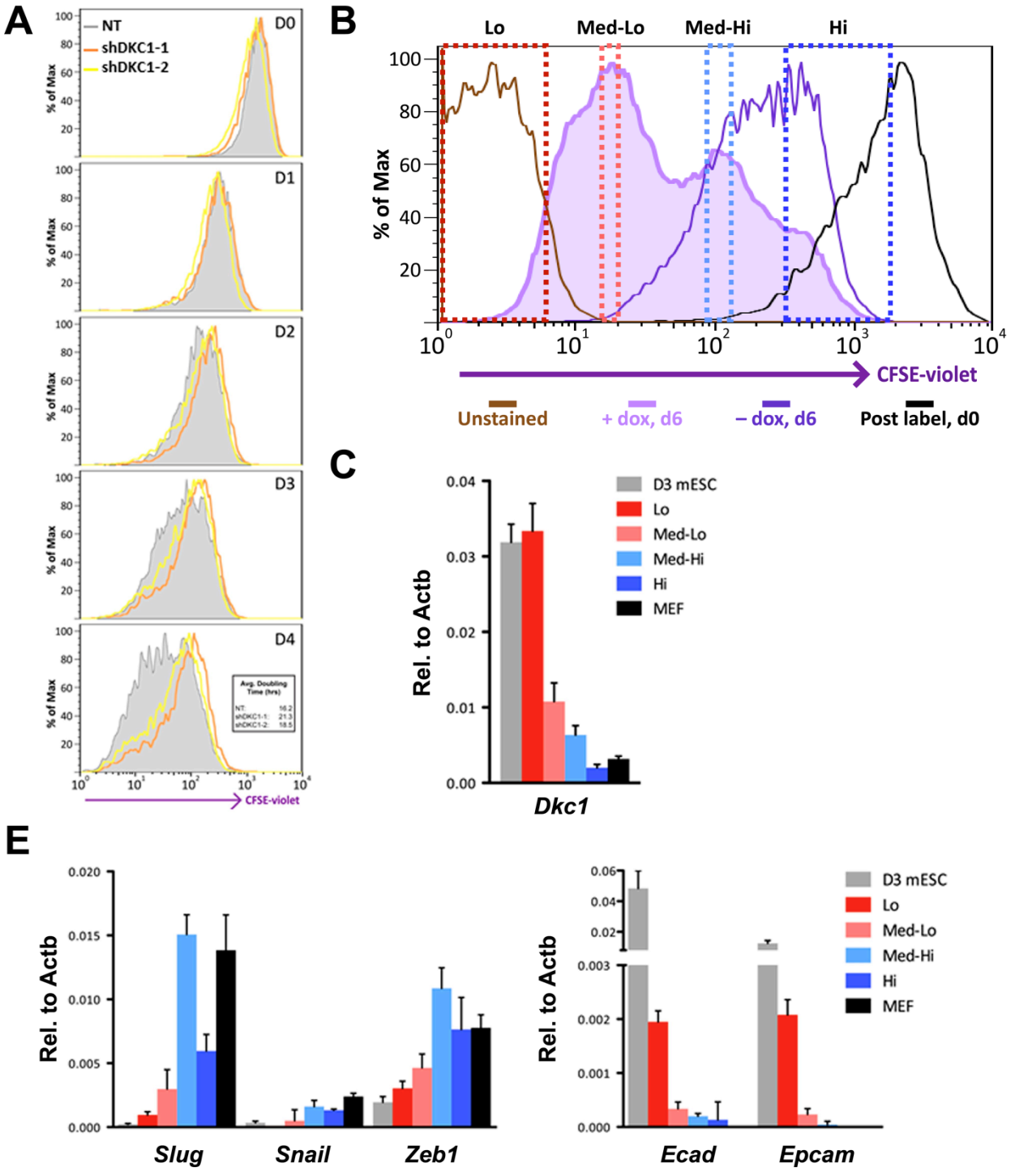
A
NT
shDKC1-1
shDKC1-2
D0
D1
D2
D3
D4
% of Max
CFSE-violet
B
Lo
Med-Lo
Med-Hi
Hi
% of Max
CFSE-violet
Unstained
+ dox, d6
– dox, d6
Post label, d0
C
D3 mESC
Lo
Med-Lo
Med-Hi
Hi
MEF
Rel. to Actb
Dkc1
E
Rel. to Actb
Slug
Snail
Zeb1
Ecad
Epcam

Figure 8. Fast cycling somatic cell state conducive to iPS cell generation requires DKC1. (**A**) MEFs depleted of DKC1 by two independent shRNAs (shDKC1-1 and shDKC1-2), along with MEFs infected with control non-target lentiviruses, were analyzed using the CellTrace CFSE Proliferation Assay (Life Technologies). The doubling time for each population was calculated using the mean fluorescence intensity of each timepoint over 96 hr. (**B**) Induced MEFs (light purple) are treated with dox for 4 days, labeled with CFSE, and continuously cultured in the presence of dox for an additional 48 hr prior to FACS. Populations for ultrafast (Lo), fast (Med-Lo), medium slow (Med-Hi), and slow (Hi) cycling dox-induced MEFs are sorted based on CFSE intensity and denoted by dashed boxes. CFSE-intensity of MEFs immediately after labeling (black), unlabeled MEFs (brown), and uninduced MEFs 48 hr post-labeling (dark purple) are shown and used as controls. (**C**) mRNA levels of *Dkc1* and *Oct4* in sorted MEF populations (Lo-Hi) are compared to D3 ES cells and uninduced MEFs by qPCR. Results are normalized to *Actb*. Error bars represent standard deviation, n = 3. (**D**) Ultrafast (CFSE-Lo) cycling MEFs undergo early MET. mRNA levels of mesenchymal genes (*Slug*, *Snail*, and *Zeb1*; left), and epithelial genes (*Ecad* and *Epcam*; right) in sorted CFSE-labeled cell populations are compared to D3 ES cells and uninduced MEFs by qPCR. Data are analyzed as in (**C**).

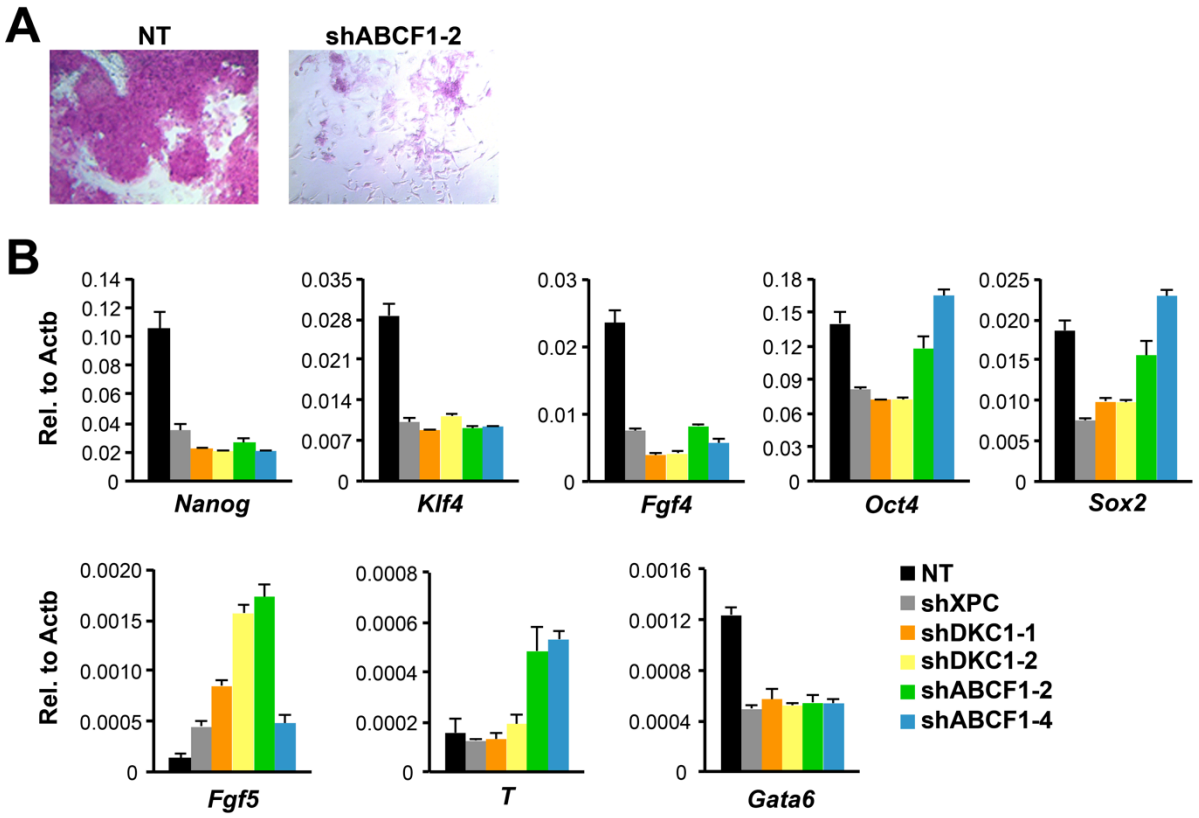

Figure 9. ABCF1 is required for stem cell maintenance. (**A**) ES cell colony morphology and alkaline phosphatase activity are compromised in ABCF1 depleted D3 cells. D3 mESCs are transduced with control non-target (NT) lentiviruses or with lentiviruses targeting ABCF1 (shABCF1-2). MOI = 25. (**B**) XPC, DKC, and ABCF1 depletion comprises pluripotency gene expression and induces expression of some differentiation markers. Quantification of pluripotency genes (*Nanog, Klf4, Fgf4, Oct4,* and *Sox2)*, primitive ectoderm (*Fgf5*), mesoderm (*T*), and primitive endoderm (*Gata6*) mRNA levels are analyzed by qPCR and normalized to β-actin (*Actb*). Error bars represent standard deviation, n = 3.

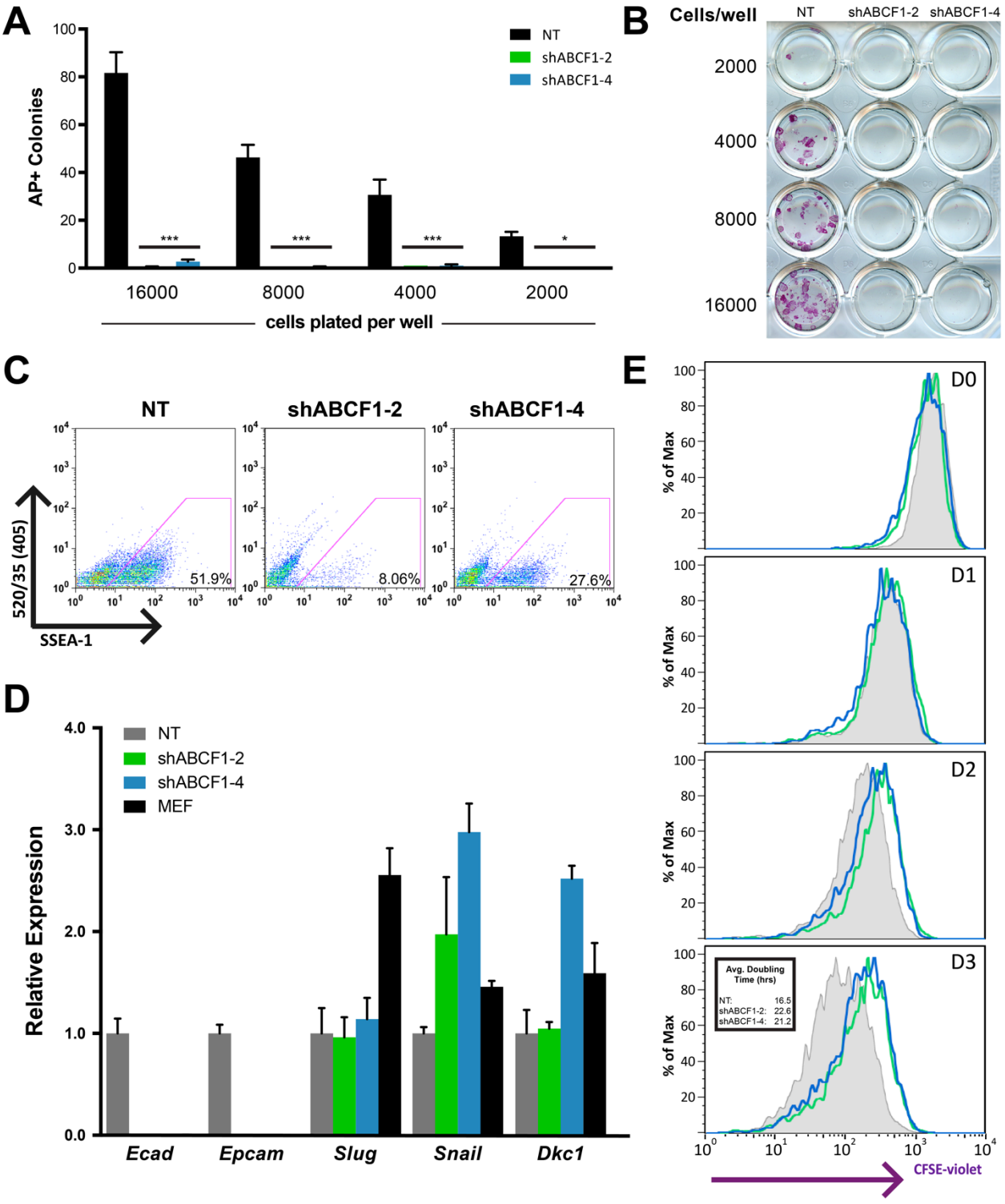
A
AP+ Colonies
NT
shABCF1-2
shABCF1-4
***
***
***
*
16000
8000
4000
2000
cells plated per well
B
Cells/well
NT
shABCF1-2
shABCF1-4
2000
4000
8000
16000
C
NT
shABCF1-2
shABCF1-4
520/35 (405)
SSEA-1
51.9%
8.06%
27.6%
D
Relative Expression
NT
shABCF1-2
shABCF1-4
MEF
Ecad
Epcam
Slug
Snail
Dkc1
E
D0
D1
D2
D3
% of Max
Avg. Doubling Time (hrs)
NT: 16.5
shABCF1-2: 22.6
shABCF1-4: 21.2
CFSE-violet

Figure 10. ABCF1 is required for efficient somatic cell reprogramming. (**A**) Depletion of ABCF1 hinders somatic cell reprogramming. CF-1 MEFs are transduced with lentiviruses expressing either a control non-target shRNA (NT) or two independent shRNAs targeting ABCF1 (shABCF1-2 and shABCF1-4), together with lentiviruses expressing OCT4, KLF4, SOX2, and c-MYC. Cells are plated at the indicated number in 24-well plates; cellular reprogramming is initiated by the addition of doxycycling (dox). Cells are stained for alkaline phosphatase (AP) activity and counted after 14 days (11 days with dox followed by 3 days of dox withdrawal) post induction. Error bars represent the standard deviation, n = 3. *** $P < .001$, * $P < .05$. (**B**) Representative AP staining of experiment shown in (**A**). (**C**) Single cells suspensions of 14 dpi CF-1 MEFs as described in (**A**) are stained with anti-mouse SSEA-1 and analyzed by flow cytometry. (**D**) Epithelial (*Ecad, Epcam*) and mesenchymal marker (*Slug*, *Snail*) expression in ABCF1-depleted CF-1 MEFs 14 dpi, analyzed by qPCR, is consistent with cells failing to undergo MET. mRNA levels are compared with that of uninduced MEFs. Values are normalized to expression levels in control non-target (NT) knockdown samples. Error bars represent standard deviation, n = 3. (**E**) MEFs depleted of ABCF1 have a slower growth rate, prior to iPS induction. The doubling time for each population was calculated using the CellTrace CFSE Prolifersation Assay (Life Technologies), where the mean fluorescence intensity was calculated every 24 hours post labeling, up to 72 hours.

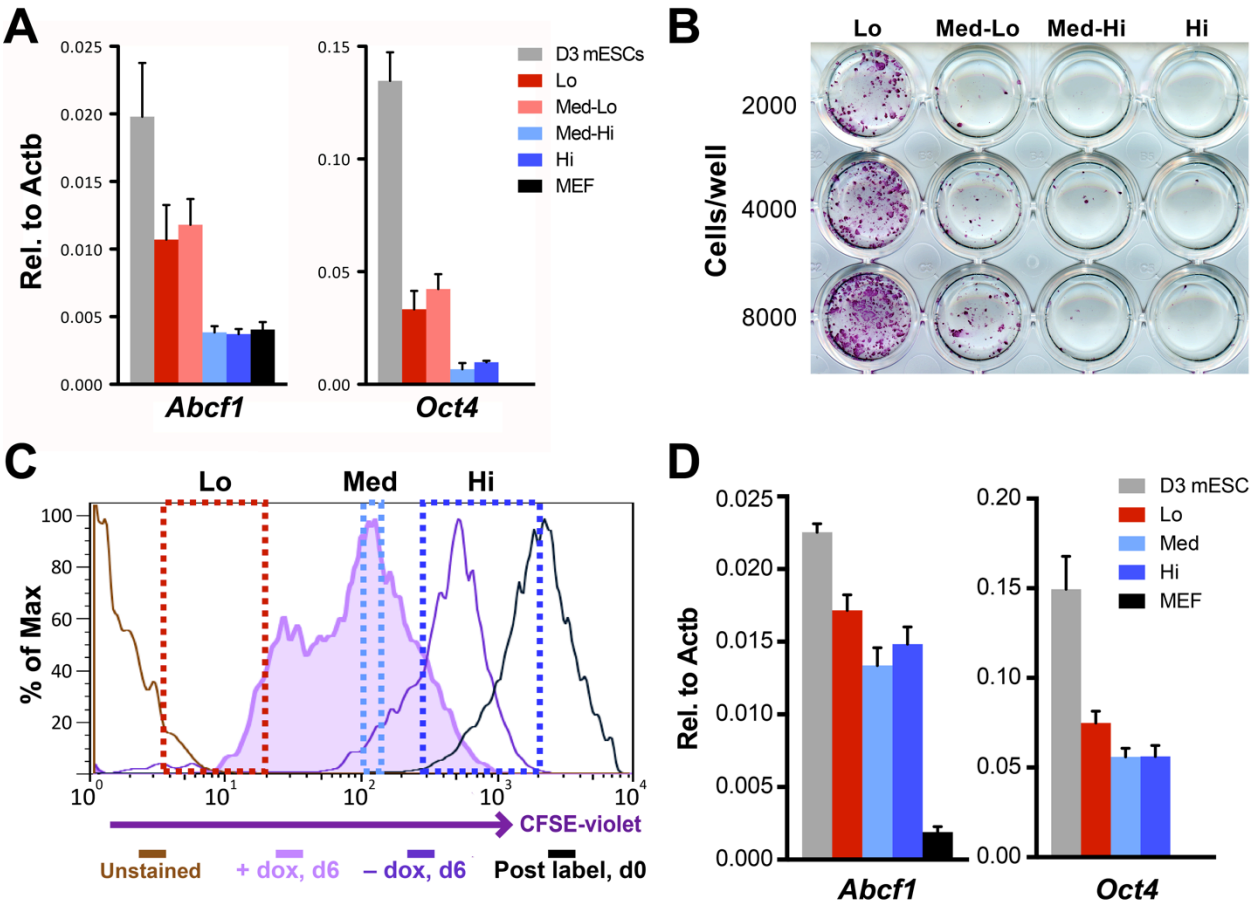

Figure 11. Fast cycling somatic cell state conducive to iPS cell generation is not strictly correlated with *Abcf1* expression. (**A**) mRNA levels of Abcf1 and Oct4 in sorted MEF populations (Lo-Hi) are compared to D3 mES cells and uninduced MEFs by qPCR. Sorted populations are described in Figure 8B. Results are normalized to *Actb*. Error bars represent standard deviation, n = 3. (**B**) Ultrafast cycling MEF population contains the bulk of reprogramming activity. Sorted MEF populations (CFSE-Lo, Med-Lo, Med-Hi, and Hi) are plated on feeders in 24-well plates at indicated cell numbers. AP-positive colonies are stained 14 dpi. (**C**) CFSE labeling and sorting of MEFs carrying an integrated dox-inducible transgene expressing OCT4, KLF4, SOX2, and c-MYC (OKSM MEFs). Populations for ultrafast (Lo), fast (Med), and slow (Hi) cycling dox-induced MEFs are sorted based on CFSE intensity and denoted by dashed boxes. CFSE-intensity of MEFs immediately after labeling (black) and unlabeled MEFs (brown) are shown and used as controls. (D) mRNA levels of Abcf1 and Oct4 in sorted MEF population containing an integrated OKSM cassette by qPCR. Results are normalized to *Actb.* Error bars represent standard deviation, n = 3.

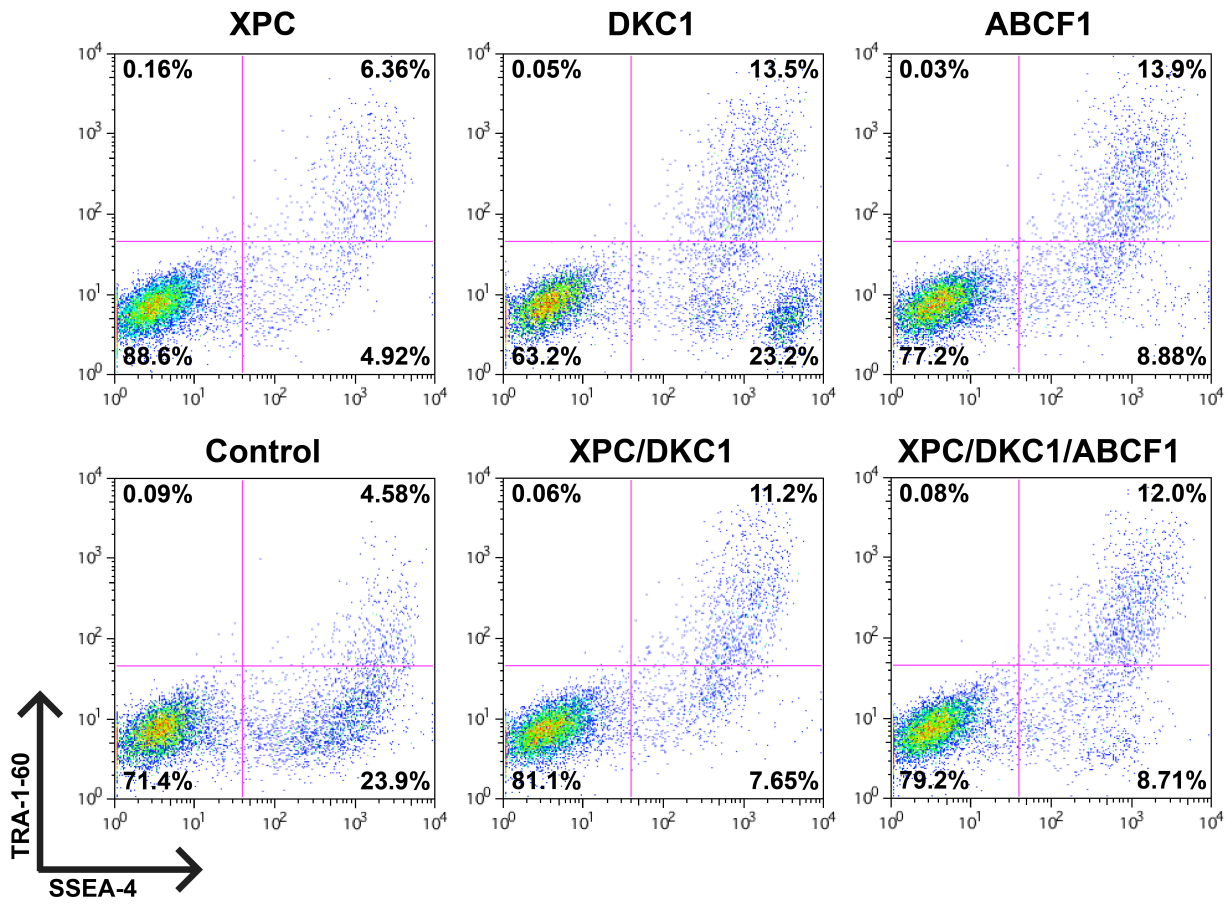

Figure 12. Overexpression of stem cell coactivator complexes (XPC, DKC1, ABCF1) increases the accumulation of late stage iPS cell populations during human iPS reprogramming. Human dermal fibroblasts (HDFs) were nucleofected with constructs containing cDNA for XPC, DKC1, and/or ABCF1 on day 0 of iPS induction. 28 days post induction cells were dissociated and labeled using anti-SSEA-4 and anti-TRA-1-60, early and late stage human iPS cells markers respectively, and analyzed by flow cytometry. These analyses revealed a great % population of late stage TRA-1-60+ iPS cells when reprogramming was done on cells overexpressing one or more of the Stem Cell Coactivator complexes.

Materials and Methods

DNA constructs and antibodies

cDNAs for human and mouse DKC1, GAR1, NHP2, NOP10, and ABCF1 were obtained from cDNA libraries generated from total RNAs isolated from human NTERA-2 (NT2) and mouse ES D3 cells. Mammalian expression plasmids encoding all four subunits of the DKC1 complex were derived from the pHAGE-EF1α-STEMCCA construct (Sommer et al., 2009), wherein OCT4, KLF4, SOX2, and c-MYC were replaced with N-terminal FLAG-tagged DKC1, NHP2, GAR1, and NOP10, respectively (pHAGE-EF1α-DKC1). Expression plasmid for overexpressing the XPC complex (pHAGE-EF1α-XPC) was described (Fong et al., 2011). For expressing the DKC1 complex in insect Sf9 cells, N-terminal His_6- tagged human DKC1 (wild-type and various disease-associated mutants), untagged GAR1, N-terminal FLAG-tagged NHP2, and untagged NOP10 were inserted into a modified pFastBAC Dual vector (Invitrogen, Carlsbad, CA). For expressing partial and holo DKC1 complexes in *E. coli*, untagged human DKC1, N-terminal HA-tagged GAR1, N-terminal FLAG-tagged NHP2, C-terminal His_6-tagged NOP10, and N-terminal His_6- tagged human ABCF1 were cloned into a pST44 polycistronic expression plasmid (Tan et al., 2005). Of note, GAR1 cDNA was reengineered using Quikchange II Site Directed Mutagenesis Kit (Agilent, Santa Clara, CA) to replace codon-pairs for diglycine residues with sequences that are more favorable for translation (Li et al., 2012). For the NAF1-containing intermediate complex, N-terminal HA-tagged NAF1 was inserted in place of HA-GAR1. Polyclonal antibodies against GAR1 (11,711), NHP2 (15,128), FBL (16,021), NOP58 (14,409), and CETN2 (15,877) were purchased from ProteinTech Group; XPC (122A), RAD23B (306A) from Bethyl Laboratories (Montgomery, TX); DKC1 (H-300), TFIIB (C-18), and OCT4 (N-19) from Santa Cruz Biotechnology (Dallas, TX); SOX2 (AB5603) from Millipore (Billerica, MA). Purified rabbit IgGs were purchased from Jackson ImmunoResearch Laboratories (West Grove, PA). Monoclonal antibodies against β-actin (AC-74) were purchased from Sigma Aldrich (St. Louis, MO), DKC1 (H-3) from Santa Cruz Biotechnology, and NOP10 (6547-1) from Epitomics (Burlingame, CA). Anti-FLAG (M2) monoclonal antibodies were purchased from Sigma Aldrich and anti-HA antibodies (MMS-101P) from Covance (Dedham, CA). Antibody against mouse RAD23 was generated in guinea pigs (Fong et al., 2011).

Cell culture

The human embryonal carcinoma NTERA-2 (NT2) cell line was obtained from ATCC. NT2, 293T, and HeLa cells were cultured in DMEM high glucose with GlutaMAX (Invitrogen) supplemented with 10% fetal bovine serum (FBS; HyClone, Piscataway, NJ). Large scale culture of NT2 cells were described (Fong et al., 2011). Mouse ES cell line D3 was purchased from ATCC (Manassas, VA) and adapted to feeder-free condition as described (Fong et al., 2011). Differentiation of D3 cells was induced by maintaining cells in LIF-free ES cell medium containing 2–5 mM all-trans retinoic acid

(Sigma Aldrich) for up to 7 days. Human ES cell line H9 (WiCell, Madison, WI) was maintained in feeder-independent conditions, using Synthemax SC-II Substrate (Corning) and grown in TeSR-E8 (Stemcell Technologies, Canada). Media was changed daily and cell cultures were passaged using Dispase (Stemcell Technologies), according to the manufacturer's protocol.

Purification of SCC-A/SCC-B

All steps were performed at 4°C. Nuclear extracts were prepared from 400 l of NT2 cells. Partially purified P11-phosphocellulose 1 M KCl and Ni-NTA flowthrough (Ni-FT) fractions were prepared as described (Fong et al., 2011). The Ni-FT fraction was dialyzed against buffer D at 0.2 M KCl with 0.0025% NP-40 and 10% glycerol (all buffers from then on contained 0.0025% NP-40 and 10% glycerol unless otherwise stated). This Ni-FT fraction was applied to a Poros 20 HQ column (Applied Biosystems, Carlsbad, CA), subjected to a 4 column volume (CV) linear gradient from 0.2 M to 0.4 M KCl (Q0.3), washed at 0.52 M KCl, and developed with a 13 CV linear gradient from 0.52 M to 1.0 M KCl. Transcriptionally active Q0.3 fraction (0.32–0.4 M) were pooled and applied directly to hydroxyapatite (HAP) type II ceramic resin (Bio-Rad, Hercules, CA), washed first at 0.38 M, then lowered to 0.1 M KCl in 3 CV. HAP column buffer was then exchanged and washed extensively with buffer D at 0.03 M KPi, pH 6.8 without KCl and NP-40. The HAP column was subjected to a 20 CV linear gradient from 0.03 M to 0.6 M KPi. Active HAP fractions eluting from 0.2–0.3 M KPi were pooled and separated on a Superose 6 XK 16/70 gel filtration column (130 ml, GE Healthcare, Piscataway, NJ) equilibrated with buffer D + 0.1 mM EDTA at 0.15 M KCl. Active Superose 6 fractions with an apparent molecular mass of 400–600 kDa were pooled and supplemented with 0.25 mg/ml insulin (Roche, Indianapolis, IN). Pooled fractions were applied to a Poros 20 HE column (Applied Biosystems) equilibrated in buffer D + 0.1 mM EDTA at 0.15 M KCl, subjected to a 34 CV linear gradient from 0.15 M to 1 M KCl. SCC-A containing HE fractions eluted from 0.56–0.62 M KCl. For affinity purification of endogenous DKC1 complexes, Ni-FT derived from 200 l of NT2 cells was applied to a Poros 20 HQ column, subjected to a 22 CV linear gradient from 0.2 M to 1 M KCl. Fractions with low levels of DKC1 were first concentrated using a Spin-X UF concentrator (Corning, Tewksbury, MA) before they were used for immune-affinity purification. Various Poros 20 HQ fractions (adjusted to 0.05% NP-40) were incubated with 10 µg of anti-DKC1 monoclonal antibody immobilized on Protein G Sepharose (GE Healthcare) for 16 hr in the presence of RNase inhibitors (RNasin Plus, Promega, Madison, WI), washed extensively with 0.6 M KCl HEMG buffer (25 mM HEPES, pH 7.9, 0.1 mM EDTA, 12.5 mM $MgCl_2$, 10% glycerol) with 0.2% NP-40, then equilibrated with 0.3 M KCl HEMG with 0.1% NP-40 before elution with peptides.

Mass spectrometry analysis

Peak Poros 20 Heparin fractions were pooled, concentrated using a Spin-X centrifugal concentrator, separated by SDS-PAGE, stained, protein bands excised, digested with trypsin, and extracted. Peptide pools from each gel slice were analyzed by matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF MS; Bruker Reflex III). Selected mass values were used to search protein databases linked to PROWL (Rockefeller University) using ProFound and protein databases linked to ExPASy (Swiss Institute of Bioinformatics, Geneva) using PeptIdent.

In vitro transcription assay

In vitro transcription reactions, DNA template, purification of activators OCT4 and SOX2, general transcription factors, RNA polymerase, and recombinant XPC complex were described (Fong et al., 2011).

5′ end radiolabeling of RNA

DKC1-associated small RNAs were isolated using TRIzol reagent (Life Technologies, Carlsbad, CA). RNAs were treated with tobacco acid pyrophosphatase (TAP) (Epicentre, Madison, WI) to remove the 5′ m7G cap followed by dephosphorylation with APex Alkaline Phosphatase (Epicentre). Purified RNAs were labeled with T4 polynucleotide kinase (PNK) (New England Biolabs, Ipswich, MA) and γ-^{32}P-ATP in the presence of RNase inhibitors (RNAsin Plus, Promega) at 37°C for 1.5 hr. RNAs were precipitated and washed with 75% ethanol to remove free γ-^{32}P-ATP. Labeled RNAs were separated on a 6% denaturing Urea-polyacrylamide gel, and visualized by radiography.

Reconstitution and purification of the DKC1 complexes

Recombinant Bacmid DNAs for expressing wild-type and mutant DKC1 complexes were generated from pFastBAC constructs (described above) according to manufacturer's instructions (Invitrogen). Recombinant baculovirus for the infection of Sf9 cells was generated using the Bac-to-Bac Baculovirus Expression System (Invitrogen). Baculoviruses were amplified three times in Sf9 cells. 2 l of Sf9 cells (~2 × 10^6/ml) were infected with baculoviruses, collected at 48 hr post infection, washed once with ice-cold PBS, lysed in six packed cell volume of 0.3 M NaCl buffer HGN (50 mM HEPES, pH 7.9, 10% glycerol, 0.5% NP-40), and sonicated briefly. Cleared lysate was supplemented with 10 mM imidazole and incubated with Ni-NTA resin pre-equilibrated with 0.5 NaCl HGN and 10 mM imidazole for 16 hr. Resin slurries were poured into gravity columns, washed with 0.5 NaCl HGN (0.1% NP-40) with 20 mM imidazole, and bound DKC1 complexes were eluted with buffer 0.3 M NaCl HGN (0.1% NP-40)

containing 0.25 M Imidazole. Peak fractions were loaded immediately to a gravity column containing Heparin Sepharose 6 Fast Flow (GE Healthcare) pre-equilibrated with 0.3 M NaCl HEGN (25 mM HEPES, pH 7.9, 0.1 mM EDTA, 10% glycerol, 0.02% NP-40). Column was washed extensively at 0.3 M NaCl HEGN, then with 0.5 M NaCl HEGN. The DKC1 complexes were eluted with 1 M NaCl HEGN. Peak fractions containing all four subunits of the DKC1 complex, as determined by western blotting, were pooled and incubated with anti-FLAG (M2) agarose (Sigma Aldrich) for 3–4 hr, washed at 0.5 M NaCl HEGN and re-equilibrated with micrococcal nuclease (MNase) digestion buffer (25 mM Tris–HCl, pH 7.9, 20 mM NaCl, 60 mM KCl, 2 mM $CaCl_2$, 0.01% NP-40, 10% glycerol). Bound DKC1 complexes were treated with 300 U of MNase (Thermo Scientific, Waltham, MA) or buffer at room temperature and nutated for 1 hr. MNase digestion was terminated with 20 mM EGTA. Mock and MNase-treated DKC1 complexes were washed extensively with 0.6 M NaCl HEMG with 0.2% NP-40 and 20 mM EGTA and equilibrated with 0.3 M NaCl HEMG with 0.1% NP-40 followed by FLAG peptide elution. For purification of bacterial DKC1 complexes, pST44 expression plasmids were transformed into BL21-Codon Plus RIPL competent cells (Agilent). Expression of hetero-dimeric (FLAG-DKC1/NOP10-His_6), -trimeric (untagged DKC1/FLAG-NHP2/ NOP10-His_6), holo (untagged DKC1/HA-GAR1/FLAG-NHP2/ NOP10-His6) DKC1 complexes as well as NAF1-containing intermediate DKC1 complex (untagged DKC1/HA-NAF1/FLAG-NHP2/ NOP10-His_6) were induced at 30°C for 4 hr with 0.5 mM IPTG. Cell pellets were lysed in high salt lysis buffer HSLB (50 mM Tris–HCl pH 7.9, 0.5 M NaCl, 0.6% TritonX-100, 0.05% NP-40, 10% glycerol) with imidazole (10 mM) and lysozyme (0.5 mg/ml). Sonicated lysates were cleared by ultracentrifugation and incubated with Ni-NTA resin for 16 hr. Bound proteins were washed extensively with HSLB with 20 mM imidazole, equilibrated with 0.25 M NaCl HGN (25 mM HEPES, pH 7.9, 10% glycerol, 0.01% NP-40) with 20 mM imidazole, and eluted with 0.25 M imidazole in 0.25 M NaCl HGN. Peak fractions were pooled and applied to a Poros 50 Heparin (HE) column, washed extensively with 0.25 M and 0.5 M NaCl HGN, and subjected to a 4 CV linear gradient from 0.5 M to 1 M NaCl. Fractions containing the desired subunits of the DKC1 complexes were detected by western blotting, pooled, and incubated with anti-FLAG agarose for 3–4 hr at 4°C. Bound proteins were washed extensively at 0.7 M NaCl HGN with 0.1% NP-40 and re-equilibrated with 0.3 M NaCl HGN with 0.1% NP-40 before elution with FLAG peptides. For holo and NAF1-containing DKC1 complexes, HE fractions were first incubated with anti-HA resin, washed and eluted with HA peptides before proceeding to the anti-FLAG affinity immunoprecipitation step as described. 32, 25, 5, and 2 l of *E. coli* cultures were required to generate ~0.5 µg of purified holo DKC1, NAF1 intermediate, hetero-trimeric, and–dimeric complexes, respectively.

Coimmunoprecipitation assay

pHAGE-EF1α-STEMCCA, pHAGE-EF1α-mXPC, and pHAGE-EF1α-mDKC1 expression plasmids were co-transfected into 293T cells using Lipofectamine 2000 (Invitrogen). Transfected cells on 10 cm dishes were lysed directly on plates with 1 ml of lysis buffer (200 mM NaCl, 50 mM HEPES-KOH, pH 7.9, 0.1 mM EDTA, 0.5% NP-40

and 10% glycerol) 40 hr post-transfection. Cell lysates were collected and homogenized by passing through a 25-gauge needle five times. Lysates were cleared by centrifugation at 15k rpm for 25 min at 4°C. 3 µg of anti-RAD23B antibodies were coupled to Protein A sepharose (GE Healthcare) in PBS containing 0.05% NP-40 for 1 hr at room temperature. Antibody-coupled beads were washed and equilibrated with lysis buffer before incubating with 0.5 ml of cleared cell lysates for 16 hr at 4°C. Sepharose beads were then washed extensively with lysis buffer and bound proteins were eluted with SDS/sample buffer and analyzed by western blotting.

shRNA-mediated knockdown of DKC1 and ABCF1 by lentiviral infection

For lentivirus production, non-target control and pLKO plasmids targeting mouse DKC1, ABCF1, and XPC were co-transfected with packaging vectors into 293T cells using lipofectamine 2000 (Invitrogen). Supernatants were collected at 48 hrs and 72 hrs post transduction. Virus preparation, titer determination, and infection of D3 mouse ES cells were performed as described (Fong et al., 2011), except at a multiplicity of infection (MOI) of 25. For knockdown reprogramming experiments, MEFs were transduced at a MOI of 5 prior to iPS cell induction. Detection of alkaline phosphatase activity of knockdown ES cells was carried out using a commercial kit (Millipore).

Chromatin immunoprecipitation

Mouse ES cell line D3 and human ES cell line H9 were first crosslinked with ethylene glycol bis[succinimidylsuccinate] (EGS, 3 mM, Pierce) for 30 min and then with formaldehyde (1%) for 5 min in fixing buffer (50 mM HEPES, pH 7.5, 0.1 M NaCl, 1 mM EDTA, 0.5 mM EGTA) to capture protein–protein and protein-DNA interactions (Zeng et al., 2006). Crosslinking was then terminated by glycine (0.125 M). Cells were washed twice with PBS, scraped, and centrifuged at 150×g for 5 min at 4°C, resuspended in lysis buffer (50 mM HEPES, pH 7.9, 0.14 M NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) with Halt Protease Inhibitor Cocktail (Pierce, Waltham, MA), and nutated at 4°C for 10 min. Nuclei were pelleted at 1700×g for 5 min, washed twice with wash buffer (10 mM Tris–HCl, pH 8.1, 0.2 M NaCl, 1 mM EDTA, 0.5 mM EGTA) and twice with shearing buffer (0.1% SDS, 1 mM EDTA, 10 mM Tris–HCl, pH 8.1). Nuclei were resuspended in shearing buffer, transferred to Covaris TC 12 × 12 mm tubes with AFA Fiber, and sonicated with a Covaris S2 Focused Ultrasonicator to obtain DNA fragments averaging 300–500 bp in length. Cleared chromatin extracts were adjusted to 0.15 M NaCl and 1% Triton X-100 and immunoprecipitated overnight at 4°C with 3 µg of purified rabbit IgGs or anti-DKC1 antibody. Immunoprecipitated DNA was captured with pre-equilibrated Protein A sepharose (GE Healthcare), washed extensively with high salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM HEPES, pH 7.9, 0.5 M NaCl), LiCl wash buffer (100 mM Tris–HCl, pH 7.5, 0.5 M LiCl, 1% NP-40, 1% sodium deoxycholate), and TE buffer (10 mM Tris–HCl, pH 8.0, 0.1 mM EDTA). Supernatant from control IgG immunoprecipitates was saved as input. Input

chromatin and immunoprecipitated DNA were reversed crosslinked overnight at 50°C with Proteinase K (Invitrogen), RNase A (Thermo Scientific), and 0.3 M NaCl. DNA was purified using a Qiaquick PCR Purification Kit (Qiagen, Netherlands). Purified DNA was quantified by real time PCR with SYBR Select Master Mix for CFX (Life Technologies) and gene specific primers using a CFX Touch Real-Time PCR Detection System (Bio-Rad). Gene specific primer sequences are provided in the appendix.

RNA isolation, reverse transcription, and real time PCR analysis

Total RNA was extracted and purified using TRIzol reagent (Life Technologies) followed by DNase I treatment (Invitrogen). cDNA synthesis was performed with 1 µg of total RNA using iScript cDNA Synthesis Kit (Bio-Rad) and diluted 10-fold. Real time PCR analysis was carried out with SYBR Select Master Mix for CFX (Life Technologies) using the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Results were normalized to β-actin. Gene specific primer sequences are provided in the appendix.

Somatic cell reprogramming and flow cytometry

CF-1 MEFs (Charles River, Wilmington, MA) were transduced with inducible STEMCCA and rtTA lentivirus-containing supernatants overnight in 8 µg/ml polybrene (Sigma Aldrich). Alternatively, MEFs isolated from mice carrying an integrated dox-inducible transgene expressing OCT4, KLF4, SOX2, and c-MYC (Jackson Laboratories, Bar Habor, ME) were also used. Doxycycline (Sigma Aldrich; 2 µg/ml) was supplemented to complete mouse ES cell media to induce expression of OKSM. Reprogramming was assayed by alkaline phosphatase staining (Millipore), NANOG staining (Abcam, United Kingdom, ab80892), or by flow cytometry analysis using anti-CD90.2/Thy1.2 (Biolegends, San Diego, CA) and anti-SSEA1 (Biolegends, San Diego, CA) on a BD LSRFortessa, performed according to the manufacturers' protocols.

CFSE labeling of mouse embryonic fibroblasts

To determine the doubling time, cells were labeled with CFSE-Violet (Life Technologies) at a working concentration of 5.0 µM, according to the manufacturers' protocol. Cells were analyzed for remaining fluorescence on a BD LSRFortessa every day for 3-4 days. Induced MEFs were labeled with CFSE-Violet (Life Technologies) at a working concentration of 7.5 µM, as described in Guo et al., 2014. CFSE-labeled MEFs were sorted into distinct fast to slow dividing populations at the UC Berkeley Li Ka Shing Flow Cytometry Facility. MEFs cultured in the absence of doxycycline were used as controls.

REFERENCES

Agarwal, S., Loh, Y.-H., McLoughlin, E.M., Huang, J., Park, I.-H., Miller, J.D., Huo, H., Okuka, M., Dos Reis, R.M., Loewer, S., et al. (2010). Telomere elongation in induced pluripotent stem cells from dyskeratosis congenita patients. Nature *464*, 292–296.

Aladjem, M.I., Spike, B.T., Rodewald, L.W., Hope, T.J., Klemm, M., Jaenisch, R., and Wahl, G.M. (1998). ES cells do not activate p53-dependent stress responses and undergo p53- independent apoptosis in response to DNA damage. Curr Biol *8*, 145–155.

Alawi, F., Lin, P., Ziober, B., and Patel, R. (2011). Dyskerin expression correlates with active proliferation independently of telomerase. Head Neck *33*, 1041–1051.

Ambrosetti, D.C., Schöler, H.R., Dailey, L., and Basilico, C. (2000). Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. J. Biol. Chem. *275*, 23387–23397.

Ang, Y.S., Tsai, S.Y., Lee, D.F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., et al. (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell *145*, 183–187.

Anokye-Danso, F., Trivedi, C.M., Juhr, D., Gupta, M., Cui, Z., Tian, Y., Zhang, Y., Yang, W., Gruber, P.J., Epstein, J.A., et al. (2011). Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. Cell Stem Cell *8*, 376–388.

Araki, M., Masutani, C., Takemura, M., Uchida, A., Sugasawa, K., Kondoh, J., Ohkuma, Y., and Hanaoka, F. (2001). Centrosome Protein Centrin 2/Caltractin 1 Is Part of the Xeroderma Pigmentosum Group C Complex That Initiates Global Genome Nucleotide Excision Repair. J. Biol. Chem. *276*, 18665–18672.

Avilion, A. a, Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. *17*, 126–140.

Avvakumov, G. V, Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature *455*, 822–825.

Ballarino, M., Morlando, M., Pagano, F., Fatica, A., and Bozzoni, I. (2005). The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in Saccharomyces cerevisiae. Mol. Cell. Biol. *25*, 5396–5403.

Banito, A., Rashid, S.T., Acosta, J.C., Banito, A., Rashid, S.T., Acosta, J.C., Li,

S., Pereira, C.F., Geti, I., Pinho, S., et al. (2009). Senescence impairs successful reprogramming to pluripotent stem cells. 2134–2139.

Bar-Nur, O., Brumbaugh, J., Verheul, C., Apostolou, E., Pruteanu-Malinici, I., Walsh, R.M., Ramaswamy, S., and Hochedlinger, K. (2014). Small molecules facilitate rapid and synchronous iPSC generation. Nat. Methods *11*, 1170–1176.

Barrès, R., Osler, M.E., Yan, J., Rune, A., Fritz, T., Caidahl, K., Krook, A., and Zierath, J.R. (2009). Non-CpG Methylation of the PGC-1α Promoter through DNMT3B Controls Mitochondrial Density. Cell Metab. *10*, 189–198.

Barreto, G., Schäfer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Döderlein, G., Maltry, N., Wu, W., Lyko, F., et al. (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature *445*, 671–675.

Batista, L.F.Z., Pech, M.F., Zhong, F.L., Nguyen, H.N., Xie, K.T., Zaug, A.J., Crary, S.M., Choi, J., Sebastiano, V., Cherry, A., et al. (2011). Telomere shortening and loss of self-renewal in dyskeratosis congenita induced pluripotent stem cells. Nature *474*, 399–402.

Bell, a C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482–485.

Bernardes de Jesus, B.M., Bjørås, M., Coin, F., and Egly, J.M. (2008). Dissection of the molecular defects caused by pathogenic mutations in the DNA repair factor XPC. Mol. Cell. Biol. *28*, 7225–7235.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042–1047.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Blasco, M., Lee, H.W., Hande, M.P., Samper, E., Lansdorp, P.M., DePinho, R. a, and Greider, C.W. (1997). Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. Cell *91*, 25–34.

Blau, H.M., Chiu, C.P., and Webster, C. (1983). Cytoplasmic activation of human nuclear genes in stable heterocaryons. Cell *32*, 1171–1180.

Boorstein, R.J., Cummings, A., Marenstein, D.R., Chan, M.K., Ma, Y., Neubert, T.A., Brown, S.M., and Teebor, G.W. (2001). Definitive Identification of Mammalian 5-Hydroxymethyluracil DNA N-Glycosylase Activity as SMUG1. J. Biol. Chem. *276*, 41991–41997.

Bortvin, A., Eggan, K., Skaletsky, H., Akutsu, H., Berry, D.L., Yanagimachi, R., Page, D.C., and Jaenisch, R. (2003). Incomplete reactivation of Oct4-related genes in

mouse embryos cloned from somatic nuclei. Development *130*, 1673–1680.

Bostick, M., Kim, J.K., Estève, P.-O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760–1764.

Boyer, L. a, Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1994). Sp1 elements protect a CpG island from de novo methylation. Nature *371*, 435–438.

Bransteitter, R., Pham, P., Scharff, M.D., and Goodman, M.F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc. Natl. Acad. Sci. U. S. A. *100*, 4102–4107.

Brook, F. a, and Gardner, R.L. (1997). The origin and efficient derivation of embryonic stem cells in the mouse. Proc. Natl. Acad. Sci. U. S. A. *94*, 5709–5712.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., Van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209–1222.

Buganim, Y., Faddah, D. a, and Jaenisch, R. (2013). Mechanisms and models of somatic cell reprogramming. Nat. Rev. Genet. *14*, 427–439.

Bunick, C.G., Miller, M.R., Fuller, B.E., Fanning, E., and Chazin, W.J. (2006). Biochemical and structural domain analysis of xeroderma pigmentosum complementation group C protein. Biochemistry *45*, 14965–14979.

Cardoso, M.C., and Leonhardt, H. (1999). DNA methyltransferase is actively retained in the cytoplasm during early development. J. Cell Biol. *147*, 25–32.

Carey, B.W., Markoulaki, S., Beard, C., Hanna, J., and Jaenisch, R. (2010). Single-gene transgenic mouse strains for reprogramming adult somatic cells. Nat. Methods *7*, 56–59.

Carlson, L.L., Page, A.W., and Bestor, T.H. (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. Genes Dev. *6*, 2536–2541.

De Carvalho, D.D., You, J.S., and Jones, P. a (2010). DNA methylation and cellular reprogramming. Trends Cell Biol. *20*, 609–617.

Cattoglio, C., Zhang, E.T., Grubisic, I., Chiba, K., Fong, Y.W., and Tjian, R.

(2015). Functional and mechanistic studies of XPC DNA-repair complex as transcriptional coactivator in embryonic stem cells. Proc. Natl. Acad. Sci. U. S. A. *112*, E2317–E2326.

Chambers, I., and Tomlinson, S.R. (2009). The transcriptional foundation of pluripotency. Development *136*, 2311–2322.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643–655.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chen, L., and Madura, K. (2002). Rad23 promotes the targeting of proteolytic substrates to the proteasome. Mol. Cell. Biol. *22*, 4902–4913.

Chen, L., Macmilan, A.M., and Verdine, G.L. (1993). Mutational Separation of DNA Binding from Catalysis in a DNA Cytosine Methyltransferase. J. Am. Chem. Soc. *115*, 5318–5319.

Chen, T., Yuan, D., Wei, B., Jiang, J., Kang, J., Ling, K., Gu, Y., Li, J., Xiao, L., and Pei, G. (2010). E-cadherin-mediated cell-cell contact is critical for induced pluripotent stem cell generation. Stem Cells *28*, 1315–1325.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. Cell *133*, 1106–1117.

Chia, N.-Y., Chan, Y.-S., Feng, B., Lu, X., Orlov, Y.L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.-S., et al. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature *468*, 316–320.

Chin, M.H., Mason, M.J., Xie, W., Volinia, S., Singer, M., Peterson, C., Ambartsumyan, G., Aimiuwu, O., Richter, L., Zhang, J., et al. (2009). Induced Pluripotent Stem Cells and Embryonic Stem Cells Are Distinguished by Gene Expression Signatures. Cell Stem Cell *5*, 111–123.

Choi, J., Southworth, L.K., Sarin, K.Y., Venteicher, A.S., Ma, W., Chang, W., Cheung, P., Jun, S., Artandi, M.K., Shah, N., et al. (2008). TERT promotes epithelial proliferation through transcriptional control of a Myc- and Wnt-related developmental program. PLoS Genet. *4*, 0124–0138.

Cobaleda, C., Jochum, W., and Busslinger, M. (2007). Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. Nature *449*, 473–477.

Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A.

(2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes Dev. *22*, 746–755.

Cortázar, D., Kunz, C., Saito, Y., Steinacher, R., and Schär, P. (2007). The enigmatic thymine DNA glycosylase. DNA Repair (Amst). *6*, 489–504.

Cortázar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F., et al. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419–423.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67–79.

Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T. a, Shliaha, P. V, Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., et al. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature *495*, 370–374.

Cowan, C.A., Atienza, J., Melton, D.A., and Eggan, K. (2005). Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. Science *309*, 1369–1373.

Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. J. Cell Biol. *153*, 773–784.

Daley, G.Q., and Scadden, D.T. (2008). Prospects for stem cell-based therapy. Cell *132*, 544–548.

Darzacq, X., Jady, B.E., Verheggen, C., Kiss, a M., Bertrand, E., and Kiss, T. (2002). Cajal body-specic small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. Embo J. *21*, 2746–2756.

Darzacq, X., Kittur, N., Roy, S., Shav-Tal, Y., Singer, R.H., and Meier, U.T. (2006). Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. J. Cell Biol. *173*, 207–218.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987–1000.

Dean, W., Santos, F., Stojkovic, M., Zakhartchenko, V., Walter, J., Wolf, E., and Reik, W. (2001). Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. Proc. Natl. Acad. Sci. U. S. A. *98*, 13734–13738.

Dimos, J.T., Rodolfa, K.T., Niakan, K.K., Weisenthal, L.M., Mitsumoto, H.,

Chung, W., Croft, G.F., Saphier, G., Leibel, R., Goland, R., et al. (2008). Generated from Patients with ALS Can Be Differentiated into Motor Neurons. Science (80-.). *321*, 1218–1221.

Ding, L., Paszkowski-Rogacz, M., Nitzsche, A., Slabicki, M.M., Heninger, A.K., Vries, I. de, Kittler, R., Junqueira, M., Shevchenko, A., Schulz, H., et al. (2009). A Genome-Scale RNAi Screen for Oct4 Modulators Defines a Role of the Paf1 Complex for Embryonic Stem Cell Identity. Cell Stem Cell *4*, 403–415.

Doege, C. a, Inoue, K., Yamashita, T., Rhee, D.B., Travis, S., Fujita, R., Guarnieri, P., Bhagat, G., Vanti, W.B., Shih, A., et al. (2012). Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. Nature *488*, 652–655.

Dupont, S., Morsut, L., Aragona, M., Enzo, E., Giulitti, S., Cordenonsi, M., Zanconato, F., Le Digabel, J., Forcato, M., Bicciato, S., et al. (2011). Role of YAP/TAZ in mechanotransduction. Nature *474*, 179–183.

Egan, E.D., and Collins, K. (2012). Biogenesis of telomerase ribonucleoproteins. RNA *18*, 1747–1759.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res. *10*, 2709–2721.

Ender, C., Krek, A., Friedländer, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., and Meister, G. (2008). A Human snoRNA with MicroRNA-Like Functions. Mol. Cell *32*, 519–528.

Engel, N., Tront, J.S., Erinle, T., Nguyen, N., Latham, K.E., Sapienza, C., Hoffman, B., and Liebermann, D. a (2009). Conserved DNA methylation in Gadd45a(-/-) mice. Epigenetics *4*, 98–99.

Errabolu, R., Sanders, M. a, and Salisbury, J.L. (1994). Cloning of a cDNA encoding human centrin, an EF-hand protein of centrosomes and mitotic spindle poles. J. Cell Sci. *107 (Pt 1*, 9–16.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156.

Fan, J., Robert, C., Jang, Y.-Y., Liu, H., Sharkis, S., Baylin, S.B., and Rassool, F.V. (2011). Human induced pluripotent cells resemble embryonic stem cells demonstrating enhanced levels of DNA repair and efficacy of nonhomologous end-joining. Mutat. Res. *713*, 8–17.

Fazzio, T.G., Huff, J.T., and Panning, B. (2008). An RNAi Screen of Chromatin Proteins Identifies Tip60-p400 as a Regulator of Embryonic Stem Cell Identity. Cell *134*, 162–174.

Federation, A.J., Bradner, J.E., and Meissner, A. (2014). The use of small molecules in somatic-cell reprogramming. Trends Cell Biol. *24*, 179–187.

Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. Cell Stem Cell *11*, 477–490.

Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.L., Walter, J., et al. (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. Cell Stem Cell *13*, 351–359.

Filion, G.J.P., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., and Defossez, P. (2006). A Family of Human Zinc Finger Proteins That Bind Methylated DNA and Repress Transcription. Mol. Cell. Biol. *26*, 169.

Filipowicz, W., and Pogacic, V. (2002). Biogenesis of small nucleolar ribonucleoproteins. Curr. Opin. Cell Biol. *14*, 319–327.

Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suvà, M.L., and Bernstein, B.E. (2015). Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature *529*, 110–114.

Fong, H., Hohenstein, K. a, and Donovan, P.J. (2008). Regulation of self-renewal and pluripotency by Sox2 in human embryonic stem cells. Stem Cells *26*, 1931–1938.

Fong, Y.W., Inouye, C., Yamaguchi, T., Cattoglio, C., Grubisic, I., and Tjian, R. (2011). A DNA repair complex functions as an Oct4/Sox2 coactivator in embryonic stem cells. Cell *147*, 120–131.

Fong, Y.W., Cattoglio, C., Yamaguchi, T., and Tjian, R. (2012). Transcriptional regulation by coactivators in embryonic stem cells. Trends Cell Biol. *22*, 292–298.

Fong, Y.W., Cattoglio, C., and Tjian, R. (2013). The Intertwined Roles of Transcription and Repair Proteins. Mol. Cell *52*, 291–302.

Fong, Y.W., Ho, J.J., Inouye, C., and Tjian, R. (2014). The dyskerin ribonucleoprotein complex as an OCT4/SOX2 coactivator in embryonic stem cells. Elife *3*, 1–30.

Fu, D., and Collins, K. (2007). Purification of Human Telomerase Complexes Identifies Factors Involved in Telomerase Biogenesis and Telomere Length Regulation. Mol. Cell *28*, 773–785.

Funaki, S., Nakamura, T., Nakatani, T., Umehara, H., Nakashima, H., and Nakano, T. (2014). Inhibition of maintenance DNA methylation by Stella. Biochem. Biophys. Res. Commun. *453*, 455–460.

Galan-Caridad, J.M., Harel, S., Arenzana, T.L., Hou, Z.E., Doetsch, F.K., Mirny, L.A., and Reizis, B. (2007). Zfx Controls the Self-Renewal of Embryonic and Hematopoietic Stem Cells. Cell *129*, 345–357.

Gallinari, P., and Jirincny, J. (1996). A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. Nature *383*, 735–738.

Gao, Y., Chen, J., Li, K., Wu, T., Huang, B., Liu, W., Kou, X., Zhang, Y., Huang, H., Jiang, Y., et al. (2013). Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. Cell Stem Cell *12*, 453–469.

Gao, Z., Cox, J.L., Gilmore, J.M., Ormsbee, B.D., Mallanna, S.K., Washburn, M.P., and Rizzino, A. (2012). Determination of protein interactome of transcription factor Sox2 in embryonic stem cells engineered for inducible expression of four reprogramming factors. J. Biol. Chem. *287*, 11384–11397.

Golipour, A., David, L., Liu, Y., Jayakumaran, G., Hirsch, C.L., Trcka, D., and Wrana, J.L. (2012). A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. Cell Stem Cell *11*, 769–782.

González, F., Georgieva, D., Vanoli, F., Shi, Z.-D., Stadtfeld, M., Ludwig, T., Jasin, M., and Huangfu, D. (2013). Homologous recombination DNA repair genes play a critical role in reprogramming to a pluripotent state. Cell Rep. *3*, 651–660.

Grozdanov, P.N., Roy, S., Kittur, N., and Meier, U.T. (2009). SHQ1 is required prior to NAF1 for assembly of H/ACA small nucleolar and telomerase RNPs. RNA *15*, 1188–1197.

Gu, B.W., Ge, J., Fan, J.M., Bessler, M., and Mason, P.J. (2013). Slow growth and unstable ribosomal RNA lacking pseudouridine in mouse embryonic fibroblast cells expressing catalytically inactive dyskerin. FEBS Lett. *587*, 2112–2117.

Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606–610.

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.P., Hu, B., et al. (2014a). Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. Cell Stem Cell *15*, 447–458.

Guo, F., Ding, Y., Caberoy, N., Alvarado, G., Wang, F., Chen, R., and Li, W. (2015). ABCF1 extrinsically regulates retinal pigment epithelial cell phagocytosis. Mol. Biol. Cell *26*, 2311–2320.

Guo, G., von Meyenn, F., Santos, F., Chen, Y., Reik, W., Bertone, P., Smith, A., and Nichols, J. (2016). Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. Stem Cell Reports *6*, 437–446.

Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., et al. (2014b). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat Neurosci *17*, 215–222.

Guo, S., Zi, X., Schulz, V.P., Cheng, J., Zhong, M., Koochaki, S.H.J., Megyola, C.M., Pan, X., Heydari, K., Weissman, S.M., et al. (2014c). Nonstochastic reprogramming from a privileged somatic cell state. Cell *156*, 649–662.

Gurdon, J.B. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. J. Embryol. Exp. Morphol. *10*, 622–640.

Haines, T.R., Rodenhiser, D.I., and Ainsworth, P.J. (2001). Allele-specific non-CpG methylation of the Nf1 gene during early mouse development. Dev. Biol. *240*, 585–598.

Han, D.W., Do, J.T., Gentile, L., Stehling, M., Lee, H.T., and Schöler, H.R. (2008). Pluripotential reprogramming of the somatic genome in hybrid cells occurs with the first cell cycle. Stem Cells *26*, 445–454.

Han, J., Yuan, P., Yang, H., Zhang, J., Soh, B.S., Li, P., Lim, S.L., Cao, S., Tay, J., Orlov, Y.L., et al. (2010). Tbx3 improves the germ-line competency of induced pluripotent stem cells. Nature *463*, 1096–1100.

Hananian, J., and Cleaver, J.E. (1980). Xeroderma pigmentosum exhibiting neurological disorders and systemic lupus erythematosus. Clin. Genet. *17*, 39–45.

Hanna, J., Wernig, M., Markoulaki, S., Sun, C., Meissner, A., Cassady, J.P., Beard, C., Brambrink, T., Wu, L., Townes, T.M., et al. (2007). Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. Science *318*, 1920–1923.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *462*, 595–601.

Hardeland, U., Steinacher, R., Jiricny, J., and Schär, P. (2002). Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. EMBO J. *21*, 1456–1464.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature *405*, 486–489.

Harland, R.M. (1982). Inheritance of DNA methylation in microinjected eggs of Xenopus laevis. Proc. Natl. Acad. Sci. U. S. A. *79*, 2323–2327.

Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature *345*, 458–460.

Hashimoto, H., Hong, S., Bhagwat, A.S., Zhang, X., and Cheng, X. (2012a). Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. Nucleic Acids Res. *40*, 10203–10214.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012b). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. *40*, 4841–4849.

Havugimana, P.C., Hart, G.T., Nepusz, T., Yang, H., Turinsky, A.L., Li, Z., Wang, P.I., Boutz, D.R., Fong, V., Phanse, S., et al. (2012). A census of human soluble protein complexes. Cell *150*, 1068–1081.

Hay, D.C., Sutherland, L., Clark, J., and Burdon, T. (2004). Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. Stem Cells *22*, 225–235.

He, Y., and Ecker, J.R. (2015). Non-CG Methylation in the Human Genome. Annu. Rev. Genomics Hum. Genet. *16*, 55–77.

He, J., Navarrete, S., Jasinski, M., Vulliamy, T., Dokal, I., Bessler, M., and Mason, P.J. (2002). Targeted disruption of Dkc1, the gene mutated in X-linked dyskeratosis congenita, causes embryonic lethality in mice. Oncogene *21*, 7740–7744.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303–1307.

Heiss, N.S., Knight, S.W., Vulliamy, T.J., Klauck, S.M., Wiemann, S., Mason, P.J., Poustka, A., and Dokal, I. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. Nat. Genet. *19*, 32–38.

Hejnar, J., Hájková, P., Plachy, J., Elleder, D., Stepanets, V., and Svoboda, J. (2001). CpG island protects Rous sarcoma virus-derived vectors integrated into nonpermissive cells from DNA methylation and transcriptional suppression. Proc. Natl. Acad. Sci. U. S. A. *98*, 565–569.

Helmrich, A., Ballarino, M., Nudler, E., and Tora, L. (2013). Transcription-replication encounters, consequences and genomic instability. Nat. Struct. Mol. Biol. *20*, 412–418.

Hendrich, B., and Bird, a (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol. Cell. Biol. *18*, 6538–6547.

Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A., and Bird, A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse

development. Genes Dev. *15*, 710–723.

Heng, J.C.D., Feng, B., Han, J., Jiang, J., Kraus, P., Ng, J.H., Orlov, Y.L., Huss, M., Yang, L., Lufkin, T., et al. (2010). The Nuclear Receptor Nr5a2 Can Replace Oct4 in the Reprogramming of Murine Somatic Cells to Pluripotent Cells. Cell Stem Cell *6*, 167–174.

Ho, L., Ronan, J.L., Wu, J., Staahl, B.T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A.I., Ranish, J., and Crabtree, G.R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proc. Natl. Acad. Sci. U. S. A. *106*, 5181–5186.

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

Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes Dev. *23*, 837–848.

Hu, X., Zhang, L., Mao, S.Q., Li, Z., Chen, J., Zhang, R.R., Wu, H.P., Gao, J., Guo, F., Liu, W., et al. (2014). Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. Cell Stem Cell *14*, 512–522.

Hyslop, L., Stojkovic, M., Armstrong, L., Walter, T., Stojkovic, P., Przyborski, S., Herbert, M., Murdoch, A., Strachan, T., and Lako, M. (2005). Downregulation of NANOG induces differentiation of human embryonic stem cells to extraembryonic lineages. Stem Cells *23*, 1035–1043.

Inoue, A., and Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science *334*, 194.

Ito, S., D'Alessio, A.C., Taranova, O. V, Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129–1133.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J. a, He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303.

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

Jády, B.E., Ketele, A., and Kiss, T. (2012). Human intron-encoded Alu RNAs are processed and packaged into Wdr79-associated nucleoplasmic box H/ACA RNPs. Genes Dev. *26*, 1897–1910.

Jaenisch, R., and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell *132*, 567–582.

Jia, W., Chen, W., and Kang, J. (2013). The functions of microRNAs and long non-coding RNAs in embryonic and induced pluripotent stem cells. Genomics. Proteomics Bioinformatics *11*, 275–283.

Jiang, J., Chan, Y.-S., Loh, Y.-H., Cai, J., Tong, G.-Q., Lim, C.-A., Robson, P., Zhong, S., and Ng, H.-H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat. Cell Biol. *10*, 353–360.

Jin, S.-G., Guo, C., and Pfeifer, G.P. (2008). GADD45A does not promote DNA demethylation. PLoS Genet. *4*, e1000013.

Jobert, L., Skjeldam, H.K., Dalhus, B., Galashevskaya, A., Vågbø, C.B., Bjørås, M., and Nilsen, H. (2013). The Human Base Excision Repair Enzyme SMUG1 Directly Interacts with DKC1 and Contributes to RNA Quality Control. Mol. Cell *49*, 339–345.

Jones, P.L., Veenstra, G.J., Wade, P. a, Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, a P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat. Genet. *19*, 187–191.

Judson, R.L., Babiarz, J.E., Venere, M., and Blelloch, R. (2009). Embryonic stem cell-specific microRNAs promote induced pluripotency. Nat. Biotechnol. *27*, 459–461.

Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., Cedar, H., and Razin, a (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germline. Genes Dev. *6*, 705–714.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kaji, K., Caballero, I.M., MacLeod, R., Nichols, J., Wilson, V. a, and Hendrich, B. (2006). The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat. Cell Biol. *8*, 285–292.

Kareta, M.S., Botello, Z.M., Ennis, J.J., Chou, C., and Chédin, F. (2006). Reconstitution and mechanism of the stimulation of de novo methylation by human DNMT3L. J. Biol. Chem. *281*, 25893–25902.

Kawai, T., and Akira, S. (2010). The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat. Immunol. *11*, 373–384.

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

Keramari, M., Razavi, J., Ingman, K.A., Patsch, C., Edenhofer, F., Ward, C.M., and Kimber, S.J. (2010). Sox2 is essential for formation of trophectoderm in the preimplantation embryo. PLoS One *5*, e13952.

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

Kim, K., Doi, a, Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I.R., et al. (2010). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285–290.

Kirillov, A., Kistler, B., Mostoslavsky, R., Cedar, H., Wirth, T., and Bergman, Y. (1996). A role for nuclear NF-kappaB in B-cell-specific demethylation of the Igkappa locus. Nat. Genet. *13*, 435–441.

Kiss, T., Fayet, E., Jády, B.E., Richard, P., and Weber, M. (2006). Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. Cold Spring Harb. Symp. Quant. Biol. *71*, 407–417.

Kiss, T., Fayet-Lebaron, E., and Jady, B.E. (2010). Box H/ACA Small Ribonucleoproteins. Mol. Cell *37*, 597–606.

Klein, I., Sarkadi, B., and Váradi, A. (1999). An inventory of the human ABC proteins. Biochim. Biophys. Acta *1461*, 237–262.

Knight, S.W., Heiss, N.S., Vulliamy, T.J., Greschner, S., Stavrides, G., Pai, G.S., Lestringant, G., Varma, N., Mason, P.J., Dokal, I., et al. (1999). X-linked dyskeratosis congenita is predominantly caused by missense mutations in the DKC1 gene. Am. J. Hum. Genet. *65*, 50–58.

Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C. a, Mostoslavsky, G., et al. (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell *8*, 200–213.

Kondo, E., Gu, Z., Horii, A., and Fukushige, S. (2005). The Thymine DNA Glycosylase MBD4 Represses Transcription and Is Associated with Methylated p16 INK4a and hMLH1 Genes The Thymine DNA Glycosylase MBD4 Represses Transcription and Is Associated with Methylated p16 INK4a and hMLH1 Genes. Mol. Cell. Biol. *25*, 4388–4396.

Kuroda, T., Tada, M., Kubota, H., Kimura, H., Hatano, S., Suemori, H., Nakatsuji, N., and Tada, T. (2005). Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. Mol. Cell. Biol. *25*, 2475–2485.

de Laat, W.L., Jaspers, N.G., and Hoeijmakers, J.H. (1999). Molecular mechanism of nucleotide excision repair. Genes Dev. *13*, 768–785.

Larijani, M., Frieder, D., Sonbuchner, T.M., Bransteitter, R., Goodman, M.F., Bouhassira, E.E., Scharff, M.D., and Martin, A. (2005). Methylation protects cytidines from AID-mediated deamination. Mol. Immunol. *42*, 599–604.

Lee, V.D., and Huang, B. (1993). Molecular cloning and centrosomal localization of human caltractin. Proc. Natl. Acad. Sci. U. S. A. *90*, 11039–11043.

Lee, J., Sayed, N., Hunter, A., Au, K.F., Wong, W.H., Mocarski, E.S., Pera, R.R., Yakubov, E., and Cooke, J.P. (2012). Activation of innate immunity is required for efficient nuclear reprogramming. Cell *151*, 547–558.

Lee, M.N., Roy, M., Ong, S.-E., Mertins, P., Villani, A.-C., Li, W., Dotiwala, F., Sen, J., Doench, J.G., Orzalli, M.H., et al. (2013). Identification of regulators of the innate immune response to cytosolic DNA and retroviral infection by an integrative approach. Nat. Immunol. *14*, 179–185.

Lei, H., Oh, S.P., Okano, M., Jüttermann, R., Goss, K. a, Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development *122*, 3195–3205.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. *20*, 311–316.

Leonhardt, H., Page, a W., Weier, H.U., and Bestor, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell *71*, 865–873.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to Methylated DNA. Cell *69*, 905–914.

Li, L., and Ye, K. (2006). Crystal structure of an H/ACA box ribonucleoprotein particle. Nature *443*, 302–307.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926.

Li, G.-W., Oh, E., and Weissman, J.S. (2012). The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature *484*, 538–541.

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., et al. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell *7*, 51–63.

Li, S., Duan, J., Li, D., Yang, B., Dong, M., and Ye, K. (2011). Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. Genes

Dev. *25*, 2409–2421.

Lian, I., Kim, J., Okazawa, H., Zhao, J., Zhao, B., Yu, J., Chinnaiyan, A., Israel, M.A., Goldstein, L.S.B., Abujarour, R., et al. (2010). The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. Genes Dev. *24*, 1106–1118.

Liang, B., and Li, H. (2011). Structures of ribonucleoprotein particle modification enzymes. Q. Rev. Biophys. *44*, 95–122.

Liang, J., Wan, M., Zhang, Y., Gu, P., Xin, H., Jung, S.Y., Qin, J., Wong, J., Cooney, A.J., Liu, D., et al. (2008). Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat. Cell Biol. *10*, 731–739.

Lin, Y., Li, X.-Y., Willis, A.L., Liu, C., Chen, G., and Weiss, S.J. (2014). Snail1-dependent control of embryonic stem cell pluripotency and lineage commitment. Nat. Commun. *5*, 3070.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature *471*, 68–73.

Lister, R., Mukamel, E. a, Nery, J.R., Urich, M., Puddifoot, C. a, Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. Science *341*, 1237905.

Listerman, I., Gazzaniga, F.S., and Blackburn, E.H. (2014). An investigation of the effects of the core protein telomerase reverse transcriptase on Wnt signaling in breast cancer cells. Mol. Cell. Biol. *34*, 280–289.

Liu, W., Coleman, R. a, Ma, E., Grob, P., Yang, J.L., Zhang, Y., Dailey, G., Nogales, E., and Tjian, R. (2009). Structures of three distinct activator − TFIID complexes Structures of three distinct activator – TFIID complexes. Genes Dev. 1510–1521.

Liu, X., Sun, H., Qi, J., Wang, L., He, S., Liu, J., Feng, C., Chen, C., Li, W., Guo, Y., et al. (2013). Sequential introduction of reprogramming factors reveals a time-sensitive requirement for individual factors and a sequential EMT-MET mechanism for optimal reprogramming. Nat. Cell Biol. *15*, 829–838.

Liu, Z., Scannell, D.R., Eisen, M.B., and Tjian, R. (2011). Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. Cell *146*, 720–731.

Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat. Genet. *38*, 431–440.

Lu, X., Han, D., Boxuan Simen, Z., Song, C.-X., Zhang, L.-S., Doré, L.C., and He, C. (2015). Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. Cell Res. 386–389.

Lyons, A.B., and Parish, C.R. (1994). Determination of lymphocyte division by flow cytometry. J. Immunol. Methods *171*, 131–137.

Ma, H., Morey, R., O'Neil, R.C., He, Y., Daughtry, B., Schultz, M.D., Hariharan, M., Nery, J.R., Castanon, R., Sabatini, K., et al. (2014). Abnormalities in human pluripotent cells due to reprogramming mechanisms. Nature *511*, 177–183.

Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Spl sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev. *8*, 2282–2292.

Maillard, O., Solyom, S., and Naegeli, H. (2007). An aromatic sensor with aversion to damaged strands confers versatility to DNA repair. PLoS Biol. *5*, 717–728.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. J. Biol. Chem. *286*, 35334–35338.

Marin, M., Karis, A., Visser, P., Gosveld, F., and Philipsen, S. (1997). Transcription Factor Sp1 Is Essential for Early Embryonic Development but Dispensable for Cell Growth and Differentiation. Cell *89*, 619–628.

Marrone, A., Walne, A., and Dokal, I. (2005). Dyskeratosis congenita: Telomerase, telomeres and anticipation. Curr. Opin. Genet. Dev. *15*, 249–257.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA Genes to the Core Transcriptional Regulatory Circuitry of Embryonic Stem Cells. Cell *134*, 521–533.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. U. S. A. *78*, 7634–7638.

Mason, P.J., and Bessler, M. (2011). The genetics of dyskeratosis congenita. Cancer Genet. *204*, 635–645.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell

Biol *9*, 625–U26.

Maurano, M.T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K., and Stamatoyannopoulos, J.A. (2015). Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Rep. *12*, 1184–1195.

Le May, N., Mota-Fernandes, D., Vélez-Cruz, R., Iltis, I., Biard, D., and Egly, J.M. (2010). NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack. Mol. Cell *38*, 54–66.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501–502.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell *58*, 499–507.

Meier, U.T. (2005). The many facets of H/ACA ribonucleoproteins. Chromosoma *114*, 1–14.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49–55.

Mitchell, J.R., Collins, K., and Wood, E. (1999). A telomerase component is defective in the human disease dyskeratosis congenita. Nature *402*, 551–555.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631–642.

Mochizuki, Y., He, J., Kulkarni, S., Bessler, M., and Mason, P.J. (2004). Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing. Proc. Natl. Acad. Sci. U. S. A. *101*, 10756–10761.

Molina-Estevez, F.J., Lozano, M.L., Navarro, S., Torres, Y., Grabundzija, I., Ivics, Z., Samper, E., Bueren, J.A., and Guenechea, G. (2013). Brief report: Impaired cell reprogramming in nonhomologous end joining deficient cells. Stem Cells *31*, 1726–1730.

Molinier, J., Ramos, C., Fritsch, O., and Hohn, B. (2004). CENTRIN2 modulates homologous recombination and nucleotide excision repair in Arabidopsis. Plant Cell *16*, 1633–1643.

Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and Regional Changes

in DNA Methylation in the Embryonic, Extraembryonic and Germ-Cell Lineages During Mouse Embryo Development. Development *99*, 371–382.

Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: Implications for epigenetic reprogramming. J. Biol. Chem. *279*, 52353–52360.

Morgan, M.T., Maiti, A., Fitzgerald, M.E., and Drohat, A.C. (2011). Stoichiometry and affinity for thymine DNA glycosylase binding to specific and nonspecific DNA. Nucleic Acids Res. *39*, 2319–2329.

Mori, M., Triboulet, R., Mohseni, M., Schlegelmilch, K., Shrestha, K., Camargo, F.D., and Gregory, R.I. (2014). Hippo signaling regulates microprocessor and links cell-density-dependent mirna biogenesis to cancer. Cell *156*, 893–906.

Mu, D., Hsu, D.S., and Sancar, A. (1996). Reaction mechanism of human DNA repair excision nuclease. J. Biol. Chem. *271*, 8285–8294.

Muller, L.U.W., Milsom, M.D., Harris, C.E., Vyas, R., Brumme, K.M., Parmar, K., Moreau, L.A., Schambach, A., Park, I.H., London, W.B., et al. (2012). Overcoming reprogramming resistance of Fanconi anemia cells. Blood *119*, 5449–5457.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386–389.

Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J.J., Wiebauer, K., and Jiricny, J. (1996). Cloning and Expression of Human G / T Mismatch-specific Thymine-DNA Glycosylase * the formation of G / T mismatches . We have shown previ-. *271*, 12767–12774.

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Anselmi, F., Parlato, C., Medana, C., DalBello, F., and Oliviero, S. (2015). Single-Base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA Methylation Dynamics. Cell Rep. *10*, 674–683.

Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C. a, Turner, B.M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, a (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat. Genet. *23*, 58–61.

Ng, J.M.Y., Vrieling, H., Sugasawa, K., Ooms, M.P., Grootegoed, J.A., Vreeburg, J.T.M., Visser, P., Beems, R.B., Gorgels, T.G.M.F., Hanaoka, F., et al. (2002). Developmental Defects and Male Sterility in Mice Lacking the Ubiquitin-Like DNA Repair Gene mHR23B. Mol. Cell Biol. *22*, 1233–1245.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D.,

Chambers, I., Schöler, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379–391.

Niida, H., Matsumoto, T., Satoh, H., Shiwa, M., Tokutake, Y., Furuichi, Y., and Shinkai, Y. (1998). Severe growth defect in mouse cells lacking the telomerase RNA component. Nat. Genet. *19*, 203–206.

Nishi, R., Okuda, Y., Watanabe, E., Mori, T., Iwai, S., Masutani, C., Sugasawa, K., and Hanaoka, F. (2005). Centrin 2 Stimulates Nucleotide Excision Repair by Interacting with Xeroderma Pigmentosum Group C Protein. Society *25*, 5664–5674.

Niwa, H., Miyazaki, J., and Smith, a G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. *24*, 372–376.

Nyce, J., Liu, L., and Jones, P.A. (1986). Variable effects of DNA-synthesis inhibitors upon DNA methylation in mammalian cells. Nucleic Acids Res. *14*, 4353–4367.

Ohgushi, M., and Sasai, Y. (2011). Lonely death dance of human pluripotent stem cells: ROCKing between metastable cell states. Trends Cell Biol. *21*, 274–282.

Ohi, Y., Qin, H., Hong, C., Blouin, L., Polo, J.M., Guo, T., Qi, Z., Downey, S.L., Manos, P.D., Rossi, D.J., et al. (2011). Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nat. Cell Biol. *13*, 541–549.

Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat. Am. Inc. *19*, 219–220.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.

Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K., et al. (2011). A more efficient method to generate integration-free human iPS cells. Nat. Methods *8*, 409–412.

Okuda, Y., Nishi, R., Ng, J.M.Y., Vermeulen, W., van der Horst, G.T.J., Mori, T., Hoeijmakers, J.H.J., Hanaoka, F., and Sugasawa, K. (2004). Relative levels of the two mammalian Rad23 homologs determine composition and stability of the xeroderma pigmentosum group C protein complex. DNA Repair (Amst). *3*, 1285–1295.

Orkin, S.H., and Hochedlinger, K. (2011). Chromatin connections to pluripotency and cellular reprogramming. Cell *145*, 835–850.

Palmieri, S.L., Werner, P., Hess, H., and Scholer, H.R. (1994). Oct-4 Transcription Factor is Differentially Expressed in the Mouse Embryo during

Establishment of the two first Extraembryonic Cell Lineages Involved in Implantation. Dev. Biol. 259–267.

Park, J.-I., Venteicher, A.S., Hong, J.Y., Choi, J., Jun, S., Shkreli, M., Chang, W., Meng, Z., Cheung, P., Ji, H., et al. (2009). Telomerase modulates Wnt signalling by association with target gene chromatin. Nature *460*, 66–72.

Paytubi, S., Wang, X., Lam, Y.W., Izquierdo, L., Hunter, M.J., Jan, E., Hundal, H.S., and Proud, C.G. (2009). ABC50 promotes translation initiation in mammalian cells. J. Biol. Chem. *284*, 24061–24073.

Pereira, C.F., Terranova, R., Ryan, N.K., Santos, J., Morris, K.J., Cui, W., Merkenschlager, M., and Fisher, A.G. (2008). Heterokaryon-based reprogramming of human B lymphocytes for pluripotency requires Oct4 but not Sox2. PLoS Genet. *4*.

Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S.K., Eisen, D., Truss, M., Steinbacher, J., Hackner, B., Kotljarova, O., et al. (2014). Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. Nat. Chem. Biol. *10*, 574–581.

Picanço-Castro, V., Russo-Carbolante, E., Reis, L.C.J., Fraga, A.M., de Magalhães, D.A.R., Orellana, M.D., Panepucci, R.A., Pereira, L.V., and Covas, D.T. (2011). Pluripotent reprogramming of fibroblasts by lentiviral mediated insertion of SOX2, C-MYC, and TCL-1A. Stem Cells Dev. *20*, 169–180.

Pijnappel, W.W.M.P., Esch, D., Baltissen, M.P. a, Wu, G., Mischerikow, N., Bergsma, A.J., van der Wal, E., Han, D.W., Bruch, H. Vom, Moritz, S., et al. (2013). A central role for TFIID in the pluripotent transcription circuitry. Nature *495*, 516–519.

Planello, A.C., Ji, J., Sharma, V., Singhania, R., Mbabaali, F., Müller, F., Alfaro, J. a, Bock, C., De Carvalho, D.D., and Batada, N.N. (2014). Aberrant DNA methylation reprogramming during induced pluripotent stem cell generation is dependent on the choice of reprogramming factors. Cell Regen. *3*, 4.

Polo, J.M., Liu, S., Figueroa, M.E., Kulalert, W., Eminli, S., Tan, K.Y., Apostolou, E., Stadtfeld, M., Li, Y., Shioda, T., et al. (2010). Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat. Biotechnol. *28*, 848–855.

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B. a, Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. Cell *151*, 1617–1632.

Popescu, A., Miron, S., Blouquit, Y., Duchambon, P., Christova, P., and Craescu, C.T. (2003). Xeroderma pigmentosum group C protein possesses a high affinity binding site to human centrin 2 and calmodulin. J. Biol. Chem. *278*, 40252–40261.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D. a, and Cairns, B.R.

(2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell *135*, 1201–1212.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., et al. (2013). Deterministic direct reprogramming of somatic cells to pluripotency. Nature *502*, 65–70.

Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R.C., and Melton, D. a (2003). Transcriptional Profiling of Embryonic and Adult Stem Cells. October *302*, 37102–37102.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, a P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc. Natl. Acad. Sci. U. S. A. *97*, 5237–5242.

Reardon, J.T., Mu, D., and Sancar, A. (1996). Overproduction, Purification, and Characterization of the XPC Subunit of the Human DNA Repair Excision Nuclease. J. Biol. Chem. *271*, 19451–19456.

Rhee, I., Bachman, K.E., Park, B.H., Jair, K., Yen, R.-W.C., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W., et al. (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature *416*, 552–556.

Richard, M., Drouin, R., and Beaulieu, A.D. (1998). ABC50 a novel human ATP-binding cassette protein found in tumor necrosis factor-α-stimulated synoviocytes. Genomics *53*, 137–145.

Richards, M., Tan, S., Tan, J., Chan, W., and Bongso, A. (2004). The transcriptome profile of human embryonic stem cells as defined by SAGE. Stem Cells *22*, 51–64.

Riedl, T., Hanaoka, F., and Egly, J.M. (2003). The comings and goings of nucleotide excision repair factors on damaged DNA. EMBO J. *22*, 5293–5303.

Rodda, D.J., Chew, J.-L., Lim, L.-H., Loh, Y.-H., Wang, B., Ng, H.-H., and Robson, P. (2005). Transcriptional regulation of nanog by OCT4 and SOX2. J. Biol. Chem. *280*, 24731–24737.

Ruiz, S., Panopoulos, A.D., Herrerías, A., Bissig, K.-D., Lutz, M., Berggren, W.T., Verma, I.M., and Izpisua Belmonte, J.C. (2011). A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. Curr. Biol. *21*, 45–52.

Sado, T., Fenner, M.H., Tan, S.S., Tam, P., Shioda, T., and Li, E. (2000). X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. Dev. Biol. *225*, 294–303.

Salomon, R., and Kaye, A.M. (1970). Methylation of mouse DNA in vivo: di- and tripyrimidine sequences containing 5-methylcytosine. Biochim. Biophys. Acta *204*, 340–351.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.-K., Beyer, T. a, Datti, A., Woltjen, K., Nagy, A., and Wrana, J.L. (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell *7*, 64–77.

Santi, D., Garrett, C., and Barr, P. (1983). On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. Cell *33*, 9–10.

Sato, N., Munoz Sanjuan, I., Heke, M., Uchida, M., Naef, F., and Brivanlou, A.H. (2003). Molecular signature of human embryonic stem cells and its comparison with the mouse. Dev. Biol. *260*, 404–413.

Schlegelmilch, K., Mohseni, M., Kirak, O., Pruszak, J., Rodriguez, J.R., Zhou, D., Kreger, B.T., Vasioukhin, V., Avruch, J., Brummelkamp, T.R., et al. (2011). Yap1 acts downstream of ??-catenin to control epidermal proliferation. Cell *144*, 782–795.

Schneider, E., and Hunke, S. (1998). ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains. FEMS Microbiol. Rev. *22*, 1–20.

Schubert, T., Pusch, M.C., Diermeier, S., Benes, V., Kremmer, E., Imhof, A., and Längst, G. (2012). Df31 protein and snoRNAs maintain accessible higher-order structures of chromatin. Mol. Cell *48*, 434–444.

Schultz, M.D., He, Y., Whitaker, J.W., Hariharan, M., Mukamel, E.A., Leung, D., Rajagopal, N., Nery, J.R., Urich, M.A., Chen, H., et al. (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. Nature *523*, 212–216.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T. a, Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908–912.

Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.-L., Zhang, K., and Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. Cell *153*, 692–706.

Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., and Zhang, Y. (2014). Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. Cell Stem Cell *15*, 459–470.

Shimizu, Y., Uchimura, Y., Dohmae, N., Saitoh, H., Hanaoka, F., and Sugasawa, K. (2010). Stimulation of DNA Glycosylase Activities by XPC Protein Complex: Roles of Protein-Protein Interactions. J. Nucleic Acids *2010*.

Siegfried, Z., Eden, S., Mendelsohn, M., Feng, X., Tsuberi, B.Z., and Cedar, H. (1999). DNA methylation represses transcription in vivo. Nat. Genet. *22*, 203–206.

Silva, J., and Smith, A. (2008). Capturing Pluripotency. Cell *132*, 532–536.

Silva, A.J., Ward, K., and White, R. (1993). Mosaic methylation in clonal tissue. Dev. Biol. *156*, 391–398.

Silva, J., Chambers, I., Pollard, S., and Smith, A. (2006). Nanog promotes transfer of pluripotency after cell fusion. Nature *441*, 997–1001.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog Is the Gateway to the Pluripotent Ground State. Cell *138*, 722–737.

Simonsson, S., and Gurdon, J. (2004). DNA demethylation is necessary for the epigenetic reprogramming of somatic cell nuclei. Nat. Cell Biol. *6*, 984–990.

Sindhu, C., Samavarchi-Tehrani, P., and Meissner, A. (2012). Transcription factor-mediated epigenetic reprogramming. J. Biol. Chem. *287*, 30922–30931.

Smet-Nocca, C., Wieruszeski, J.-M., Chaar, V., Leroy, A., and Benecke, A. (2008). The thymine-DNA glycosylase regulatory domain: residual structure and DNA binding. Biochemistry *47*, 6519–6530.

Smith, S.L., Everts, R.E., Tian, X.C., Du, F., Sung, L.-Y., Rodriguez-Zas, S.L., Jeong, B.-S., Renard, J.-P., Lewin, H. a, and Yang, X. (2005). Global gene expression profiles reveal significant nuclear reprogramming by the blastocyst stage after cloning. Proc. Natl. Acad. Sci. U. S. A. *102*, 17582–17587.

Smith, Z.D., Nachman, I., Regev, A., and Meissner, A. (2010). Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. Nat. Biotechnol. *28*, 521–526.

Soldner, F., Laganière, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S., et al. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell *146*, 318–331.

Sommer, C. a, Stadtfeld, M., Murphy, G.J., Hochedlinger, K., Kotton, D.N., and Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells *27*, 543–549.

Song, J., Teplova, M., Ishibe-Murakami, S., and Patel, D.J. (2012). Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science *335*, 709–712.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of

the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994–1004.

Sperger, J.M., Chen, X., Draper, J.S., Antosiewicz, J.E., Chon, C.H., Jones, S.B., Brooks, J.D., Andrews, P.W., Brown, P.O., and Thomson, J. a (2003). Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. Proc. Natl. Acad. Sci. U. S. A. *100*, 13350–13355.

Stadtfeld, M., and Hochedlinger, K. (2010). Induced pluripotency: history, mechanisms, and applications. Genes Dev. *24*, 2239–2263.

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell *2*, 230–240.

Steinacher, R., and Schär, P. (2005). Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. Curr. Biol. *15*, 616–623.

Stöger, R., Kajimura, T.M., Brown, W.T., and Laird, C.D. (1997). Epigenetic variation illustrated by DNA methylation patterns of the fragile-X gene FMR1. Hum. Mol. Genet. *6*, 1791–1801.

Su, H., Xu, T., Ganapathy, S., Shadfan, M., Long, M., Huang, T.H.-M., Thompson, I., and Yuan, Z.-M. (2014). Elevated snoRNA biogenesis is essential in breast cancer. Oncogene *33*, 1348–1358.

Sugasawa, K. (2011). Multiple DNA damage recognition factors involved in mammalian nucleotide excision repair. Biochem. *76*, 16–23.

Sugasawa, K., Masutani, C., Uchida, a, Maekawa, T., van der Spek, P.J., Bootsma, D., Hoeijmakers, J.H., and Hanaoka, F. (1996). HHR23B, a human Rad23 homolog, stimulates XPC protein in nucleotide excision repair in vitro. Mol. Cell. Biol. *16*, 4852–4861.

Sugasawa, K., Ng, J.M.Y., Masutani, C., Iwai, S., van der Spek, P.J., Eker, A.P.M., Hanaoka, F., Bootsma, D., and Hoeijmakers, J.H.J. (1998). Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. Mol. Cell *2*, 223–232.

Sugasawa, K., Shimizu, Y., Iwai, S., and Hanaoka, F. (2002). A molecular mechanism for DNA damage recognition by the xeroderma pigmentosum group C protein complex. DNA Repair (Amst). *1*, 95–107.

Szabó, P.E., Tang, S.H.E., Rentsendorj, A., Pfeifer, G.P., and Mann, J.R. (2000). Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. Curr. Biol. *10*, 607–610.

Taatjes, D.J., Näär, A.M., Andel, F., Nogales, E., and Tjian, R. (2002). Structure, function, and activator-induced conformations of the CRSP coactivator. Science *295*, 1058–1062.

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., and Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Curr. Biol. *11*, 1553–1558.

Taft, R.J., Glazov, E.A., Lassmann, T., Hayashizaki, Y., Carninci, P., and Mattick, J.S. (2009). Small RNAs derived from snoRNAs. RNA *15*, 1233–1240.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930–935.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. Cell *158*, 1254–1269.

Tan, S., Kern, R.C., and Selleck, W. (2005). The pST44 polycistronic expression system for producing protein complexes in Escherichia coli. Protein Expr. Purif. *40*, 385–395.

Tanaka, T.S. (2002). Gene Expression Profiling of Embryo-Derived Stem Cells Reveals Candidate Genes Associated With Pluripotency and Lineage Specificity. Genome Res. *12*, 1921–1928.

Terns, M., Terns, R., and Spring, C. (2006). Noncoding RNAs of the H / ACA Family Noncoding RNAs of the H / ACA Family. Cold Spring Harb. Symp. Quant. Biol. *LXXI*, 395–405.

Thiery, J.P., Acloque, H., Huang, R.Y.J., and Nieto, M.A. (2009). Epithelial-Mesenchymal Transitions in Development and Disease. Cell *139*, 871–890.

Tichy, E.D., and Stambrook, P.J. (2008). DNA repair in murine embryonic stem cells and differentiated cells. Exp. Cell Res. *314*, 1929–1936.

Tini, M., Benecke, A., Um, S.J., Torchia, J., Evans, R.M., and Chambon, P. (2002). Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. Mol. Cell *9*, 265–277.

Tsubooka, N., Ichisaka, T., Okita, K., Takahashi, K., Nakagawa, M., and Yamanaka, S. (2009). Roles of Sall4 in the generation of pluripotent stem cells from blastocysts and fibroblasts. Genes to Cells *14*, 683–694.

Tsubouchi, T., Soza-Ried, J., Brown, K., Piccolo, F.M., Cantone, I., Landeira, D., Bagci, H., Hochegger, H., Merkenschlager, M., and Fisher, A.G. (2013). DNA synthesis is required for reprogramming mediated by stem cell fusion. Cell *152*, 873–883.

Turker, M.S. (1999). The establishment and maintenance of DNA methylation patterns in mouse somatic cells. Semin. Cancer Biol. *9*, 329–337.

Turker, M.S. (2002). Gene silencing in mammalian cells and the spread of DNA methylation. Oncogene *21*, 5388–5393.

Tycowski, K.T., Shu, M.D., Kukoyi, A., and Steitz, J.A. (2009). A Conserved WD40 Protein Binds the Cajal Body Localization Signal of scaRNP Particles. Mol. Cell *34*, 47–57.

Tyzack, J.K., Wang, X., Belsham, G.J., and Proud, C.G. (2000). ABC50 interacts with eukaryotic initiation factor 2 and associates with the ribosome in an ATP-dependent manner. J. Biol. Chem. *275*, 34131–34139.

Uchida, A., Sugasawa, K., Masutani, C., Dohmae, N., Araki, M., Yokoi, M., Ohkuma, Y., and Hanaoka, F. (2002). The carboxy-terminal domain of the XPC protein plays a crucial role in nucleotide excision repair through interactions with transcription factor IIH. DNA Repair (Amst). *1*, 449–461.

Unternaehrer, J.J., Zhao, R., Kim, K., Cesana, M., Powers, J.T., Ratanasirintrawoot, S., Onder, T., Shibue, T., Weinberg, R.A., and Daley, G.Q. (2014). The epithelial-mesenchymal transition factor SNAIL paradoxically enhances reprogramming. Stem Cell Reports *3*, 691–698.

Utikal, J., Polo, J.M., Stadtfeld, M., Maherali, N., Kulalert, W., Walsh, R.M., Khalil, A., Rheinwald, J.G., and Hochedlinger, K. (2009). Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. Nature *460*, 1145–1148.

Volker, M., Moné, M.J., Karmakar, P., Van Hoffen, A., Schul, W., Vermeulen, W., Hoeijmakers, J.H.J., Van Driel, R., Van Zeeland, A.A., and Mullenders, L.H.F. (2001). Sequential assembly of the nucleotide excision repair factors in vivo. Mol. Cell *8*, 213–224.

Wade, P.A., Gegonne, A., Jones, P.L., Ballestar, E., Aubry, F., and Wolffe, A.P. (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat. Genet. *23*, 1–5.

Walne, A.J., Vulliamy, T., Marrone, A., Beswick, R., Kirwan, M., Masunari, Y., Al-Qurashi, F.H., Aljurf, M., and Dokal, I. (2007). Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-

associated protein NOP10. Hum. Mol. Genet. *16*, 1619–1629.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364–368.

Watanabe, D., Suetake, I., Tada, T., and Tajima, S. (2002). Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. Mech. Dev. *118*, 187–190.

Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J.B., Nishikawa, S., Nishikawa, S., Muguruma, K., et al. (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. Nat. Biotechnol. *25*, 681–686.

Waters, T.R., and Swann, P.F. (1998). Kinetics of the action of thymine DNA glycosylase. J. Biol. Chem. *273*, 20007–20014.

Waters, T.R., Gallinari, P., Jiricnyl, J., and Swann, P.F. (1999). Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. J. Biol. Chem. *274*, 67–74.

Weber, A.R., Schuermann, D., and Schär, P. (2014). Versatile recombinant SUMOylation system for the production of SUMO-modified protein. PLoS One *9*.

Wernig, M., Zhao, J., Pruszak, J., Hedlund, E., Fu, D., Soldner, F., Broccoli, V., Constantine-Paton, M., Isacson, O., and Jaenisch, R. (2008). Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease. Proc. Natl. Acad. Sci. U. S. A. *105*, 5856–5861.

Weston, and Parish (1990). New fluorescent dyes for lymphocyte migration studies. Analysis by flow cytometry and fluorescence microscopy. J. Immunol. Methods *133*, 87–97.

Wilcox, S.M. (2010). The Function of ABCF1 in Immunity and Mouse Development. University of British Columbia.

Wilusz, J.E., Sunwoo, H., and Spector, D.L. (2009). Long noncoding RNAs : functional surprises from the RNA world Long noncoding RNAs : functional surprises from the RNA world. 1494–1504.

Wong, J.M.Y., and Collins, K. (2006). Telomerase RNA level limits telomere maintenance in X-linked dyskeratosis congenita. Genes Dev. *20*, 2848–2858.

Woodcock, D.M., Crowther, P.J., and Diver, W.P. (1987). The majority of methylated deoxycytidines in human DNA are not in the CpG dinucleotide. Biochem. Biophys. Res. Commun. *145*, 888–894.

Wu, D.Y., and Yao, Z. (2006). Functional analysis of two Sp1/Sp3 binding sites in murine Nanog gene promoter. Cell Res. *16*, 319–322.

Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. Nat. Rev. Mol. Cell Biol. *11*, 607–620.

Wu, Q., Chen, X., Zhang, J., Loh, Y.H., Low, T.Y., Zhang, W., Zhang, W., Sze, S.K., Lim, B., and Ng, H.H. (2006). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. J. Biol. Chem. *281*, 24090–24094.

Xie, H., Ye, M., Feng, R., and Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. Cell *117*, 663–676.

Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., et al. (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell *153*, 1134–1148.

Yamanaka, S. (2009). Elite and stochastic models for induced pluripotent stem cell generation. Nature *460*, 49–52.

Yamanaka, S., and Blau, H.M. (2010). Nuclear reprogramming to a pluripotent state by three approaches. Nature *465*, 704–712.

Yang, P.K., Hoareau, C., Froment, C., Monsarrat, B., Henry, Y., and Chanfreau, G. (2005). Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly. Mol. Cell. Biol. *25*, 3295–3304.

Yang, X., Smith, S.L., Tian, X.C., Lewin, H. a, Renard, J.-P., and Wakayama, T. (2007). Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning. Nat. Genet. *39*, 295–302.

Yasuda, G., Nishi, R., Watanabe, E., Mori, T., Iwai, S., Orioli, D., Stefanini, M., Hanaoka, F., and Sugasawa, K. (2007). In vivo destabilization and functional defects of the xeroderma pigmentosum C protein caused by a pathogenic missense mutation. Mol. Cell. Biol. *27*, 6606–6614.

Yin, Q.F., Yang, L., Zhang, Y., Xiang, J.F., Wu, Y.W., Carmichael, G.G., and Chen, L.L. (2012). Long Noncoding RNAs with snoRNA Ends. Mol. Cell *48*, 219–230.

Yoder, J.A., Soman, N.S., Verdine, G.L., and Bestor, T.H. (1997). DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. J. Mol. Biol. *270*, 385–395.

Yokoi, M., Masutani, C., Maekawa, T., Sugasawa, K., Ohkuma, Y., and Hanaoka, F. (2000). The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA. J Biol Chem *275*, 9870–9875.

Yoon, A., Peng, G., Brandenburger, Y., Brandenburg, Y., Zollo, O., Xu, W., Rego, E., and Ruggero, D. (2006). Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita. Science *312*, 902–906.

Yu, J., and Thomson, J.A. (2006). Human Embryonic Stem Cells Reprogram Myeloid Precursors Following Cell–Cell Fusion. Stem Cells *24*, 168–176.

Yu, J., Vodyanik, M. a, Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G. a, Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917–1920.

Yuan, H., Corbi, N., Basilico, C., and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes Dev. *9*, 2635–2645.

Zeng, P.Y., Vakoc, C.R., Chen, Z.C., Blobel, G.A., and Berger, S.L. (2006). In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. Biotechniques *41*, 694–698.

Zhang, E.T., He, Y., Grob, P., Fong, Y.W., Nogales, E., and Tjian, R. (2015). Architecture of the human XPC DNA repair and stem cell coactivator complex. Proc. Natl. Acad. Sci.

Zhang, J., Tam, W.-L., Tong, G.Q., Wu, Q., Chan, H.-Y., Soh, B.-S., Lou, Y., Yang, J., Ma, Y., Chai, L., et al. (2006). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. Nat. Cell Biol. *8*, 1114–1123.

Zhang, X., Zhang, J., Wang, T., Esteban, M.A., and Pei, D. (2008). Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. J. Biol. Chem. *283*, 35825–35833.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. *13*, 1924–1935.

Zhao, Y., Yin, X., Qin, H., Zhu, F., Liu, H., Yang, W., Zhang, Q., Xiang, C., Hou, P., Song, Z., et al. (2008). Two Supporting Factors Greatly Improve the Efficiency of Human iPSC Generation. Cell Stem Cell *3*, 475–479.

APPENDIX A

SUPPLEMENTAL FIGURES PERTAINING TO CHAPTER 2: REGULATION OF DNA DEMETHYLATION BY THE XPC DNA REPAIR COMPLEX IN SOMATIC REPROGRAMMING

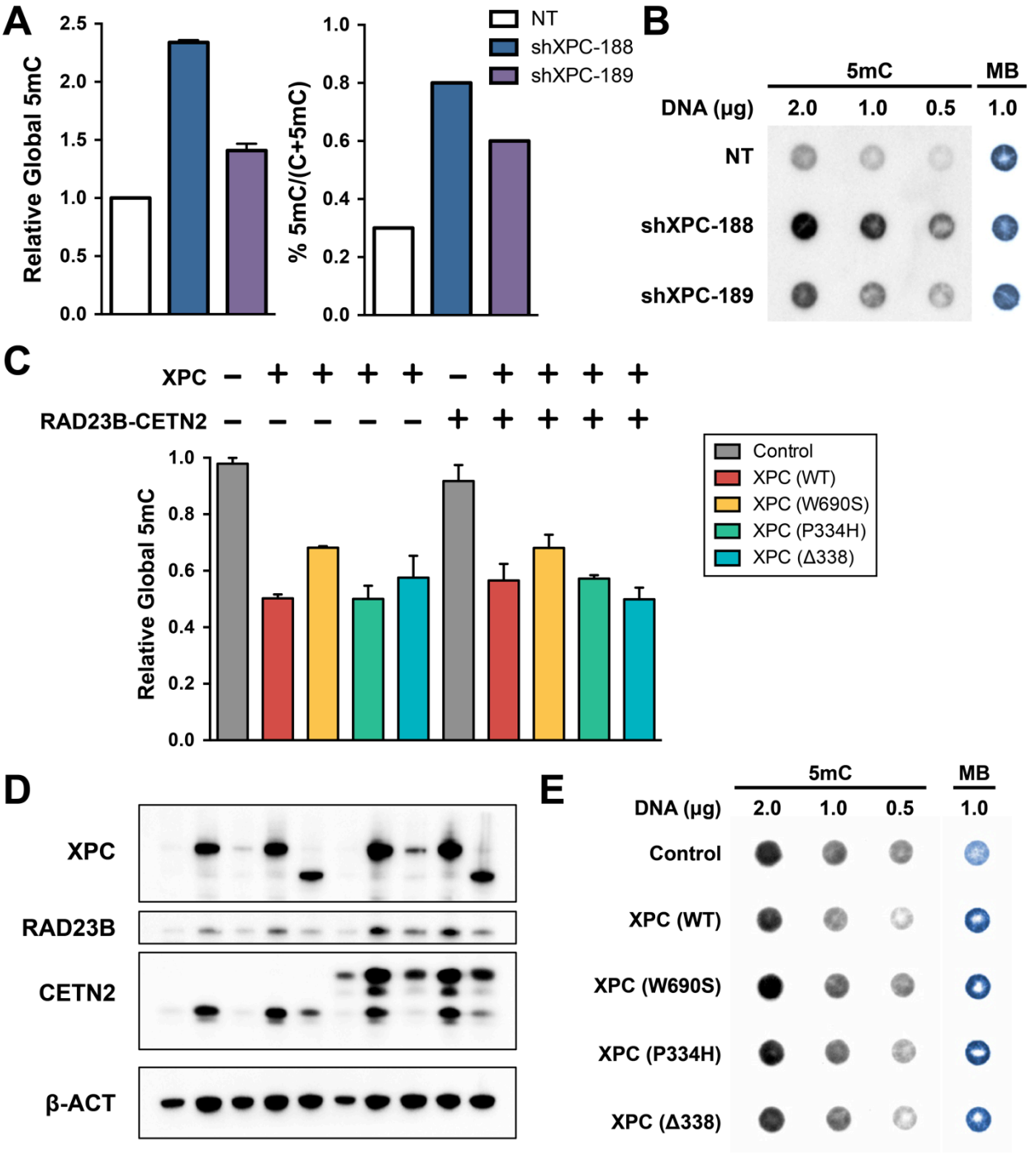

Figure 1. XPC knockdown (shXPC) in mouse ESCs leads to elevated global 5mC assayed by (**A**) 5mC-specific ELISA (left), thin layer chromatography (right), and (**B**) dot blot. (**C**) Global 5mC of HDFs overexpressing the XPC subunit with or without RAD23B-CETN2, measured by 5mC-specific ELISA. (**D**) Immunoblot analysis for HDF samples depicted in C. (**E**) 5mC dot blot of HDFs expressing the XPC subunit alone.

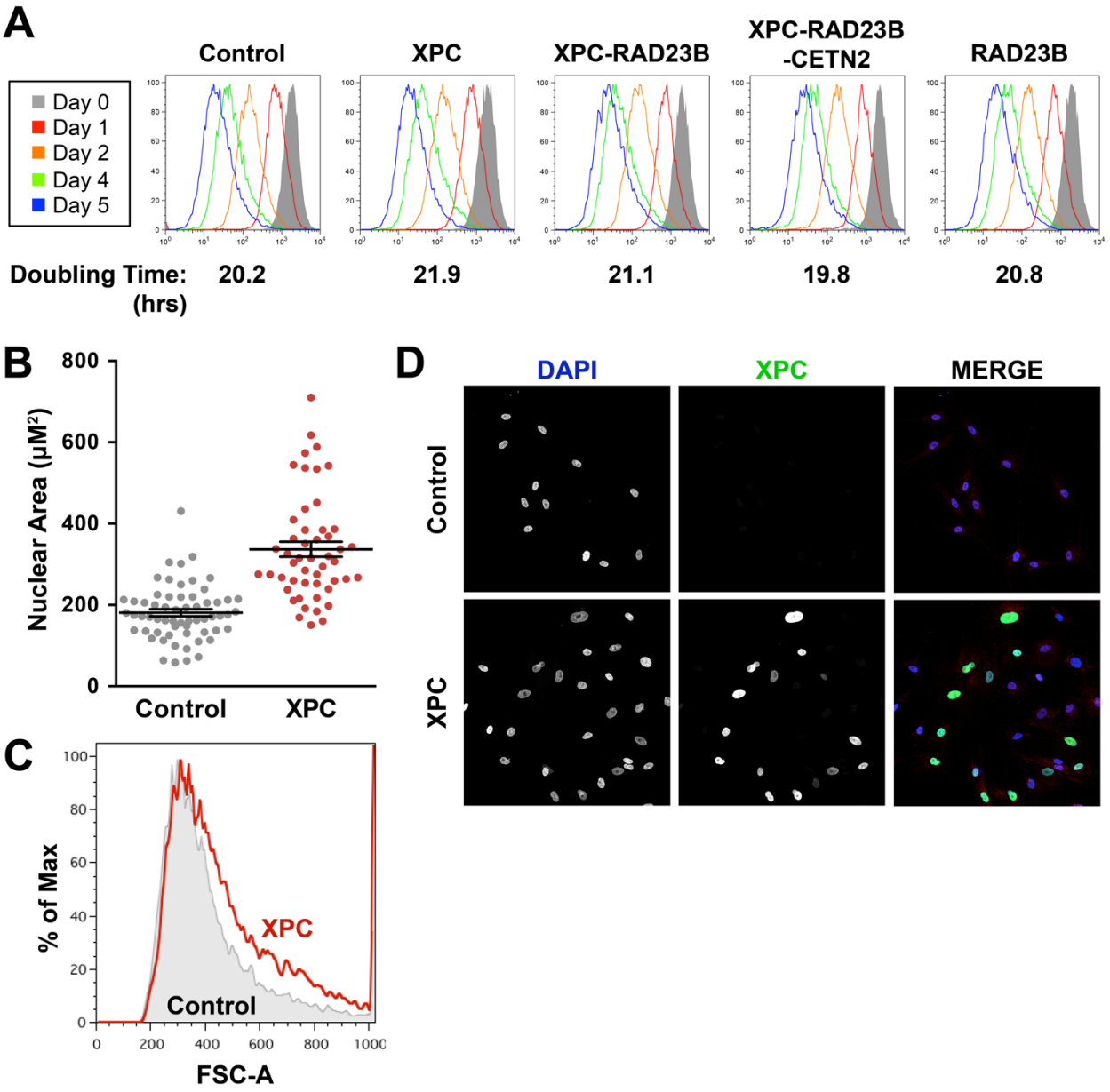

Figure 2. (**A**) HDFs expressing a control (mCherry) vector or various combinations of the XPC complex subunits have similar doublings times. The doubling time for each population was calculated using the CellTrace CFSE Proliferation Assay (Life Technologies), where the mean fluorescence intensity was calculated over a 5-day period. (**B**) Mean nuclear area of control or XPC expressing cells, calculated using the largest cross-section of each cell. (**C**) Effect of XPC overexpression on forward scatter. (**D**) Representative immunofluorescent images of control and XPC overexpressing HDFs, labeled with DAPI (blue), XPC (green), and COXIV (red). Images across the samples were taken at the same acquisition settings (excitation laser intensity, gain, and exposure time).

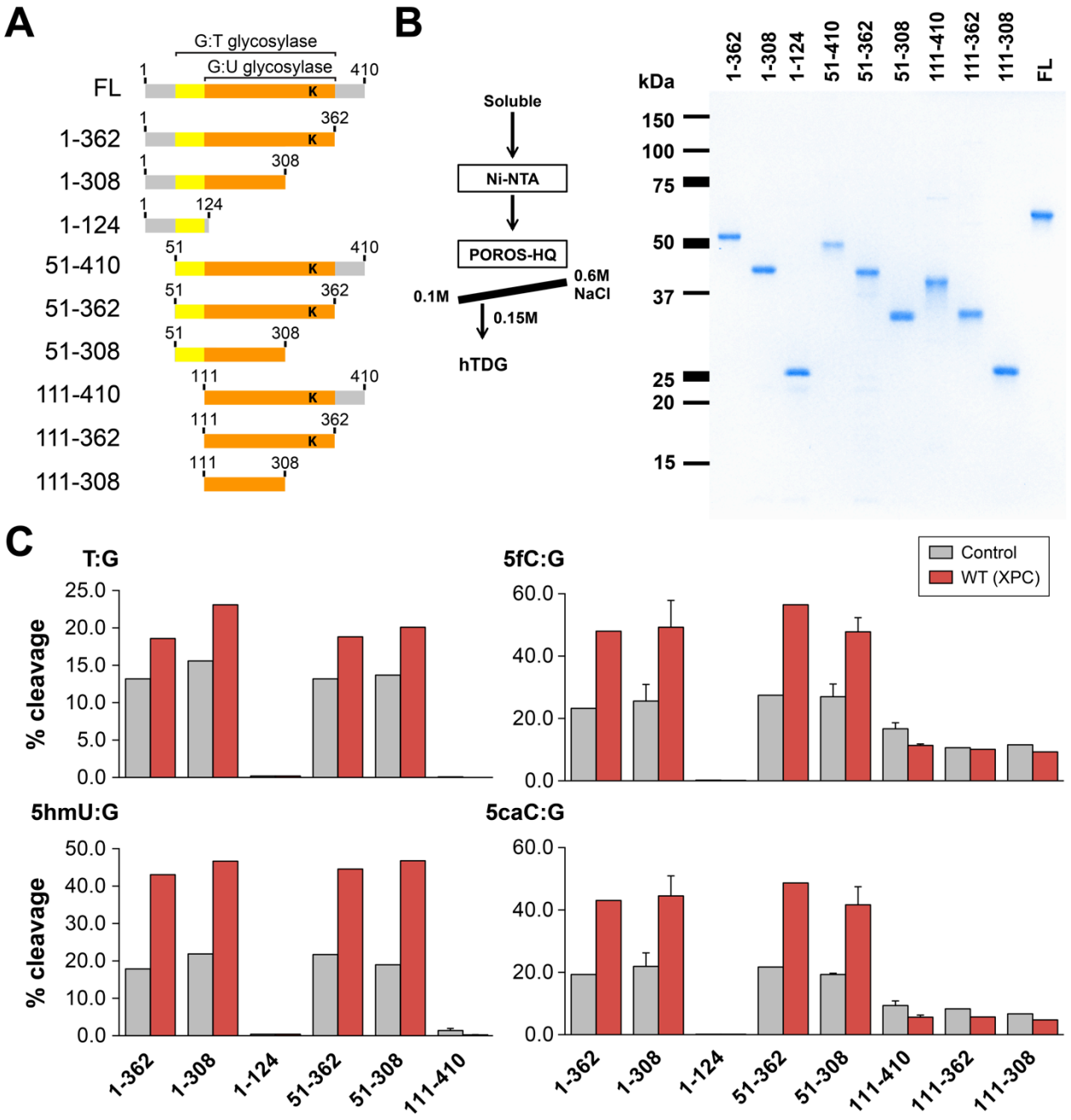

Figure 3. (**A**) Schematic representation of all human TDG truncations purified and tested using the in vitro glycosylase assay. (**B**) Schematic representation of the purification method used for recombinant TDG and SDS-PAGE and PageBlue staining of purified TDG truncations. (**C**) Relative glycosylase activity of TDG truncations in the presence or absence of XPC (WT) was assayed on all relevant DNA demethylation intermediates. Error bars represent the standard deviation.

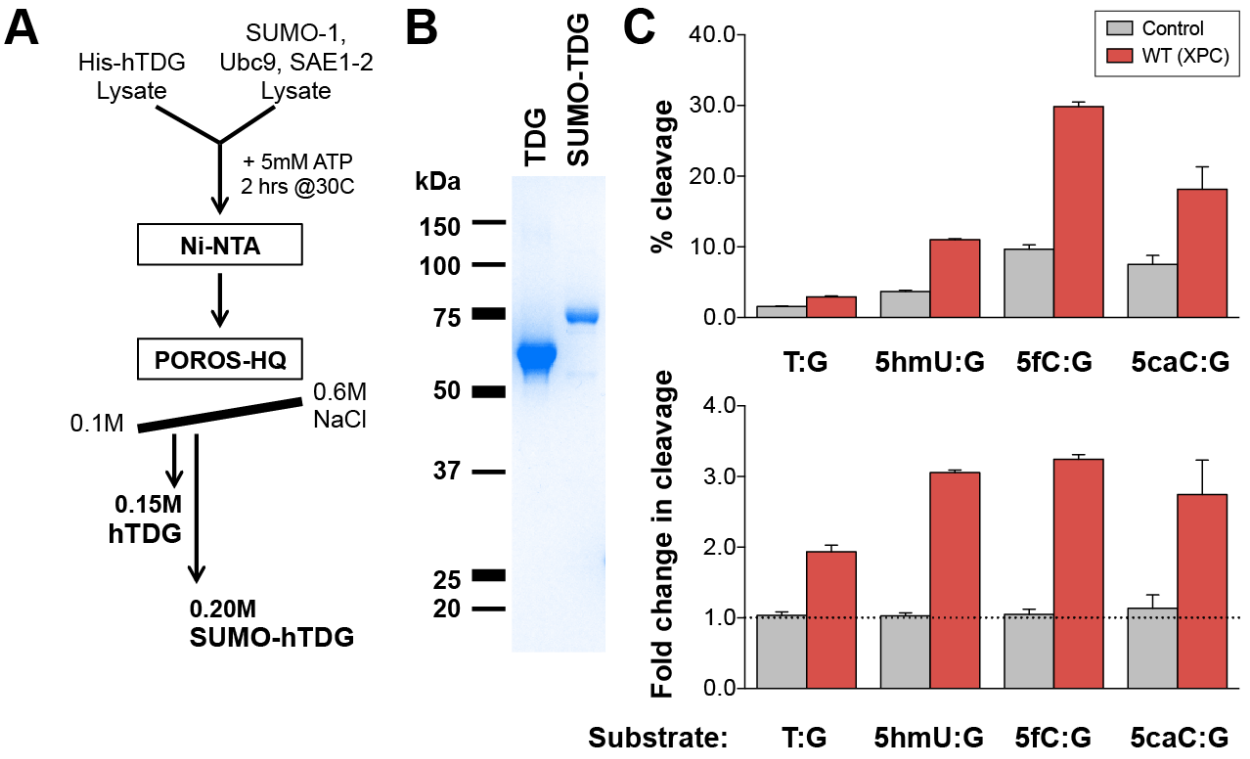

Figure 4. (**A**) In extract SUMOylation and subsequent purification scheme of SUMO-TDG. (**B**) SDS-PAGE and PageBlue staining of purified TDG and SUMO-TDG proteins. (**C**) Relative SUMO-TDG activity was determined by in vitro glycosylase assays in the presence or absence wildtype XPC. Graphs indicate amount of cleavage as total percentage of labeled substrate (top) or fold change normalized to the SUMO-TDG control (bottom). The 37mer dsDNA TDG substrate tested is designated under the graph. Error bars represent the standard deviation, n = 2.

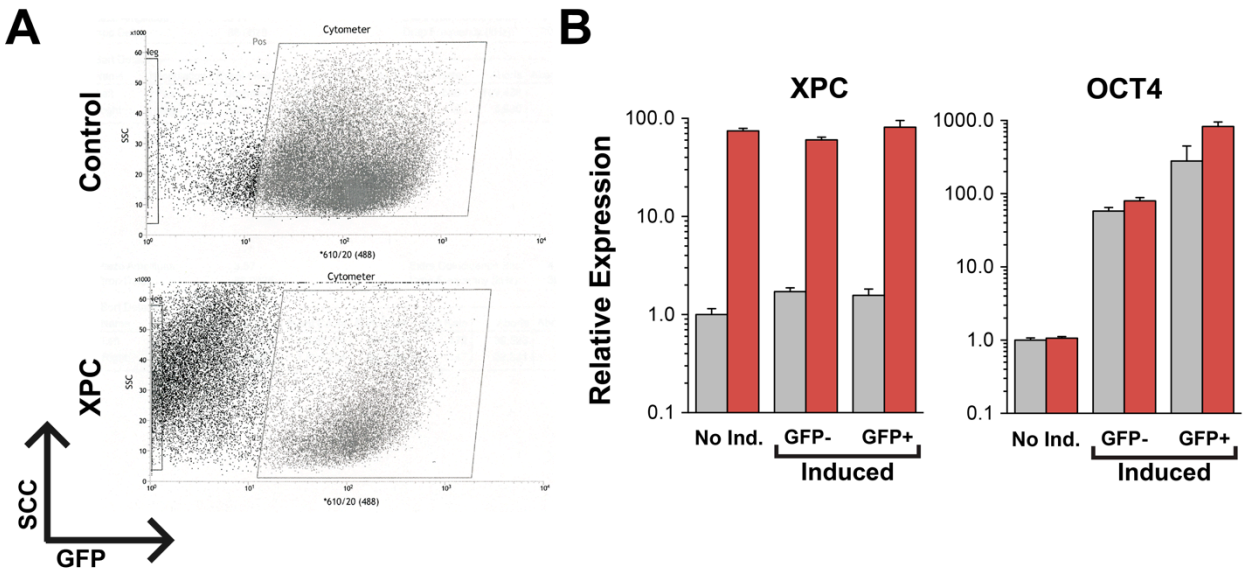

Figure 5. (**A**) Sorted pre-iPSC populations for MeDIP-seq experiment depicted in Chapter 2, Figure 6. (**B**) mRNA levels of XPC and Oct4 of uninduced HDFs (No Ind.) and partially reprogrammed cells sorted for transfection (GFP+). GFP- cells are largely non-transfected HDFs. Error bars represent standard deviation.

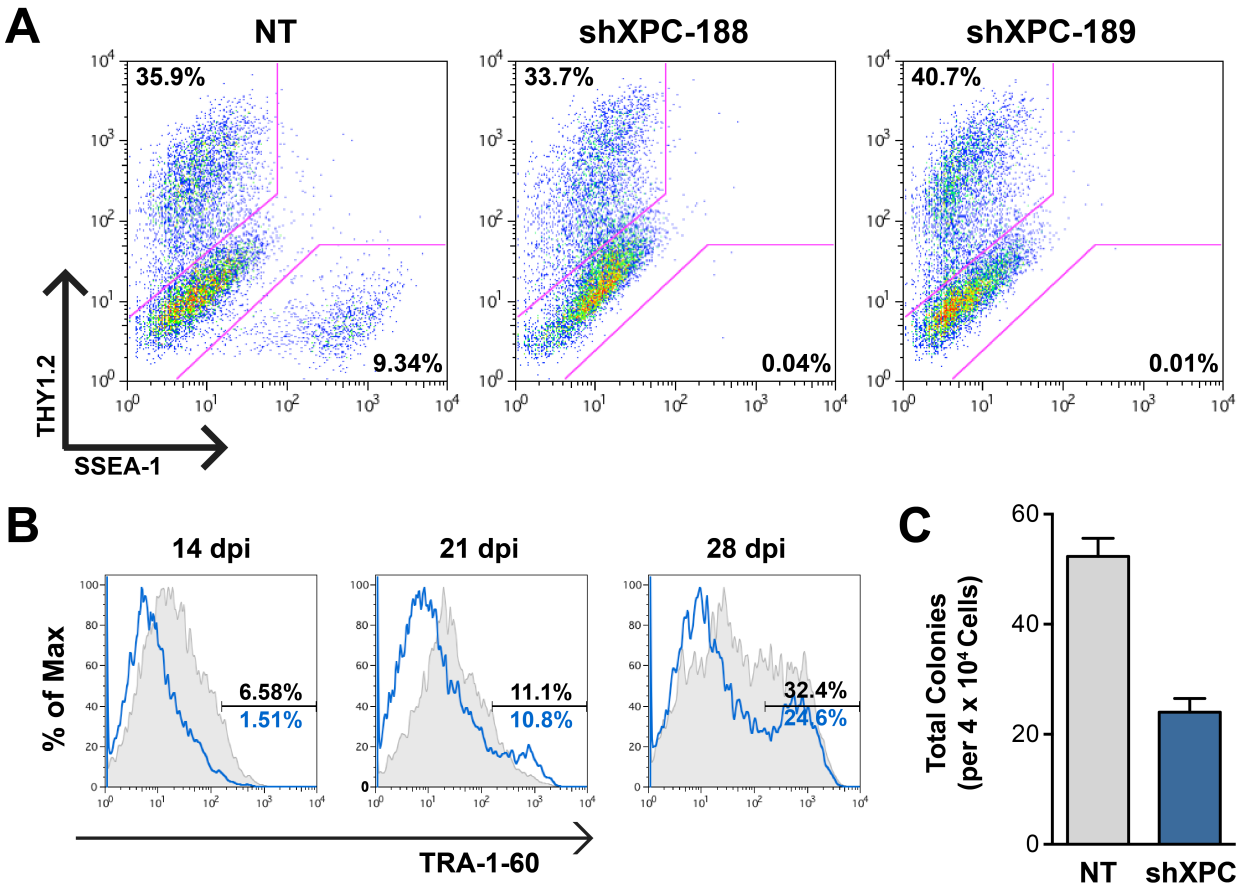

Figure 6. (**A**) Depletion of XPC in mouse embryonic fibroblasts (MEFs) abolishes reprogramming revealed by flow cytometry analysis of the fibroblast marker, Thy1.2, and the early stage mouse iPS marker, SSEA-1. (**B**) Flow cytometry analysis of the late stage human iPS marker, TRA-1-60, on HDFs expressing an shRNA against XPC at 14, 21, and 28 days post induction (dpi). (**C**) Average number of colonies between control (non-targeting) HDFs and HDFs expressing an shRNA against XPC. Colonies were scored by AP staining 28 dpi.

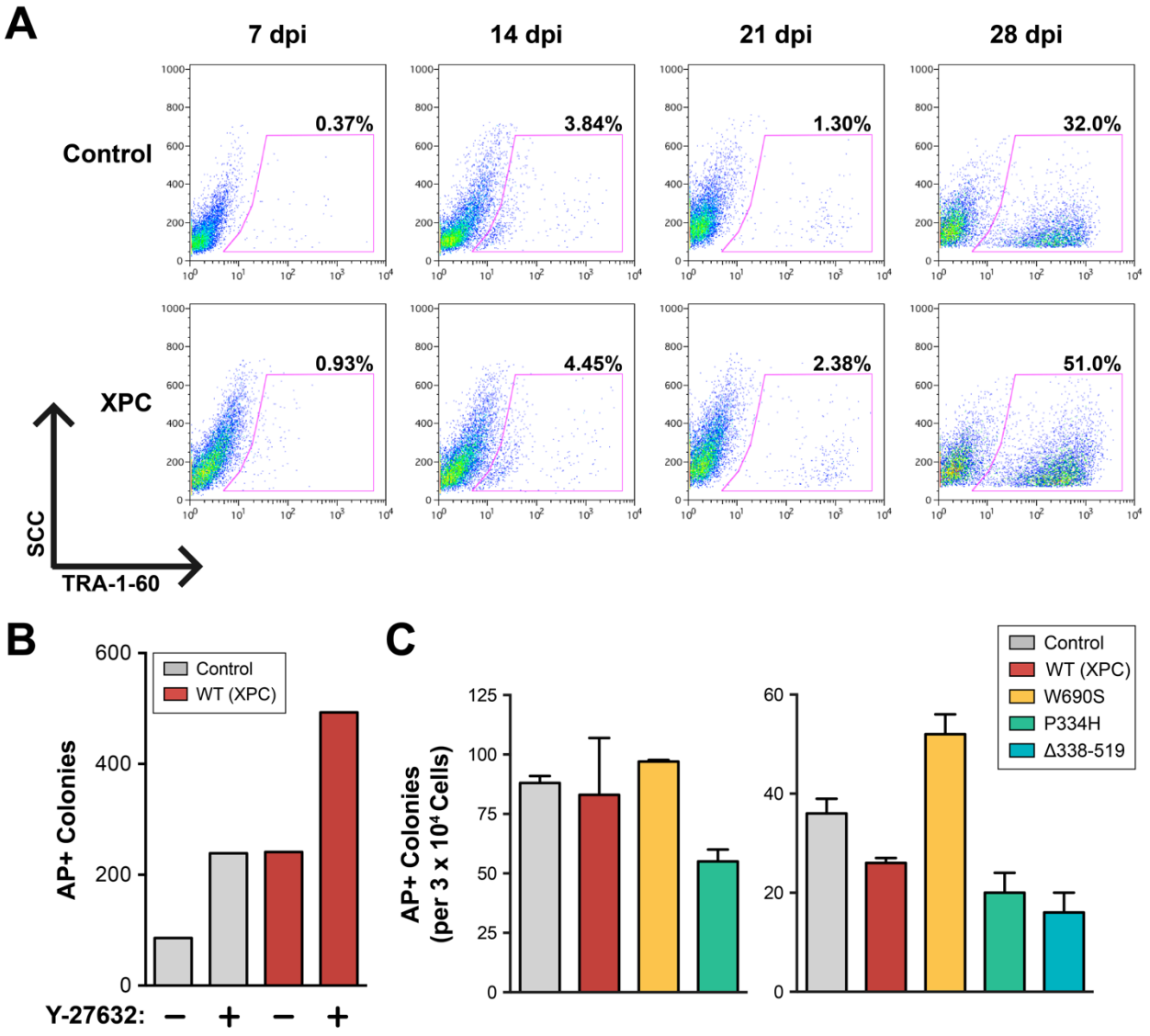

Figure 7. (**A**) Flow cytometry analysis of the late stage human iPS marker, TRA-1-60, on induced HDFs overexpressing the XPC subunit alone. dpi, days post induction. (**B**) iPSCs derived from control or XPC-RAD23B-CETN2 overexpression were challenged with single cell passaging and allowed to recover for 3-4 days before staining with alkaline phosphatase (AP). Treatment with the ROCK inhibitor, Y-27632, includes pre-treatment for 1 hour at 37C prior to trypsinization until 24 hours after plating. 2.5 x 10^4 cells were plated per 24-well. (**C**) Average number of AP+ colonies scored 28 dpi of HDFs expressing the wildtype XPC or various mutants (W690S, P334H, Δ338-519). Error bars represent the standard deviation, n = 3.

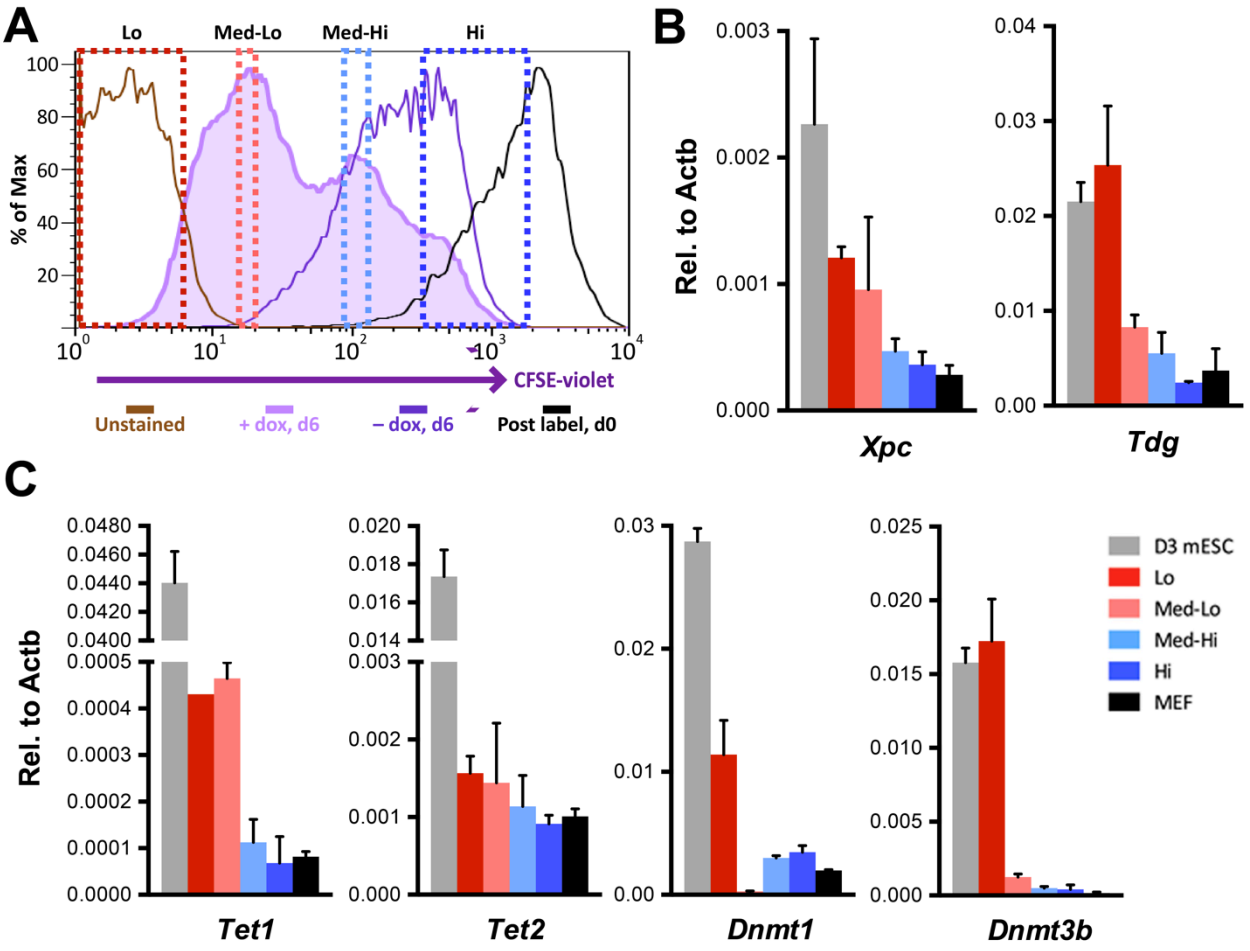

Figure 8. (**A**) Induced MEFs (light purple) are treated with doxycycline (dox) for 4 days, labeled with CFSE, and continuously cultured in the presence of dox for an additional 48 hours prior to FACS. Populations for ultrafast (Lo), fast (Med-Lo), medium slow (Med-Hi), and slow (Hi) cycling dox-induced MEFS are sorted based on CFSE intensity and denoted by dashed boxes. CFSE-intensity of MEFs immediately after labeling (black), unlabeled MEFs (brown), and uninduced MEFs 48 hours post-labeling (dark purple) are shown and used as controls. mRNA levels of (**B**) *Xpc* and *Tdg* and some (**C**) regulators of DNA methylation in sorted MEF populations (Lo-Hi) are compared to D3 ESCs and uninduced MEFs by qPCR. Results are normalized to *Actb*. Error bars represent standard deviation, n = 3.

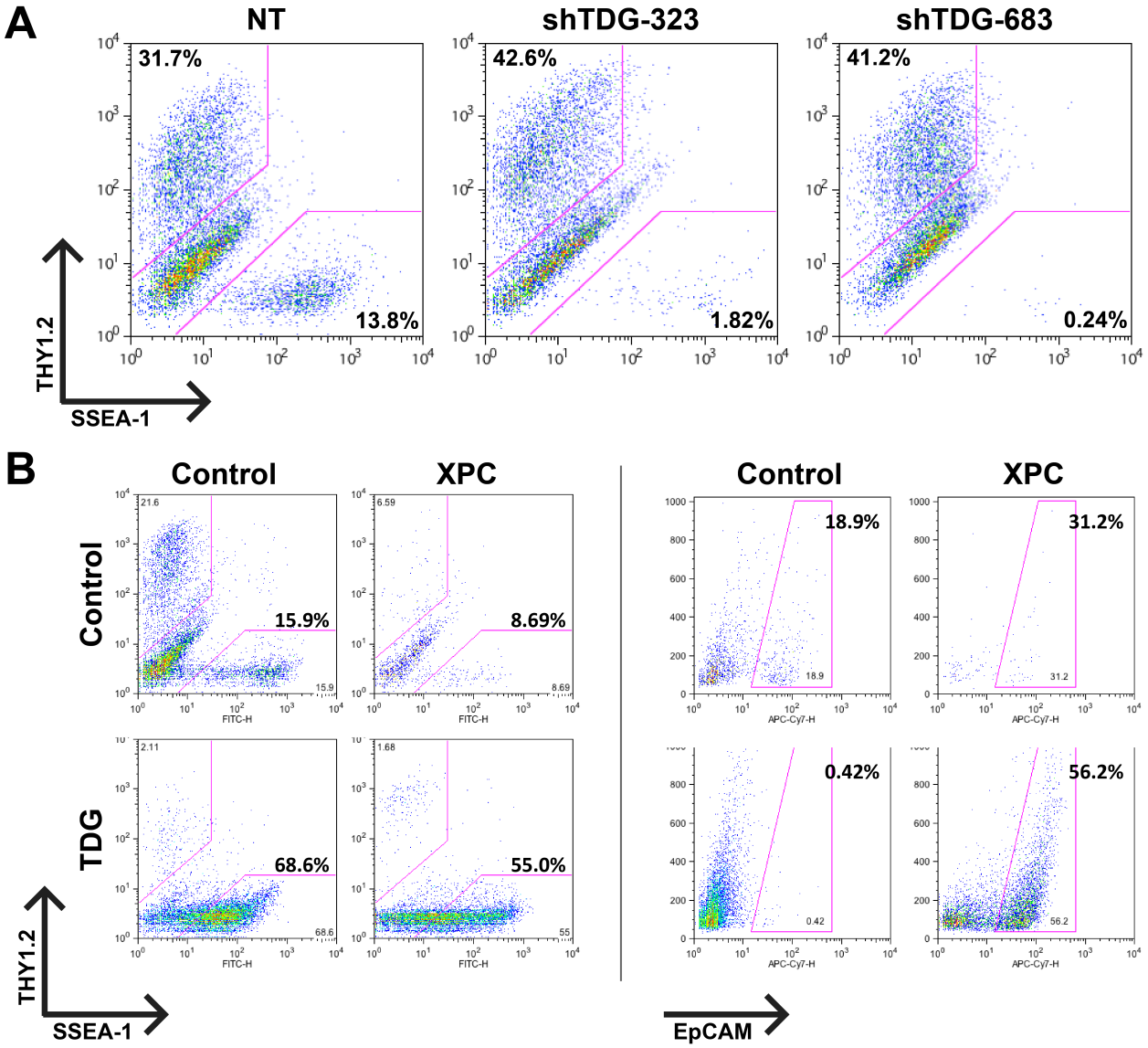

Figure 9. (**A**) Depletion of TDG in mouse embryonic fibroblasts (MEFs) abolishes reprogramming, revealed by flow cytometry analysis of the fibroblast marker, Thy1.2, and the early stage mouse iPS marker, SSEA-1. (**B**) Flow cytometry analysis of MEFs overexpressing human XPC, TDG, or XPC and TDG simultaneously. Gated Thy1.2-/SSEA-1+ populations express various levels of the late stage mouse iPS marker, EpCAM (right).

APPENDIX B

PRIMER SEQUENCES

RT-qPCR Primer Sequences

	Forward (5' → 3')	Reverse (5' → 3')
hCripto	AGAAGTGTTCCCTGTGTAAATGCTG	CACGAGGTGCTCATCCATCA
hDnmt3b	GCTAAGCTACACACAGGACTTGACAG	AGTTCGGACAGCTGGGCTTT
hGapdh	TCTGCTCCTCCTGTTCGACA	AAAAGCAGCCCTGGTGACC
hKlf4	GATGAACTGACCAGGCACTA	GTGGGTCATATCCACTGTCT
hNanog	CCAACATCCTGAACCTCAGCTAC	GCCTTCTGCGTCACACCATT
hOct4	TCGAGAACCGAGTGAGAGGC	CACACTCGGACCACATCCTTC
hRad23b	AGCCTGCAGAAAAGCCAGCAGA	CGTAAGACTGACCCGTCACAAGTGC
hSox2	CACACTGCCCCTCTCACACAT	CATTTCCCTCGTTTTTCTTTGAA
hStella	CGAATCTGTTTCCCCTCTATCG	CTCTCCTGCTGTAAAGCCACTC
hXPC	ACGCCAGAGCAGGCGAAGAC	AGGTGAACCTTGTGTGTGTCCTCA
mAbcf1	AGGTGGTGGCTGATGAAACAC	TAGCAACTTCAGTCGCTTGGTAT
mActb	GATCTGGCACCACACCTTCT	GGGGTGTTGAAGGTCTCAAA
mCdx2	AGGCTGAGCCATGAGGAGTA	CGAGGTCCATAATTCCACTCA
mDkc1	TTAGGACAACGACACCACCA	CCCAGCTGGACATAATGCTT
mDnmt1	ATCAGGTGTCAGAGCCCAAAG	TGGTGGAATCCTTCCGATAAC
mDnmt3b	CTCGCAAGGTGTGGGCTTTTGTAAC	CTGGGCATCTGTCATCTTTGCACC
mEcad	GCACTCTTCTCCTGGTCCTG	GTTGACCGTCCCTTCACAGT
mEpcam	GCTGGCAACAAGTTGCTCTCTGAA	CGTTGCACTGCTTGGCTTTGAAGA
mFgf4	GGGAGGCTACAGACAGCAAG	CTGTGAGCCACCAGACAGAA
mFgf5	TTGCGACCCAGGAGCTTAAT	CTACGCCTCTTTATTGCAGCAT
mGapdh	TCAATGAAGGGGTCGTTGAT	CGTCCCGTAGACAAAATGGT
mGata6	TTGCTCCGGTAACAGCAGTG	GTGGTCGCTTGTGTAGAAGGA
mKlf4	CAGTGGTAAGGTTTCTCGCC	GCCACCCACACTTGTGACTA
mNanog	CCTCAGCCTCCAGCAGATGC	CCGCTTGCACTTCATCCTTTG
mOct4	GAAGCAGAAGAGGATCACCTTG	TTCTTAAGGCTGAGCTGCAAG
mSlug	CACATTCGAACCCACACATTGCCT	TGTGCCCTCAGGTTTGATCTGTCT
mSnail	TTGTGTCTGCACGACCTGTGGAAA	TCTTCACATCCGAGTGGGTTTGGA
mSox2	GAGTGGAAACTTTTGTCCGAGA	GAAGCGTGTACTTATCCTTCTTCAT
mT	CTCTAATGTCCTCCCTTGTTGCC	TGCAGATTGTCTTTGGCTACTTTG
mTdg	CTTTAGCTGTGGCAGTGATGG	ATGCCCATGTAAACAGCAGTG
mTet1	CCATTCTCACAAGGACATTCACA	GCAGGACGTGGAGTTGTTCA
mTet2	GCCATTCTCAGGAGTCACTGC	CTTCTCGATTGTCTTCTCTATTGAGG
mXpc	CTGGATGACCGCAACCCGCA	GCTGGTCCAGGTGCTTCGCC
mZeb1	GCTGGCAAGACAACGTGAAAG	GCCTCAGGATAAATGACGGC

ChIP-qPCR Primer Sequences

	Forward (5' → 3')	Reverse (5' → 3')
mOct4 -2000	GGAACTGGGTGTGGGGAGGTTGTA	AGCAGATTAAGGAAGGGCTAGGACGAGAG
mOct4 -251	AGCAACTGGTTTGTGAGGTGTCCGGTGAC	TCCCCAATCCCACCCTCTAGCCTTGAC
mNanog -199	GGGTCACCTTACAGCTTCTTTTGCATTA	GGCTCAAGGCGATAGATTTAAAGGGTAG
mNanog -952	GGCAAACTTTGAACTTGGGATGTGGAAATA	CTCAGCCGTCTAAGCAATGGAAGAAGAAAT
mSox2 +3515	TTTTCGTTTTTAGGGTAAGGTACTGG	CGTGAATAATCCTATATGCATCACAAT
mSox2 -3581	CCCTGTTCCAAGTCTCTTTCTG	GATTTCAATCCAACACCATCATAG
mFgf4 -1581	CCCCAAAGCAGTTTGATGAT	CCTTAGTCTGGGCACTCCTG
mFgf4 +2932	GGGAGGCTACAGACAGCAAG	CTGTGAGCCACCAGACAGAA
hNanog -190	GCTGGGTTTGTCTTCAGGTTC	CACACCCCCTACTGACCCAC
hNanog -5000	CCCTACCCCAACCTCCATTA	GGAGGCACAGTGAGACCTTG
hOct4 -2000	GGTAGATTATGGGGCCTGGT	TGTGGAGATTCCAGCCAAAT
hOct4 +1500	TCTGTTTTGGGGTTTTGGAA	TGGCTGTGTGCTCCGTTTAT

In vitro glycosylase assay sequence

37mer substrate	
Forward	**/Cy3/GAGTCATGCCATTGGCCACATXGTGTCAGCTAGGATT**
Reverse	**CAATCCTAGCTGACACGATGTGGCCAATGGCATGACT**

where **X is replaced with T, 5hmU, 5fC, or 5caC