Dissecting Molecular Similarities and Differences Between Pluripotent Stem Cell Lines

Dissecting Molecular Similarities and Differences Between Pluripotent Stem Cell Lines

A dissertation presented

by

Jiho Choi

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Developmental and Regenerative Biology

Harvard University

Cambridge, Massachusetts

April 2015

Dissertation Advisor: Dr. Konrad Hochedlinger Author: Jiho Choi

Dissecting Molecular Similarities and Differences Between Pluripotent Stem Cell Lines

Abstract

Traditionally, pluripotent stem cells are derived from preimplantation embryos and fetal germ cells, which give rise to embryonic stem cells (ESCs) and embryonic germ cells (EGCs), respectively, In contrast, induced pluripotent stem cells (iPSCs) are derived from somatic cells upon overexpression of defined transcription factors such as *Oct4, Sox2, Klf4* and *c-Myc*. Despite their origin from different cell types, all of these pluripotent stem cell lines share the ability to self-renew indefinitely in culture while retaining the capacity to differentiate into derivatives of all three germ layers. Because pluripotent cells provide a useful tool in basic research and cell therapy, it is critical to understand the molecular similarities and differences between ESCs, EGCs and iPSCs.

The studies presented in this thesis aim to address the equivalence of different pluripotent cell types. In the first study, we performed a systematic comparison of DNA methylation and gene expression patterns between isogenic mouse ESCs and EGCs. Surprisingly, we found that global DNA methylation patterns were indistinguishable between ESC and EGC lines of the same sex, while female cell lines exhibited global hypomethylation compared to male cell lines. Mechanistically, upregulation of the X-linked gene, *dual specificity phosphatase 9* (*Dusp9*) in female cells attenuated MAP kinase signaling, resulting in global DNA hypomethylation via the reduction of Dnmt3a and Dnmt3b protein levels. In the second study, we compared isogenic, transgene-free hESC and hiPSC lines to determine whether molecular differences exist between hiPSC and hESC lines when controlling for genetic background and reprogramming

methodology. Strikingly, transcriptional variation between different genetic backgrounds was greater than variation observed between cell types (i.e., hiPSCs compared to hESCs). Moreover, the few transcriptional differences observed between isogenic hESC and hiPSC lines had no apparent functional consequences and these genes were not identified during the comparison of a larger set of independently derived non-isogenic hESC/hiPSC lines. We conclude that hESCs and hiPSCs are highly similar on a transcriptional and functional level and cannot be distinguished by a defined gene expression signature.

Together, our data demonstrated that sex rather than cell type of origin drives global epigenetic and transcriptional patterns in conventional mouse pluripotent cell lines. These results provide fundamental insights into the epigenetic regulatory mechanisms that govern pluripotency. Additionally, our comparison of isogenic hiPSCs and hESCs supports the view that cellular reprogramming technologies faithfully reset the transcriptional pattern of somatic cells and establish a pluripotent state that is molecularly and functionally equivalent to embryo-derived stem cells. These findings may provide the basis for future mechanistic studies and help to translate iPSC technologies into a therapeutic setting.

This thesis is dedicated to my lovely wife and family

for their constant love and support

Table of Contents

Chapter 1

Introduction

I. Pluripotency and Stem Cells

Cells of the early mammalian preimplantation embryo retain the remarkable capacity to differentiate into all cell types of the adult animal, a potential that is referred to as "pluripotency". During postimplantation development, pluripotent cells progressively differentiate to form all the specialized cell types of the adult body. The transient nature of pluripotent cells in early embryogenesis provides a challenge for dissecting the underlying molecular mechanisms. However, pluripotent stem cells can be derived and maintained *in vitro* from different sources such as germ cell tumors, preimplantation embryos and somatic cells following experimental manipulation. These pluripotent cell lines can be propagated indefinitely in culture (i.e. self-renewal), while maintaining the potential to give rise to all somatic cell types (i.e. differentiation); these are the hallmarks of stem cells. In the following sections, I will briefly summarize the major types of pluripotent stem cells.

1. Mouse Embryonal Carcinoma Cells

In 1954, Stevens and Little found that inbred strains of mice exhibit a high incident of teratocarcinomas, tumors of germ cell origin that contain cell types from all three germlayers, mesoderm, ectoderm and endoderm[1]. A single teratocarcinoma cell was able to initiate another tumor containing all three germ layers upon transplantation. Based on this clonal stem cell assay, the authors hypothesized the existence of pluripotent stem cells in these teratocarcinomas and designated the putative pluripotent stem cells "Embryonal Carcinoma Cells (ECCs)."[2] Subsequently, ECCs were successfully isolated and propagated *in vitro.* These cultured cell lines also developed into tumors containing a variety of differentiated cell types when subcutaneously injected into mice[3]. Moreover, certain ECC lines contributed to chimeras when injected into blastosytsts[4,5]. Therefore, the isolation of pluripotent stem cells *in vitro* culture has provided a

new and practical tool not only for dissecting the molecular mechanisms of pluripotency but also for genetic analyses of mammalian development.

2. Mouse Embryonic Stem Cells

Pioneering work on ECCs laid the foundation for the isolation of pluripotent stem cells directly from developing embryos. Mouse diapause blastocysts were initially explanted onto growth arrested fibroblasts, giving rise to pluripotent stem cells[3], which were termed "Embryonic Stem Cells (ESCs)."[6] ESCs, which are derived from the inner cell mass (ICM) of preimplantation blastocysts, can self-renew indefinitely in vitro while maintaining their pluripotent state. They are also capable of contributing to chimeric embryos when introduced into blastocysts[7]. Furthermore, ESCs can form a whole embryo when introduced into a tetraploid blastocyst whose constituent cells will only contribute to extraembryonic lineages but not embryonic lineages[8]. Resultant embryos and mice are completely derived from the injected ESCs using this "tetraploid complementation" assay.

3. Mouse Embryonic Germ Cells

Teratomas were historically defined as germ cell tumors that originate from "embryonic totipotent cells that have escaped the influence of organizers"[9]. The origin of teratomas from germ line cells was experimentally proven by injecting primordial germ cells (PGCs) from genital ridges at embryonic day 12.5 (E12.5) into the testes of strain 129 mice [10]. This finding raised the possibility that pluripotent stem cells could be directly derived from the genital ridges. Indeed, pluripotent stem cells were successfully generated from explanted PGCs from the genital ridges at E8.5[11] and E12.5[12], respectively. These cells, called "Embryonic Germ Cells (EGCs)"[11], are essentially indistinguishable from ESCs except for DNA methylation differences at imprinted loci, which reflects their origin from the germline where imprints are erased during

gametogenesis. EGCs, like ECCs and ESCs, form teratomas when injected into nude mice and contribute to germline-transmitting chimeras when injected into a host blastocysts[13,14].

II. Signaling Pathways and Pluripotent States

1. Signaling Pathways in Pluripotent Stem Cells

1.1. LIF/JAK/STAT3 Signaling Pathway

Initially, mouse ECCs and ESCs were cultured on layers of mitotically inactive mouse fibroblast feeders[3,15]. Feeder cells were essential to support the undifferentiated state of pluripotent stem cells, indicating that fibroblasts produce a signal that prevents differentiation. It was subsequently recognized that soluble factors produced by non-dividing STO cells[16] or a Buffalo liver cell line[17] also obviated the requirement for fibroblast feeders to sustain the undifferentiated state of ECCs and ESCs. These findings led to the identification of leukemia inhibitory factor (LIF), which specifically suppresses spontaneous differentiation of ESCs *in vitro* in the absence of feeder cells[18,19].

LIF is a member of the Interleukin-6 cytokine family and acts through the LIF receptor (LIFR), which heterodimerizes with the coreceptor gp130 upon binding[20]. This heterodimer then activates the Janus kinase (JAK), which reciprocally phosphorylates gp130. Subsequently, activated gp130 recruits the transcription factor, Signal Transducers and Activators of Transcription 3 (STAT3), whose concomitant phosphorylation and activation by JAK plays a central role in the maintenance of the pluripotent phenotype of ESCs[21].

Phosphorylated STAT3 translocates to the nucleus in ESCs and activates a number of target gene, such as *Klf4*[22,23], *Gbx2*[24], and *Tfcp2l1*[25,26]. Of note, overexpression of constitutively active *Stat3* or *Tfcp2l1* is sufficient to sustain self-renewal in the absence of LIF, while forced expression of *Klf4* and *Gbx2* cannot fully recapitulate STAT3 activity[21,26]. However, activation of

STAT3 alone is not sufficient to block spontaneous ESC differentiation in the absence of serum, indicating that additional factors present in serum maintain the undifferentiated state of ESCs.

1.2. TGF-β Signaling Pathway

Bone Morphogenetic Proteins (BMPs) were recognized to be the key signals provided by serum to maintain ESC self renewal in combination with LIF[27]. BMPs are members of the TGF-β superfamily, which includes TGF-β, Activin, and Nodal. TGF-β proteins signal through multiple effectors, such as the cell surface receptors "type I" and "type II" as well as SMADs. BMP binding initiates the formation of a heterotetramer between pre-dimerized type I and type II receptors, which subsequently activates intracellular SMAD proteins (SMAD1/5/8)[28]. In mouse ESCs, activated SMAD1/5/8 associated with SMAD4 translocate into the nucleus to regulate target gene expression. Inhibitors of differentiation (*Id*) genes are well-known downstream targets of BMP and overexpression of *Id* genes indeed blocks neural commitment of ESCs in the absence of serum. However, Id-expressing cells readily differentiate into non-neural lineages in the absence of LIF. Therefore, ESC self renewal requires both BMP signals to inhibit neural commitment and LIF/STAT3 signaling to block non-neural lineage differentiation[27].

1.3. Suppression of MAPK & GSK3 Pathways As an Alternative Growth Condition to Serum and LIF

The investigation of additional signaling pathways in ESCs recently led to an alternative culture system for ESCs that no longer requires serum and LIF, as discussed below. Activation of the Mitogen-Activated Protein (MAP) kinases ERK1 and ERK2 have been associated with the differentiation of mouse ESC. Accordingly, suppression of the upstream kinase, Mitogen/Extracellular signal-regulated kinase (MEK), was shown to enhance the propagation of undifferentiated ESCs in serum-LIF conditions[29,30]. Likewise, the Wnt pathway has recently been associated with ESC self renewal. Wnt-mediated activation of the transmembrane receptor

Frizzled (Fz) prevents phosphorylation of β-catenin by GSK3 and thus proteosomal degradation[31]. Consequently, β-catenin accumulates in the cytoplasm and translocates into the nucleus, forming a complex with the transcriptional repressor Tcf3[32]. The interaction between β-catenin and Tcf3 disrupts Tcf3-mediated repression of pluripotency genes, thus stabilizing the pluripotent program[33,34].

Importantly, inhibition of Tcf3 via the GSK3 inhibitor (CHIR99021) is sufficient to keep ESCs undifferentiated in the presence of the MEK inhibitor (PD0325901)[35]. This alternative culture system for ESCs is commonly referred to as "2i". The 2i culture system harnesses ESCs in a distinct transcriptional state; while ESCs cultured in conventional conditions ("serum-LIF") maintain heterogeneous expression patterns of key pluripotency factors, ESCs in 2i exhibit a uniform expression of these factors[35-38]. Functionally, this homogeneous expression of pluripotency factors renders ESCs in a stabilized pluripotent state, where ESCs consistently form high-contribution chimeras with notably greater efficiency than ESCs in serum-LIF[39,40].

2. Transcriptional Networks of Pluripotency

The pluripotent state is regulated by complex, intertwined signaling pathways, which activate transcription factors and cofactors that play a central role in the maintenance of the pluripotent state.

2.1. Core Pluripotency Networks

Transcription factors (TFs) recognize specific DNA sequences to either activate or inhibit transcription of target genes. TFs bind both to core promoters and to distal enhancers. Enhancers bound by multiple TFs concomitantly recruit cofactors forming a multiprotein complex, which interacts with RNA polymerase II to induce transcription[41].

The TFs Oct4 and Sox2 are required for the maintenance of pluripotency and thus defined as "core" pluripotency factors[42-44]. In general, core pluripotency TFs either activate pluripotency-associated genes or repress lineage-specific genes to maintain ESCs in an undifferentiated state that is poised for differentiation[45]. Moreover, precisely regulated levels of *Oct4* and *Sox2* are critical to the balance between self-renewal and differentiation[44]. Indeed, overexpression of *Oct4* induces differentiation into primitive endoderm and mesoderm whereas repression results in trophectoderm differentiation[46]. Similar to *Oct4*-null ESCs, *Sox2*-null cells differentiate into trophectoderm-like cells, a phenotype which can be rescued by *Oct4* overexpression[47,48].

In addition to Oct4 and Sox2, Nanog is considered a key TF during the acquisition of pluripotency; however, its role in the maintenance of pluripotency is more complex[49]. Recent findings demonstrated that *Nanog* is dispensable for the *in vitro* establishment of pluripotency, which implies that *Nanog* plays a distinct role from *Oct4* and *Sox2*[50,51]. Nevertheless, Oct4, Sox2, and Nanog cooperatively regulate the pluripotency expression program in ESCs by binding overlapping genomic targets including their own promoters and/or by recruiting diverse TFs and cofactors. Prominent targets of Oct4, Sox2, and Nanog are co-bound by STAT3, β-catenin, and SMAD1, which are the effectors of key signaling pathways such as LIF, Wnt and BMP4, thus providing a link between extracellular signals and intracellular TFs required for ESC self renewal[42,43,52].

2.2. Auxiliary Pluripotency Networks

In addition to the core pluripotency TFs Oct4, Sox2 and Nanog, ESCs also depend on "auxiliary" TFs, such as Esrrb, Sall4, c-Myc, Tbx3, and Prdm14 that augment the pluripotency networks but are individually dispensable for the maintenance of pluripotency. For example, overexpression of *Esrrb* phenocopies GSK3 inhibition in 2i-LIF ESCs[53]. However, in serum-LIF, *Esrrb* is dispensable for self-renewal. Another "auxiliary" TF c-Myc also promotes self-renewal

by blocking differentiation. Sustained levels of c-Myc maintain self-renewal in the absence of LIF while a dominant negative form of c-Myc antagonizes self-renewal and promoted differentiation [54]. Moreover, c-Myc co-binds a large portion of genes that are bound by Oct4, Sox2, and Nanog, thus ensuring that these genes are fully transcribed[47,55]. Consistent with these findings, overexpression of *c-Myc* significantly enhances the *in vitro* acquisition of pluripotency[56,57].

3. Metastable Pluripotent States

Pluripotent stem cells in culture have been a surrogate model to study early embryogenesis. *In vivo*, pluripotent cell populations exist until the late blastocyst stage. The ICM cells of the late blastocyst give rise to the epiblast and hypoblast that are the origin of future embryonic and extraembryonic lineages, respectively[49]. While the pre-implantation epiblast cells are considered to be in a "ground state", reflecting their unbiased developmental potential, the post-implantation epiblast cells are in a "primed state", which has restricted developmental potential and is poised to differentiate[57].

Mouse ESCs in serum-LIF reportedly exist in a heterogeneous, metastable state encompassing both the naïve state, which is functionally equivalent to the ground state of the epiblast, and the more developmentally advanced primed state. Thus, mouse ESCs in serum-LIF are in a dynamic, metastable equilibrium between the two pluripotent states[58-60]. In contrast, 2i-LIF culture endows ESCs with a uniform naïve pluripotent state[61].

The ground state epiblast normally generates the entire fetus. In fact, a single epiblast cell can contribute to all lineages when injected into a host blastocyst[57]. Similarly, naïve ESCs can participate in preimplantation development and generate all three germ layer derivatives when introduced into a blastocyst[7]. A key feature that is common to both naïve mouse ESCs cultured *in*

vitro and the ground state epiblast *in vivo* is global DNA hypomethylation[36,38,62-64], which likely reflects epigenetic resetting of pluripotent cells towards a blank state that supports differentiation into all somatic lineages. These observations therefore raise fundamental questions about the mechanisms that link culture conditions, developmental signals and global DNA hypomethylation.

III. DNA Methylation Patterns in Mouse Pluripotent Stem Cells

In mammals, the developmental potential of a fertilized embryo becomes progressively restricted by epigenetic modifications, ultimately producing the diverse cell types that constitute somatic tissues[65,66]. DNA methylation, which generally occurs on the fifth position of cytosine in the context of CpG dinucleotides, is the best studied epigenetic modification and is generally believed to repress promoter activity[67], although exceptions to this rule exist. DNA methylation patterns are faithfully propagated during cell divisions by DNA methyltransferases (Dnmts) to maintain cellular identity, genomic imprints as well as to repress transposable elements[68].

1. DNA Methyltransferases

In mammals, DNA methylatransferases 1 (Dnmt1), Dnmt3a and Dnmt3b, mediate the transfer of a methyl group from S-adenosylmethionine (SAM) to cytosine with the help of the regulatory protein, DNMT3L, which lacks catalytic activity[69,70]. Dnmt1 is primarily a maintenance methyltransferase that localizes to DNA replication foci and methylates hemimethylated CpG dinucleotides through interaction with the cofactor Uhrf1[71]. In contrast, Dnmt3a and Dnmt3b are important for establishing DNA methylation patterns *de novo* at repressed promoters in complex with other epigenetic repressors[62,65,68]. Although Dnmt3l is catalytically inactive, it enhances *de novo* methylation and increases the binding of SAM to Dnmt3a/b[72].

2. DNA Methylation Patterns during Embryo Development

During early embryogenesis, DNA methylation is dynamically remodeled leading to a globally hypomethylated state in the ICM. Subsequently, *de novo* DNA methylation occurs in a lineage-specific manner. The global erasure of DNA methylation during preimplantation development coincides with the acquisition of pluriopotency[62,65]. At this point, expression of pluripotency-associated genes is initiated while developmental genes remain repressed. In contrast, the pluripotency-associated loci are re-methylated in the committed cell types.

In the post-implantation embryo, a second round of comprehensive DNA demethylation occurs in PGCs[73]. PGCs are specified from the post-implantation epiblast and subsequently migrate into the genital ridges where they undergo global DNA demethylation at E11.5-12.5, which includes imprinted loci[74]. Consequently, EGCs derived from this developmental stage are hypomethylated at imprinted loci and non-imprinted loci[75]. EGCs derived from premigratory PGCs (E8.5) have also been reported to lack methylation imprints, albeit to a lesser extent than EGCs derived from later stages (E11.5-E12.5)[76].

3. DNA Methylation Patterns in Naïve Pluripotent Cells

Although PGCs emerge from the post-implantation epiblast, which is in a primed pluripotent state, EGCs derived from PGCs are functionally similar to ESCs[13,14]. While ESCs cultured in serum-LIF maintain relatively high methylation levels despite their origin from the otherwise hypomethylated ICM[77], both ESCs and EGCs acquire global DNA hypomethylation when switched to 2i-LIF conditions[36]. These observations imply that not only ground state epiblast but also PGCs are capable of acquiring naïve pluripotency *ex vivo* under permissive culture conditions.

4. DNA Methylation Patterns in Female Pluripotent Stem Cells

Global DNA hypomethylation in the epiblast represents the establishment of a transcriptionally permissive epigenome, which is a fundamental feature of naïve pluripotency[36,78-80]. This state is characterized by reactivation of the silenced X chromosome in female embryos. The paternal X chromosome is reactivated in the ICM[81] and subsequently, one X chromosome is randomly inactivated at the epiblast stage[82].

Because the naïve pluripotent state is transient, the study of sex-specific epigenomic differences *in vivo* is challenging. Thus, female pluripotent stem cells carrying two active X chromosomes (XX) provide a useful surrogate model to dissect the mechanism of X inactivation and DNA methylation. Notably, XX ESCs/EGCs in serum-LIF exhibit globally reduced DNA methylation levels compared to XY ESCs/EGCs[76,83-85]. XX ESCs spontaneously lose one of the two X chromosomes with extended culture and subsequently become XO; methylation levels in XO ESCs are restored to that in XY ESCs [83-85]. Therefore, hypomethylation in XX ESCs is due to the presence of two active X chromosomes rather than to the absence of a Y chromosome[83-85].

5. Mechanisms Underlying Global DNA Hypomethylation in XX ESCs and 2i culture

Global hypomethylation induced by the inhibition of MAPK and GSK3 pathways in 2i-LIF has been associated with the repression of DNA methyltransferases[78-80,85]. Similarly, XX ESCs cultured in serum-LIF seem to stabilize a naïve pluripotent state by inhibiting components of the MAPK and GSK3 pathways, leading to downregulation of *de novo* DNMTs[85]. However, the molecular mechanisms by which the presence of two active X chromosomes in female ESCs induces downregulation of the MAPK pathways and DNMT expression resulting in global hypomethylation remains to be elucidated.

IV. Human Pluripotent Stem Cells

1. Human Pluripotent Stem Cells

1.1. Human Embryonic Stem Cells

The first human ESC lines were derived from the ICM cells of blastocysts produced by *in vitro* fertilization (IVF)[86]. Unlike mouse pluripotent stem cells, a stringent test for pluripotency, such as chimera formation, is not available in humans. The current functional gold standard for hESCs is the evaluation of teratoma formation[87]. In addition, pluripotent cells can be *in vitro* differentiated into embryoid bodies (EBs), spherical structures resembling the gastrulating embryo, which can be assessed for markers of each of the three germ layers[88]. Combined with the teratoma formation assay, EB formation provides a comprehensive evaluation of the developmental potential of hESC lines.

While mouse ESCs have provided a platform to study mammalian development *in vitro*, human ESCs allow researchers to study human development, disease mechanisms, and ultimately use these cells for regenerative medicine. However, hESCs face two major obstacles to realize their full potential: (i) the derivation of hESCs, which requires the destruction of human embryos, entails ethical issues, and (ii) when hESCs are used for cell therapy, recipient patients may elicit an immune response. Recently, the successful derivation of somatic cell nuclear transfer (SCNT)-derived hESCs circumvented the issue of immune rejection by generating patient-specific hESC lines[89]. However, the use of human oocytes and the necessity to destroy blastocysts still pose ethical issues.

1.2. Human Induced Pluripotent Stem Cells

Development and cell fate determination are highly regulated cellular processes, during which cells transit from a pluripotent to one of many somatic states. Since this process is unidirectional, differentiated cell types cannot re-acquire pluripotency. However, the recent

finding of "induced pluripotency" offers a way to experimentally reverse cellular differentiation[90]. By forced expression of only four pluripotency genes, (*OCT4/SOX2/KLF4/c-MYC* or *OCT4/SOX2/LIN28/NANOG*) adult fibroblasts can be reprogrammed to pluripotent stem cells, which are termed human induced pluripotent stem cells (hiPSCs)[91,92].

Subsequently, hiPSC lines have been generated from a wide range of other somatic cell types, such as keratinocytes[93] and blood[94]. Furthermore, a variety of reprogramming methods have been developed to generate hiPSC lines. Initially, integrating retroviruses or lentiviruses were used to deliver the reprogramming factors[91,92]. More recently, alternative and safer methods have been developed to introduce reprogramming factors into host cells, such as excisable transposons[95], episomal vectors[96], and non-integrating DNA-free Sendai virus[97].

2. Signaling Pathways in Human Embryonic Stem Cells

The signaling mechanisms that regulate the pluripotency of hESCs are distinct from those of mouse ESCs; human LIF is not sufficient to maintain hESCs in a pluripotent, undifferentiated state[98]. Interestingly, while BMPs maintain mouse ESCs in a pluripotent state, BMP4 provides hESCs with an instructive signal for differentiation to trophoblast[99]. In addition, short-term treatment of hESCs with BMP4 promotes early mesoderm induction and these mesoderm progenitors can further differentiate into the hematopoietic lineage[100]. Consistent with these findings, suppression of BMP4 signaling by its antagonist Noggin sustains hESCs in an undifferentiated state[101]. Furthermore, the TGF-β ligand GDF3 contributes to the upregulation of pluripotency genes by antagonizing BMP4 in hESCs[102]. Together, these results indicate that BMP4 induces differentiation of hESCs, which is in contrast to its role in mouse ESCs.

Activation of another branch of the TGF-β signaling pathway, TGF-β/activin/nodal, plays a crucial role in maintaining the undifferentiated state of hESCs through SMAD2/3[103]. Medium enriched with ACTIVINA can support the undifferentiated state of hESCs in the absence of

feeder cells[104], while inhibition of either ACTIVIN or NODAL signaling by FOLLISTATIN and LEFTY, respectively, promotes differentiation in hESCs[105]. Collectively, these data indicate that TGF-β signaling pathway plays multiple roles in regulating the pluripotency of hESCs.

Human ESCs have traditionally been cultured in the presence of basic fibroblast growth factor (bFGF/FGF2)[86,106,107]. The dependence of hESC pluripotency on FGF2 indicates the importance of FGF signaling in hESC self-renewal. FGF2 signaling contributes to the maintenance of hESC pluripotency through the FGF/MAPK pathway. Exogenous FGF2 signaling purportedly activates MAPK pathway to enhance cell survival and cloning efficiency, while endogenous FGF2 suppresses differentiation since knock-down of endogenous FGF2 promotes differentiation[108]. However, the molecular mechanisms of how endogenous FGF2 inhibits hESC differentiation remain unclear.

Although human and mouse pluripotent stem cells have distinct signaling requirements, biological consequences of these differences have yet to be elucidated. It might simply reflect the differences between species or rather imply that these two cell types do not share the same degree of developmental potential. It has been shown that mouse pluripotent stem cells can access a naïve pluripotent state, where they are functionally equivalent to the preimplantation epiblast; "tetraploid complementation" has served as the most stringent test of pluripotency for mouse pluripotent stem cells. However, unlike mouse pluripotent stem cells, universal standards to evaluate the pluripotent state of human pluripotent stem cells are not available. The possibility that human pluripotent stem cells can acquire a naïve pluripotent state is an active field of research and will be discussed below.

3. Naïve Human Pluripotent Stem Cells

Recently, several studies reported that transgenic modifications enable hESCs/hiPSCs to acquire and maintain naïve pluripotency[109-111]. Although these naïve human cells share

transcriptional, epigenetic and functional features with their mouse counterparts, upon transgene withdrawal they reverted to a primed state. More recently, several groups demonstrated that various combinations of small molecules could force hESCs/hiPSCs into a naïve-like pluripotent state[112-116]. These putative naïve human pluripotent stem cells share similarities at the transcriptional level with human preimplantation epiblast cells and mouse naïve ESCs, which is characterized by the robust expression of naive markers including *NANOG, KLF4*, and *TFCP2L1*. However, naïve transcriptional networks are not well conserved between human and mouse pluripotent stem cells and are more similar to their respective blastocysts[117]. Furthermore, at present, universal tests for human naïve pluripotency are lacking, and thus it is not plausible to benchmark hESCs against *bona fide* naïve pluripotent state.

Whereas mouse pluripotent stem cells are useful tools for studying mammalian development and disease modeling, human pluripotent stem cells are valuable sources not only for research but also for regenerative medicine. In particular, human iPSCs hold great potential for regenerative medicine since it is possible to generate patient-specific cell lines directly from patients' somatic cells with relative ease, which circumvents the ethical and technical issues posed by hESCs. However, in order to realize their full potential, it is imperative to evaluate if hiPSCs are equivalent to hESCs that have been well characterized and thus serve as the gold standard of pluripotent stem cells.

V. Molecular Similarities and Differences Between Human Pluripotent Stem Cells

1. Transcriptional Patterns of Human Pluripotent Stem Cells

Recently, it was shown that isogenic mouse iPSCs and ESCs are transcriptionally indistinguishable with the exception of a few transcripts within the imprinted *Dlk1-Dio3* gene cluster on chromosome 12qF1[118]. It remains unclear whether hiPSCs and hESCs are equivalent at

the transcriptional level. Initially, the comparison of transcriptional profiles of a cohort of hiPSCs and hESCs revealed that 4,356 genes were differentially expressed between hESCs and hiPSCs[119]. This result implied that there might be fundamental differences at the transcriptional level between hESCs and hiPSCs. Surprisingly, however, a different study failed to find consistent differences in gene expression upon extended culture between hESCs and hiPSCs. Instead, only four genes were reported to be differentially expressed between hESCs and hiPSCs, suggesting that hESCs and hiPSCs are nearly identical[120]. Moreover, a hierarchical clustering analysis showed that transcriptional patterns of hESCs and hiPSCs were grouped based on the lab of origin rather than cell types[121]. Thus, there are no consistent conclusions about whether hESCs and hiPSCs are transcriptionally equivalent.

More importantly, even though some papers reported statistically significant differences, it is not clear if these are biologically relevant. The magnitude of reported transcriptional changes was often very small (1.5-3 fold), which may reflect genetic variation or more subtle influences of the culture environment rather than intrinsic cell type-specific differences. Consistent with this notion, a recent study suggested that although transcriptional differences could be identified between hESCs and hiPSCs, statistically significant expression differences might not lead to functional consequences[122].

Factors that may confound previously published hESC/hiPSC comparisons include the use of (i) hESCs and hiPSCs with different genetic backgrounds, (ii) different reprogramming methods, and (iii) different culture systems. To determine whether hESCs and hiPSCs are transcriptionally equivalent, it is therefore critical to eliminate these sources of variability.

2. DNA Methylation Patterns of Human Pluripotent Stem Cells

Given that the reprogramming of somatic cells into hiPSCs involves an epigenetic remodeling process, incomplete or erroneous reprogramming could result in aberrant epigenetic patterns. Aberrant DNA methylation patterns in hiPSCs are one of the epigenetic abnormalities

arising during reprogramming[123]. A recent study reported that differentially methylated regions (DMRs) were identified in hiPSCs compared to parental somatic cells. Importantly, these DNA methylation patterns differed from those of the parent somatic cells and from those of human ES cells, which might reflect aberrant epigenetic reprogramming. Interestingly, incomplete reprogramming could also result in a latent epigenetic fingerprint from the parental somatic cells, which manifests itself at the transcriptional and functional level[124]. In addition, reprogramming methods and culture conditions can affect the developmental potential of iPSC lines, as seen in mouse studies [90,125-127]. However, it remains to be seen whether this is generalizable to all human reprogramming conditions[128,129].

Recently, a reprogramming-specific epigenetic signature was identified, which consists of nine aberrantly methylated genes in hiPSCs as compared to hESC[130]. The signature could faithfully segregate hESC and hiPSC lines. Intriguingly, another group reported that the methylation profiles from 82 CpG sites associated with 66 genes could separate hiPSCs from hESCs[131]. However, no overlap between these two studies was identified. More importantly, the biological relevance of these differentially methylated genes remain unclear.

To explore the epigenetic and functional variability between hESCs and hiPSCs, Bock et al.[88] profiled 20 well-characterized hESC and 12 hiPSC lines for gene expression, DNA methylation, and differentiation propensities, which resulted in a comprehensive "reference map" of these three categories. By comparing each hESC or hiPSC line to the "reference map", the authors confirmed that although cell-line-specific outliers were slightly more prevalent among hiPSC lines than among hESC lines, they could not find any transcriptional or epigenetic deviation that was unique to all hiPSC lines. Consistent with these data, large-scale analyses of human pluripotent stem cells revealed that hESCs and hiPSCs were not separated based on microarray and global DNA methylation assay and that a subset of hiPSC clones rather than all

hiPSCs exhibited aberrant gene expression and defective differentiation potential as compared to hESCs[128].

Collectively, these results suggest that with current technology, it might not be possible to generate hiPSC lines that are completely equivalent to the ideal hESC lines[87]. Moreover, subtle abnormalities in iPSC lines could result in a major functional defect[118]. Since the standards for assessing the quality of hiPSCs are still controversial, it is critical to carefully and precisely evaluate the genetic, epigenetic and functional status of hiPSCs for the future application of these cells.

Chapter 2

Global transcriptional and epigenetic patterns in ESCs and EGCs are driven by sex rather than cell type

Attributions

I generated all ESC and EGC lines used in this thesis except for the *Dnmt1/3a/3b* triple knockout ESC line that was obtained from Alex Meissner's lab. I designed all assays and performed AP staining, immunostaining, Southern blot and Western blot assays. Jean-Pierre Etchegaray and I performed Dot blot assays. I performed gene targeting with technical assistance from Aaron Huebner. I also performed flow cytometry analysis (DNA content analysis) with technical assistance from FACS facility at MGH. For the RNA-sequencing, I collected samples and isolated RNA, and Tufts Genomics Core conducted the sequencing. Peter Park and Soohyun Lee performed bioinformatics analysis of RNA-sequencing data. For the RRBS analysis, I collected samples and isolated genomic DNA, and Patrick Boyle performed the sequencing. Alex Meissner, Jamie Webster, and Kendell Clement performed bioinformatics analysis of RRBS data. For bisulfite-sequencing, I collected samples and isolated genomic DNA, and EpigeneDx completed the sequencing.

Abstract

Blastocyst-derived embryonic stem cells (ESCs) and genital ridge-derived embryonic germ cells (EGCs) represent two classic types of pluripotent stem cell lines. Despite remarkable similarities between ESCs and EGCs, the epigenetic equivalence of these cell types remains debated. For example, previous studies suggested that EGCs are globally hypomethylated relative to ESCs, mirroring epigenetic patterns in the germline. However, subsequent reports failed to identify differences between ESCs and EGCs. Here, we compared global DNA methylation and transcriptional signatures between isogenic ESC and EGC lines with the goal to define similarities and differences and to identify possible regulators underlying the purported DNA hypomethylation of EGCs. Surprisingly, we found that global DNA methylation patterns were indistinguishable between ESC and EGC lines of the same sex, with female cell lines showing global hypomethylation compared to male cell lines. Cell fusion experiments between EGCs or ESCs and somatic cells further showed that the ratio of X chromosomes to autosomes determines methylation levels in pluripotent cells. Mechanistically, we discovered a link between elevated expression of the X-linked gene, *dual specificity phosphatase* 9 (*Dusp9*), attenuated MAP kinase signaling and reduced protein levels of Dnmt3a and Dnmt3b in female pluripotent cell lines. Together, our data demonstrate that sex rather than cell type of origin of EGC and ESC lines dictates global epigenetic and transcriptional patterns in mouse pluripotent cell lines maintained in conventional culture conditions.

Introduction

Pluripotent stem cell lines have the ability to indefinitely self-renew in culture while retaining the capacity to differentiate into derivatives of all three germ layers, thus providing a

useful tool in basic research and cell therapy[132]. They are typically derived from preimplantation embryos, yielding embryonic stem cells (ESCs), or from fetal germ cells, generating embryonic germ cells (EGCs)[3,6,11,12]. Moreover, pluripotent stem cells can be produced from somatic cells upon forced expression of transcription factors, giving rise to induced pluripotent stem cells (iPSCs)[90]. Whereas ESCs, EGCs and iPSCs are thought to be highly similar at the molecular and functional levels[118], EGCs reportedly exhibit epigenetic differences that may reflect their origin from fetal germ cells[133]. For example, EGCs derived from E11.5-E12.5 genital ridges show global DNA hypomethylation and loss of genomic imprints, which mirrors epigenetic changes occurring in fetal germ cells *in vivo*[13,133]. Furthermore, classic cell-cell fusion experiments between EGCs and somatic cells demonstrated loss of autosomal, repetitive and imprint-specific methylation patterns in hybrids, suggesting that EGCs harbor a dominant demethylation activity[133]. Importantly, ESCs lack this activity when fused to somatic cells[134]. A recent study confirmed these earlier findings and suggested that Tet1 enzyme played a role in EGC-specific demethylation of imprints in hybrid clones, although global methylation patterns were not examined in that study[135]. Moreover, a systematic comparison of DNA methylation patterns between parental ESCs and EGCs and resultant hybrids has not yet been performed, precluding attempts to further dissect the mechanisms underlying this putative demethylation process.

Accumulating evidence suggests that culture conditions can dramatically alter DNA methylation patterns in pluripotent cell lines[78-80]. While ESC lines maintained in conventional culture media ("serum/LIF") show high methylation levels that globally resemble somatic tissues, exposure to MAPK and GSK3 inhibitors in the presence of LIF ("2i/LIF") triggers a roughly four-fold reduction of global methylation patterns and the concomitant acquisition of a more naïve pluripotent state[36,78,79]. Notably, imprinted gene methylation remains unaffected by 2i/LIF culture, suggesting a gene-specific protection mechanism[36,79]. 2i culture-induced demethylation involves direct transcriptional silencing of the *de novo* methyltransferases Dnmt3a and Dnmt3b as well as transcriptional upregulation of the dioxygenases Tet1 and Tet2 by the pluripotency

regulators Prdm14 and Nanog, resulting in both passive and active demethylation of the genome within a few passages[37,79,80,136].

In addition to culture conditions, the sex of ESC lines may influence methylation patterns[83,84]. Female mouse ESC lines cultured in serum/LIF exhibit a reduction of global methylation levels and a decrease in imprinted gene methylation when compared to male ESCs. Of note, reports in the literature as to the degree of female ESC hypomethylation relative to male vary significantly, anywhere from ~10% to ~80%[80,84]. Hypomethylation in XX ESCs was directly linked to the presence of two activate X chromosomes and reduced expression levels for Dnmt3a/b[84] or Dnmt3l[83] since XO ESCs and XX somatic cells regained Dnmt3a/b expression and global methylation patterns[83,84]. More recently, Schulz et al. (2014) discovered that female ESCs grown in serum/LIF exhibit reduced MAPK signaling, elevated *Prdm14*, *Nanog* and *Tet1/2* transcription and reduced *Dnmt3a/b* mRNA levels relative to male ESCs, indicating that the presence of two active X chromosomes recapitulate certain phenotypes of ESCs maintained in a naïve pluripotent state. It remains unclear, however, whether MAPK inhibition is sufficient to trigger global hypomethylation towards a female ESC-like state, and no female-specific suppressors of the MAPK pathway were identified in that study. Moreover, the impact of sex and culture conditions on DNA methylation and gene expression patterns in late-stage (E11.5-E12.5) EGCs, relative to ESCs, remains unexplored.

Here, we revisit the previous observation that EGCs exhibit widespread DNA hypomethylation and dominant demethylation activity over ESCs with the goal of identifying possible mechanisms and mediators underlying this phenotype. Most of the previous comparisons between ESCs and EGC lines were limited by a small number of cell lines[137], developmental stages of germ cells with variable degree of epigenetic reprogramming (E8.5)[36], differences in genetic background[138], sex[133,139] or culture conditions[80]. We therefore derived multiple isogenic E3.5 ESC and E11.5-E12.5 EGC lines under equivalent growth conditions in order to compare

and contrast (i) global transcription, (ii) genome-wide DNA methylation, and (iii) the ability to induce DNA demethylation in fusion hybrids.

Results

Generation of isogenic ESCs and EGCs

We crossed female 129S6 mice to male C57B6/6J-Tg(pPGKneobpA)3Ems/J mice to derive isogenic, F1 ESC and EGC lines. Specifically, ESC lines were established from E3.5 blastocysts whereas EGC lines were produced from E11.5 and E12.5 genital ridges (Figure 2.1). We derived a total of 6 male and 4 female EGC lines as well as 3 male and 3 female ESC lines (Table 2.1). All cell lines were derived in 2i/LIF and maintained in serum/LIF culture conditions on feeders. This modification allowed us to use equivalent culture conditions for ESCs and EGCs, which is not feasible with classic derivation protocols.

ESC and EGC lines could be propagated over multiple passages while maintaining pluripotency-associated markers such as alkaline phosphatase (AP) activity and endogenous Nanog and Sox2 expression, documenting their self-renewal potential (Figure 2.2). Critically, EGC lines but not ESC lines exhibited demethylation of the *Peg3* locus, indicating successful erasure of genomic imprints, which is a hallmark of late-stage germ cells and derivative EGCs (Figure 2.3A). Together, these results show that our newly derived ESC and EGC lines exhibit the expected growth behavior, pluripotency marker profiles, and imprinting patterns.

Global methylation analysis distinguishes ESCs and EGCs by sex rather than cell type

To assess global methylation patterns in ESC and EGC lines, we first performed Southern blot analysis for repetitive elements and dot blot analysis for 5mC. In contrast to some previous suggestions, EGC lines were not uniformly hypomethylated compared to ESC lines. Instead, we

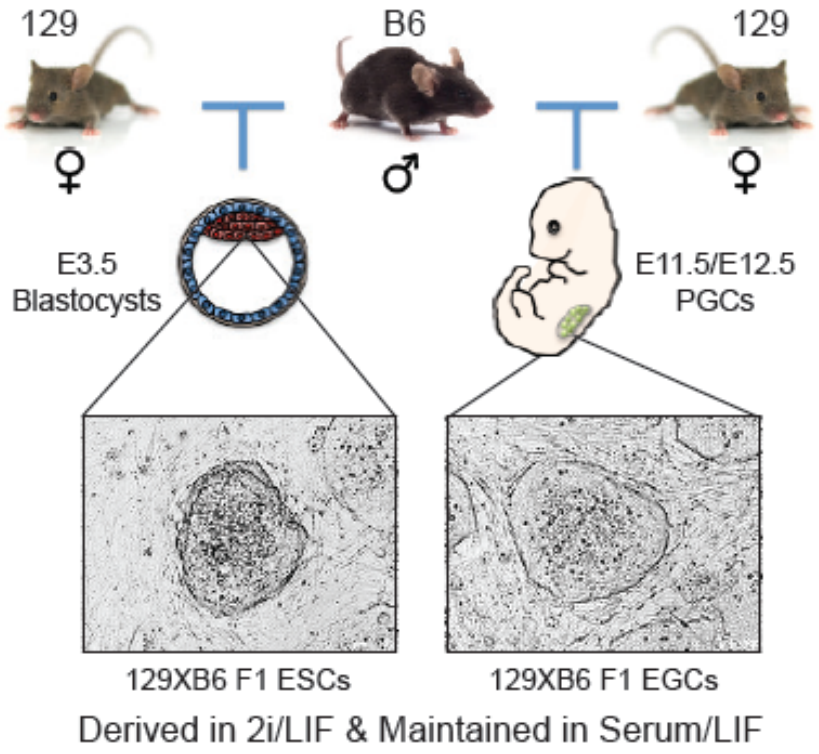


Figure 2.1. Schematic for the generation of genetically matched ESCs and EGCs.

Table 2.1. Detailed information on the established ESCs and EGCs.
ESCs and EGCs were collected between passages 4-6 for all experiments unless otherwise specified.

Cell Type	Developmental Origin	# XY Lines	# XX Lines
ESC	E3.5 Blastocysts	3	3
EGC	E11.5 PGCs	3	3
EGC	E12.5 PGCs	3	1

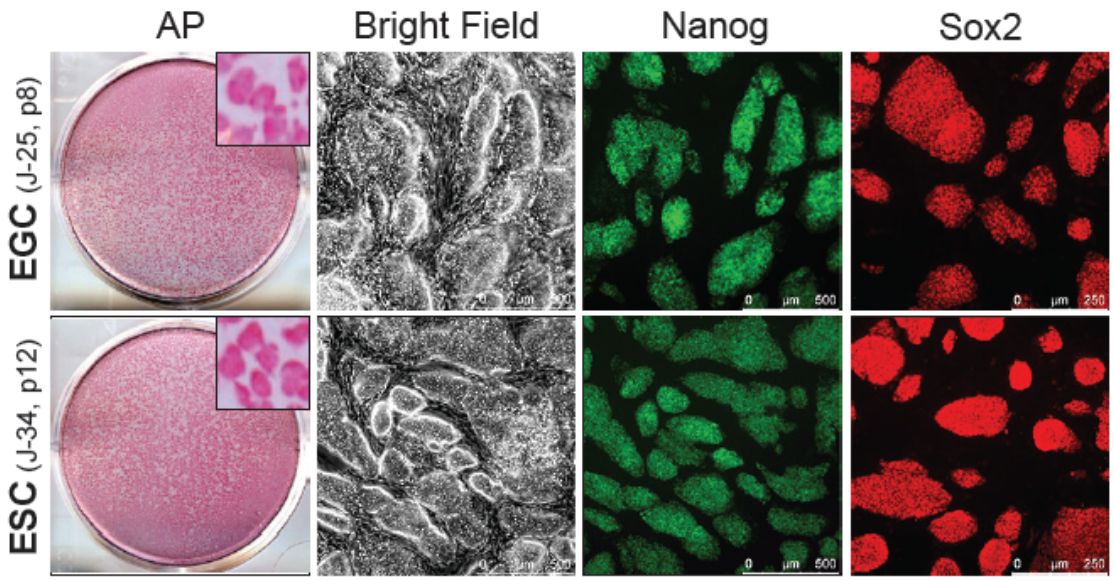


Figure 2.2. Characterization of isogenic ESC and EGC lines.
EGCs (top panels) and ESCs (bottom panels) express alkaline phosphatase (AP), NANOG and SOX2. Insets show magnification of AP^+ colonies.

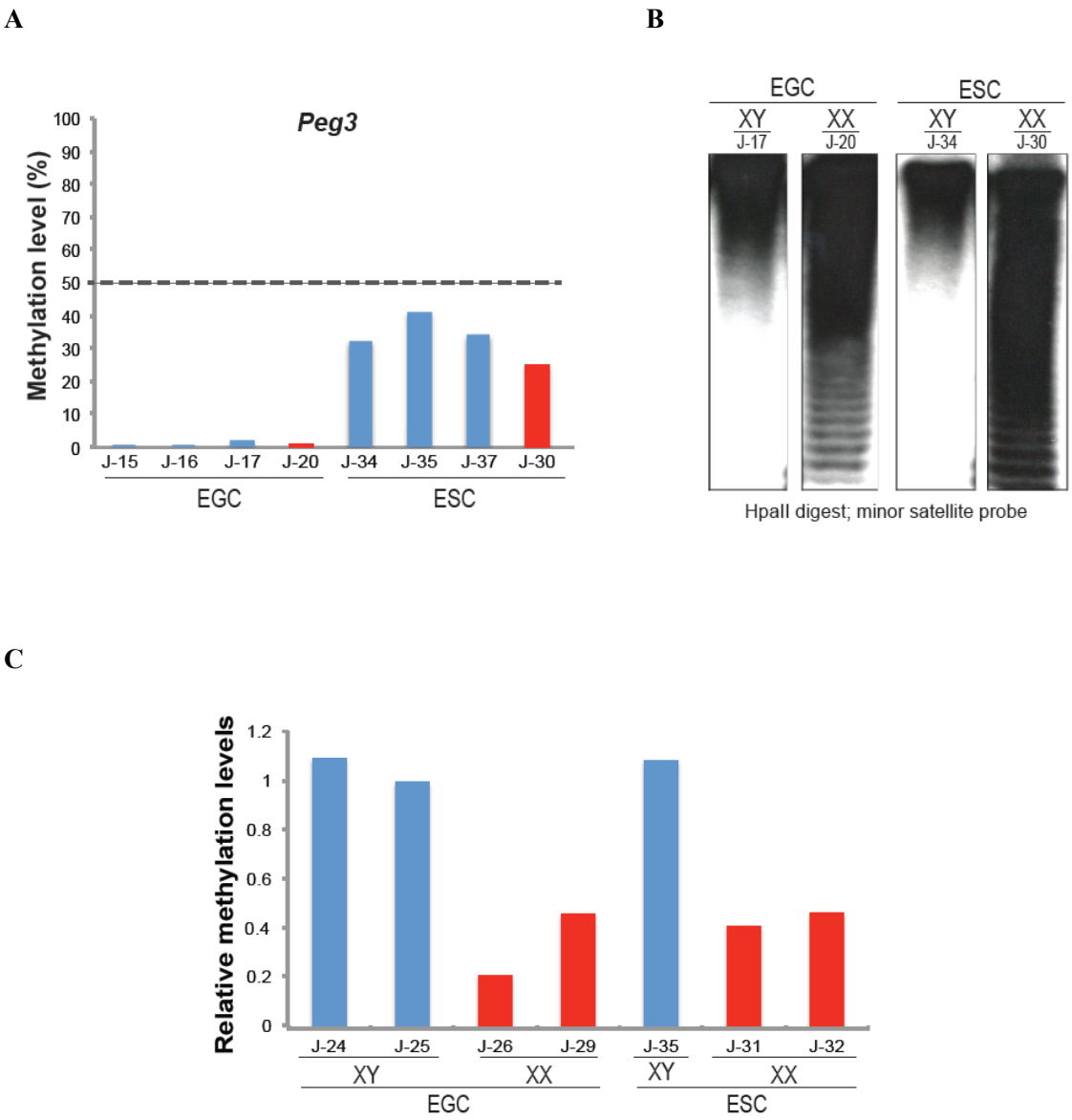


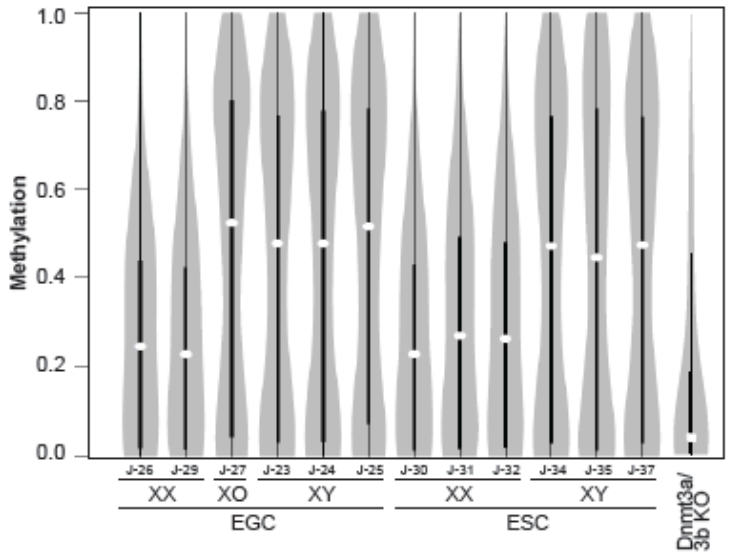
Figure 2.3. Global DNA methylation levels of isogenic ESC and EGC lines.
(A) Bisulfite methylation sequencing of the imprinted *Peg3* locus in isogenic ESC and EGC lines. Blue bars represent XY cells, red bars represent XX cells. **(B)** Southern blot analysis for minor satellite repeats using a pMR150-specific probe in ESCs and EGCs. HpaII digests unmethylated DNA; high molecular weight bands represent DNA hypermethylation while lower molecular weight bands indicate DNA hypomethylation. **(C)** Global methylation levels of isogenic ESC and EGC lines were measured using dot blot analysis for 5mC. Blue bars represent XY cells, red bars represent XX cells.

noticed that male ESC and EGC lines were equally hypermethylated while female ESC and EGC lines were equally hypomethylated, suggesting that sex rather than cell type of origin is the predominant driver behind overall DNA methylation levels in pluripotent cell lines (Figure 2.3B and C). To corroborate this result, we analyzed all cell lines by reduced representation bisulfite sequencing (RRBS), which measures methylation patterns at single-base resolution across the genome[77]. RRBS analysis confirmed these results, demonstrating that male ESC/EGC lines are globally hypermethylated relative to female ESC/EGC lines (Figure 2.4A). Accordingly, unsupervised clustering separated all cell lines based on sex rather than cell type with one exception; the EGC line J-27 clustered with male ESC and EGC lines despite the absence of a Y chromosome-specific PCR signal (Figure 2.4B and data not shown). This cell line must have originated from either female EGCs that lost one of the X chromosomes or from male EGCs that lost the Y chromosome. Regardless of the origin, the "XO" genotype is expected to be globally hypermethylated[84], thus explaining our observations. We therefore conclude, based on examining global DNA methylation patterns by three independent assays, that genetically matched ESC and EGC lines maintained in serum/LIF are distinguished by sex rather than cell type of origin.

Hypomethylation in XX ESCs and EGCs occurs evenly across most genomic features

Although total DNA methylation levels were equivalent between ESCs and EGCs of the same sex, local distribution of methylation patterns could differ in the two cell types, as exemplified by demethylation of the *Peg3* locus that was observed exclusively in EGCs (Figure 2.3A). We therefore compared methylation patterns by RRBS across several genomic features for both autosomes and the X chromosomes in ESCs and EGCs (Figure 2.5A and 2.5B). This analysis failed to detect cell type-specific methylation differences when comparing 1 kb tiles. Likewise, examination of promoter regions (CpG islands and shores) and repetitive elements (IAPs, LINEs, SINEs and LTRs) showed that female ESC and EGC lines were similarly

A



B

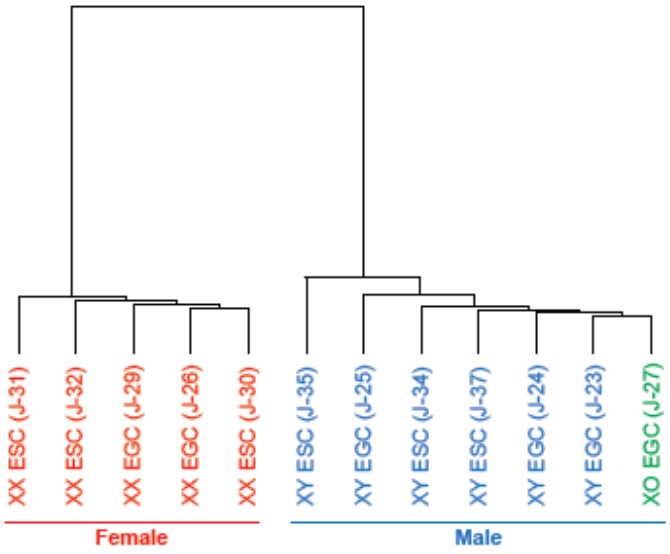


Figure 2.4. Global DNA methylation reflects sex rather than cell type.
(A) Global methylation analysis of isogenic ESC and EGC lines by reduced representation bisulfite sequencing (RRBS) using violin plot representation. High-passage *Dnmt3a/3b* double knock-out (DKO) ESCs were included as controls. **(B)** Unsupervised hierarchical clustering of samples shown in Fig. 2.4A. Blue bars represent XY cells, red bars represent XX cells; green bars represent XO cell line.

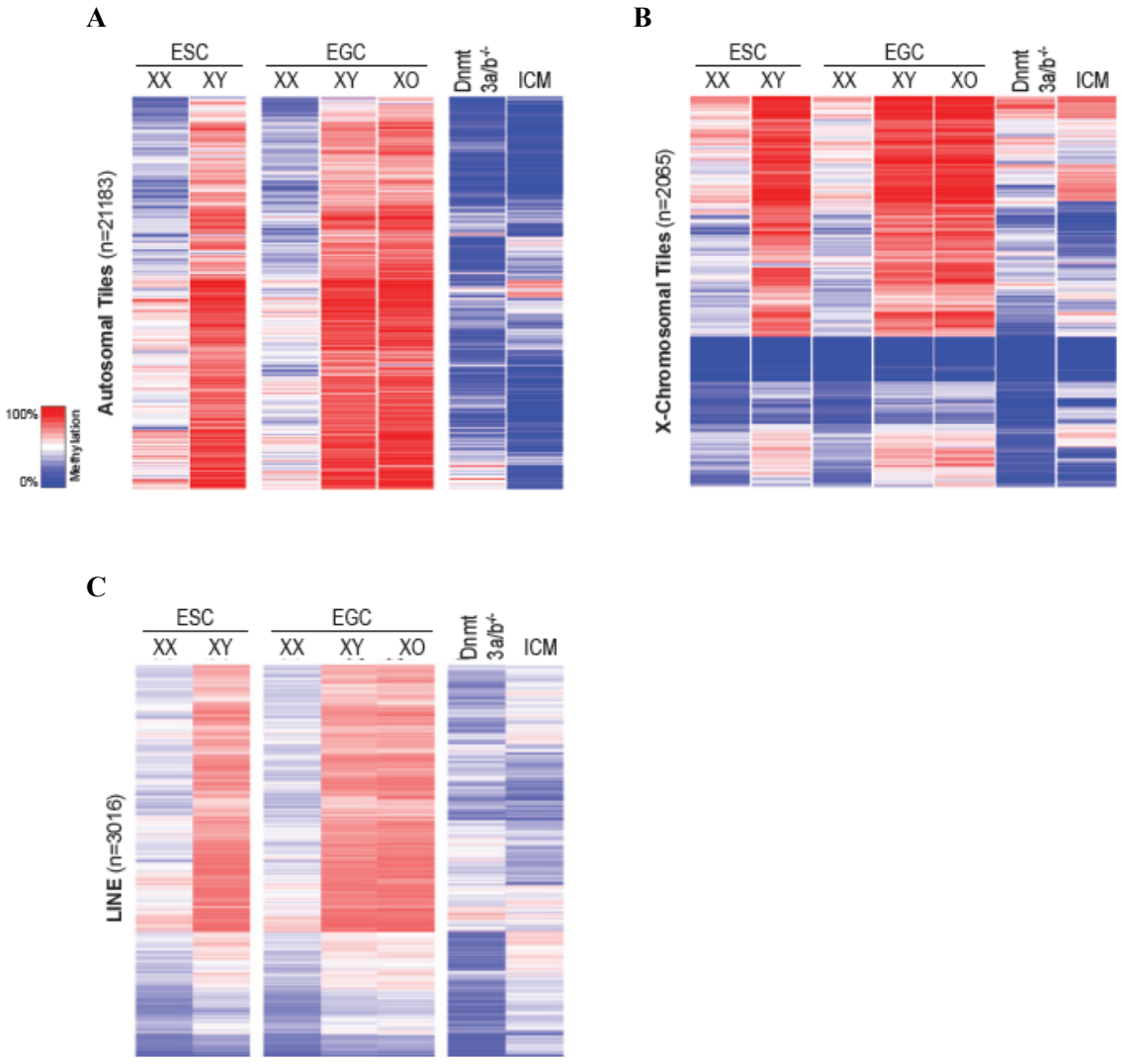


Figure 2.5. Methylation patterns are equivalent across most genomic features in sex-matched ESCs and EGCs.

(A-C) Comparison of DNA methylation patterns by RRBS across autosomes (A), the X-chromosome (B) and LINE element (C) in isogenic ESC and EGC lines, *Dnmt3a/b* double knockout ESCs and inner cell mass (ICM) cells.

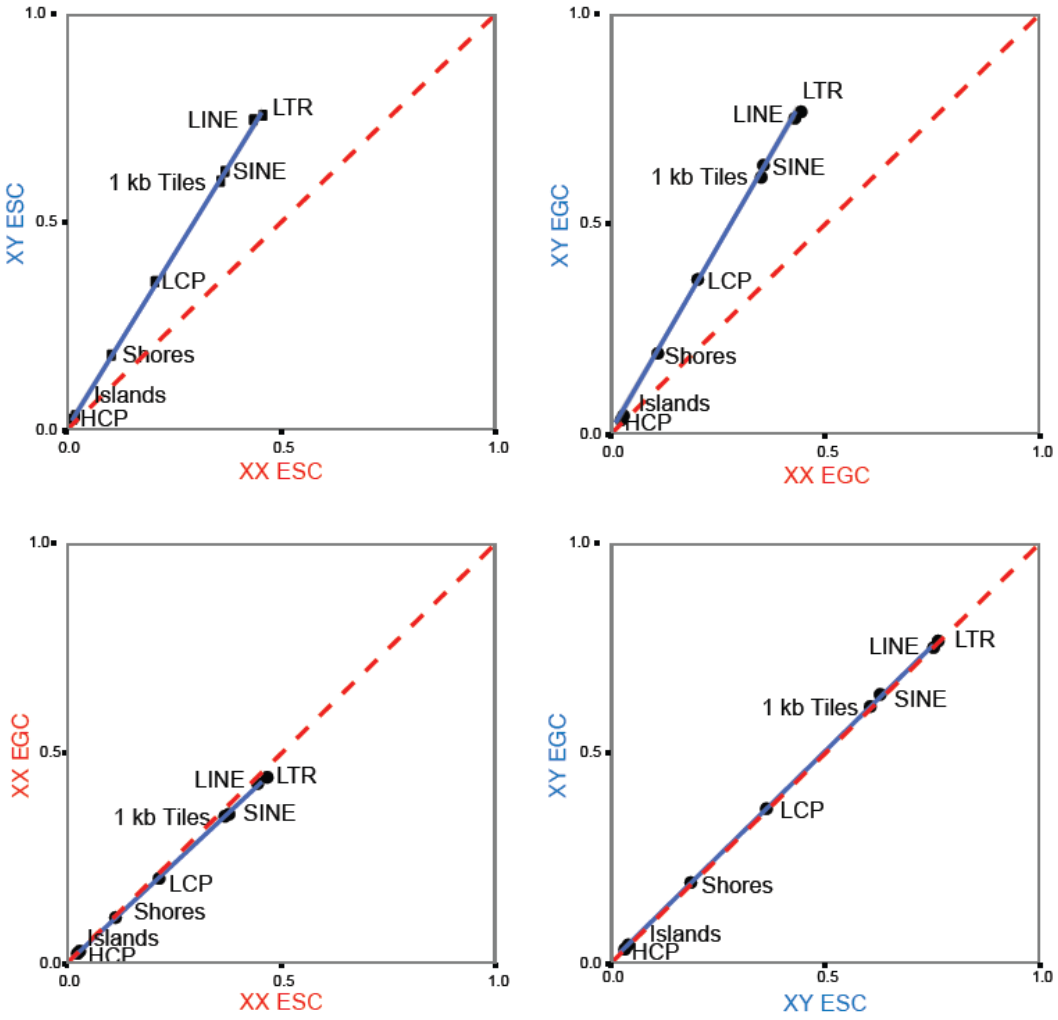


Figure 2.6. Correlation plots comparing methylation levels of indicated genetic elements between pairs of cells lines.

hypomethylated whereas male ESC and EGC lines were equally hypermethylated at these sites (Figure 2.5C, 2.6 and 2.7).

In contrast to methylation patterns across promoters and repetitive elements, CpGs associated with imprinting control regions (ICRs) showed notable cell type and sex-dependent differences. For example, ICR methylation was generally absent or extremely low in EGC lines compared to male ESC lines, as expected (e.g. Peg13, U2af1) (Figure 2.8A and B). However, female ESCs also exhibited EGC-like hypomethylation at many ICR elements (e.g. Grb10, Impact, Nnat), pointing out similarities between EGC lines and female ESC lines[80] (Figure 2.8A and B). Accordingly, unsupervised clustering of samples based on all captured imprinted genes grouped female ESCs with EGCs but separately from male ESCs. (Figure 2.8A)

Of interest, global DNA methylation patterns of female stem cell lines were more similar to those of Dnmt3a/3b double knock-out (DKO) ESC lines and inner cell mass (ICM) cells than to male stem cell methylation patterns (Figure 2.4A and 2.5A-C). This finding is in agreement with the recent suggestion that female ESCs are captured in a more naïve pluripotent state compared to male ESCs[85]. Collectively, these results show that, with the exception of a few imprinted genes such as Peg13 (Figure 2.8A and B), hypomethylation in female ESCs and female EGCs occurs evenly across the genome and resembles global levels observed in the ICM and Dnmt3a/b DKO cells.

Methylation patterns in ESC/EGC-somatic cell hybrids are driven by the ratio of X chromosomes to autosomes

In light of our finding that male and female pluripotent stem cell lines exhibit profound differences in DNA methylation patterns irrespective of cell type, we next interrogated the role of sex chromosomes in ESC-somatic and EGC-somatic hybrids (Figure 2.9A). Briefly, neomycin-resistant ESC and EGC lines of both sexes were fused to puromycin-resistant male or female

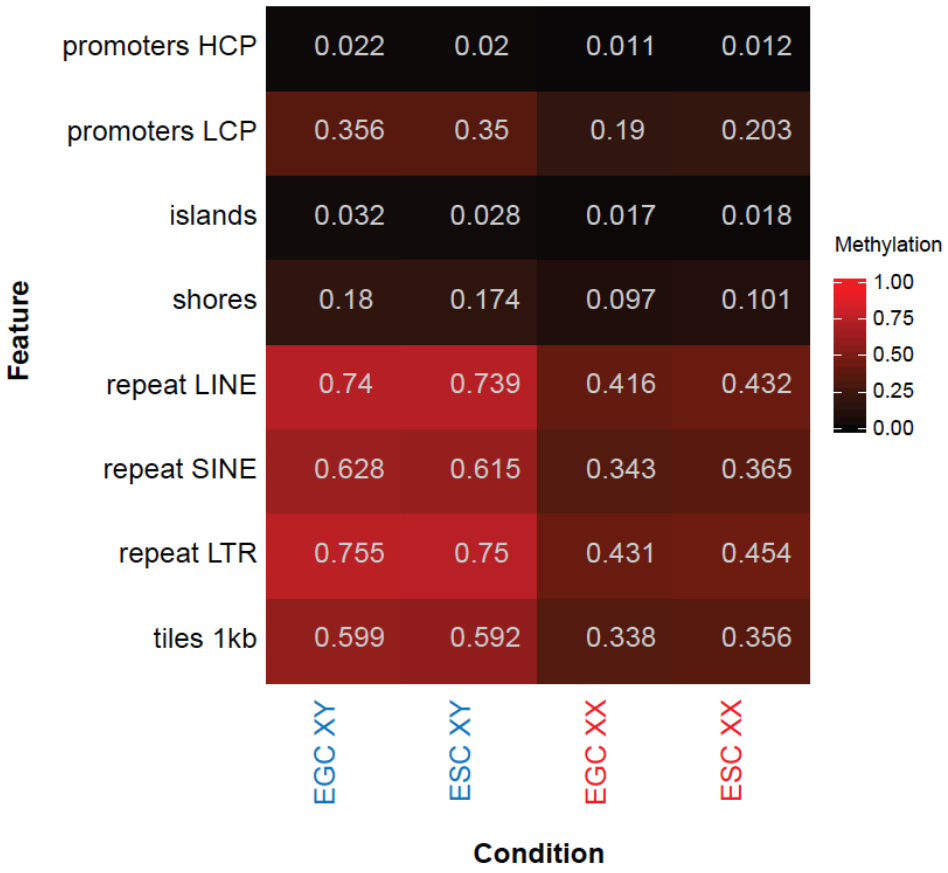
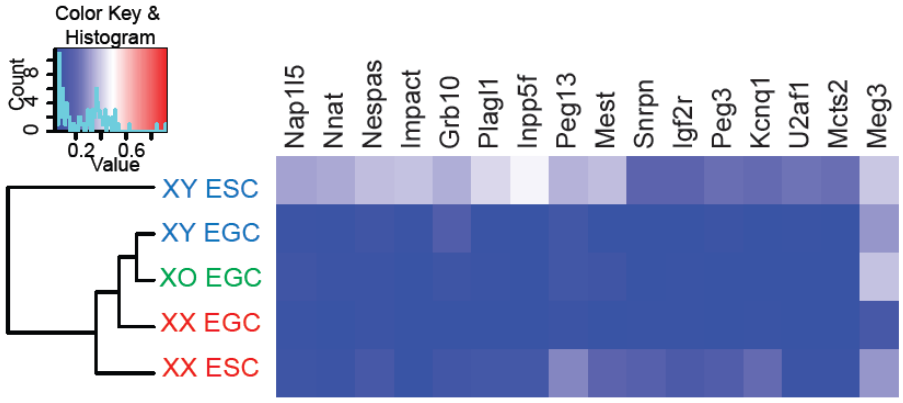


Figure 2.7. Distribution of DNA methylation patterns in female stem cells across the genome.
RRBS was used to analyze DNA methylation levels for various genetic elements from isogenic X and XY ESC/EGC lines. The average values of the representative methylation levels for 3 XY ESC, 3 XX ESC, 3 XY EGC, and 2 XX EGC lines were shown, respectively.

A



B

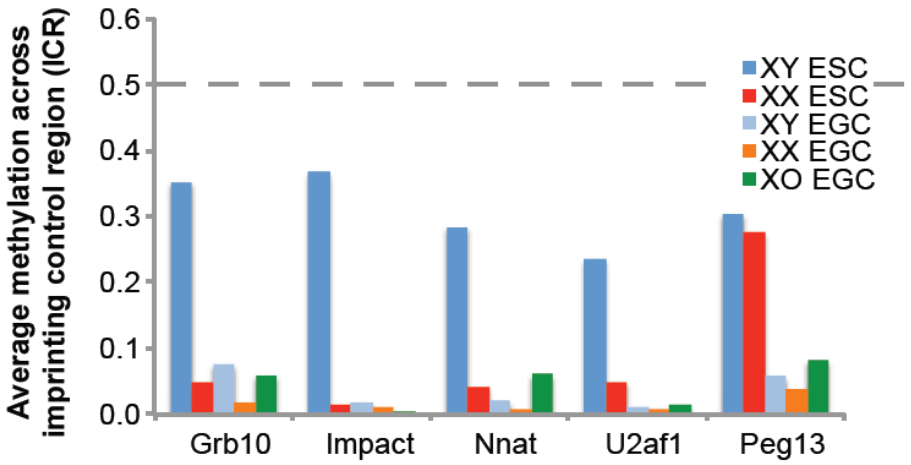


Figure 2.8. DNA methylation patterns of imprinting control regions (ICR) at imprinted loci in isogenic ESCs and EGCs.
(A) Dendgrogram and heatmap representation of imprinted loci in isogenic ESC and EGC lines. **(B)** Average methylation levels across ICRs at indicated imprinted loci. The average values of the representative methylation levels for 3 XY ESC, 3 XX ESC, 3 XY EGC, and 2 XX EGC lines were shown, respectively.

MEFs carrying an EGFP reporter in the endogenous Pou5f1 locus (termed *Oct4-GFP*) using polyethyleneglycol (PEG), followed by dual selection for G418 and puromycin (Figure 2.9A). We confirmed that hybrid clones reactivated the somatically silenced *Oct4-GFP* reporter and carried a tetraploid genome before measuring DNA methylation patterns by dot blot or RRBS analysis (Figure 2.9B, 2.9C, and 2.10A).

Several important conclusions can be drawn from these analyses. First, overall methylation levels of (4n) hybrids resembled those of the parental male or female (2n) pluripotent stem cell line, although the (2n) somatic fusion partners of either sex were consistently hypermethylated (Figure 2.11A and 2.10B). This result indicates that the female somatic genome in XXXX hybrids undergoes massive demethylation towards a female ESC/EGC-like state while the male somatic genome remains more methylated in XXYY hybrids (Figure 2.11A and 2.10B). This finding further implies that the ratio of active X chromosomes to autosomes rather than the absolute number of X chromosomes inside a pluripotent cell dictate DNA methylation levels. In agreement with this notion, XXXY hybrids carrying three active X chromosomes per tetraploid genome exhibit intermediate methylation levels between those of male (XXYY) and female (XXXX) hybrids (Figure 2.11B). Pluripotency loci such as Oct4 were equally demethylated in male and female hybrids, thus excluding the possibility that the observed sex-specific methylation differences were a consequence of incomplete reprogramming of the somatic genome (Figure 2.10A).

Second, global methylation levels of ESC hybrids were indistinguishable from those of EGC hybrids of the same sex using unsupervised clustering (Figure 2.11C). Examination of methylation patterns at promoter regions, LTRs, LINEs and SINEs in ESC and EGC hybrids corroborated this conclusion (Figure 2.10C). Thus, sex rather than cell type determines global methylation levels in both ESC-somatic and EGC-somatic hybrids.

Third, imprinted gene methylation was again the only discernible difference between ESC and EGC hybrids (Figure 2.12A-C). Whereas male ESC hybrids showed similar methylation

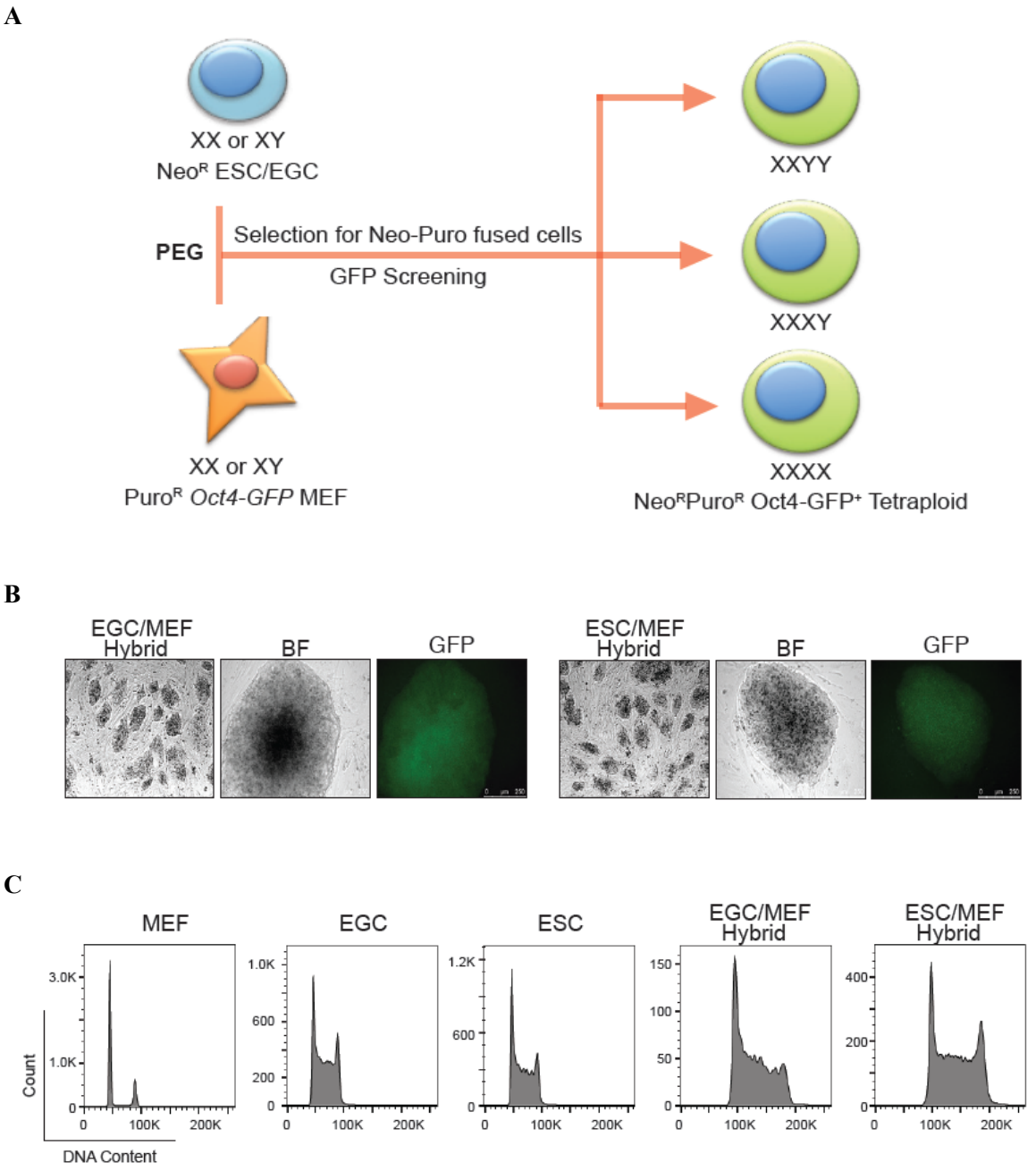


Figure 2.9. Generation of ESC/EGC-somatic fusion hybrids.
(A) Schematic for the generation of ESC/EGC-somatic fusion hybrids **(B)** Representative EGC-MEF hybrid and ESC-MEF hybrid show activation of somatically silenced *Oct4-EGFP* reporter. **(C)** Representative EGC-MEF and ESC-MEF hybrids show tetraploid DNA content compared to parental fusion partners by propidium iodide (PI) staining.

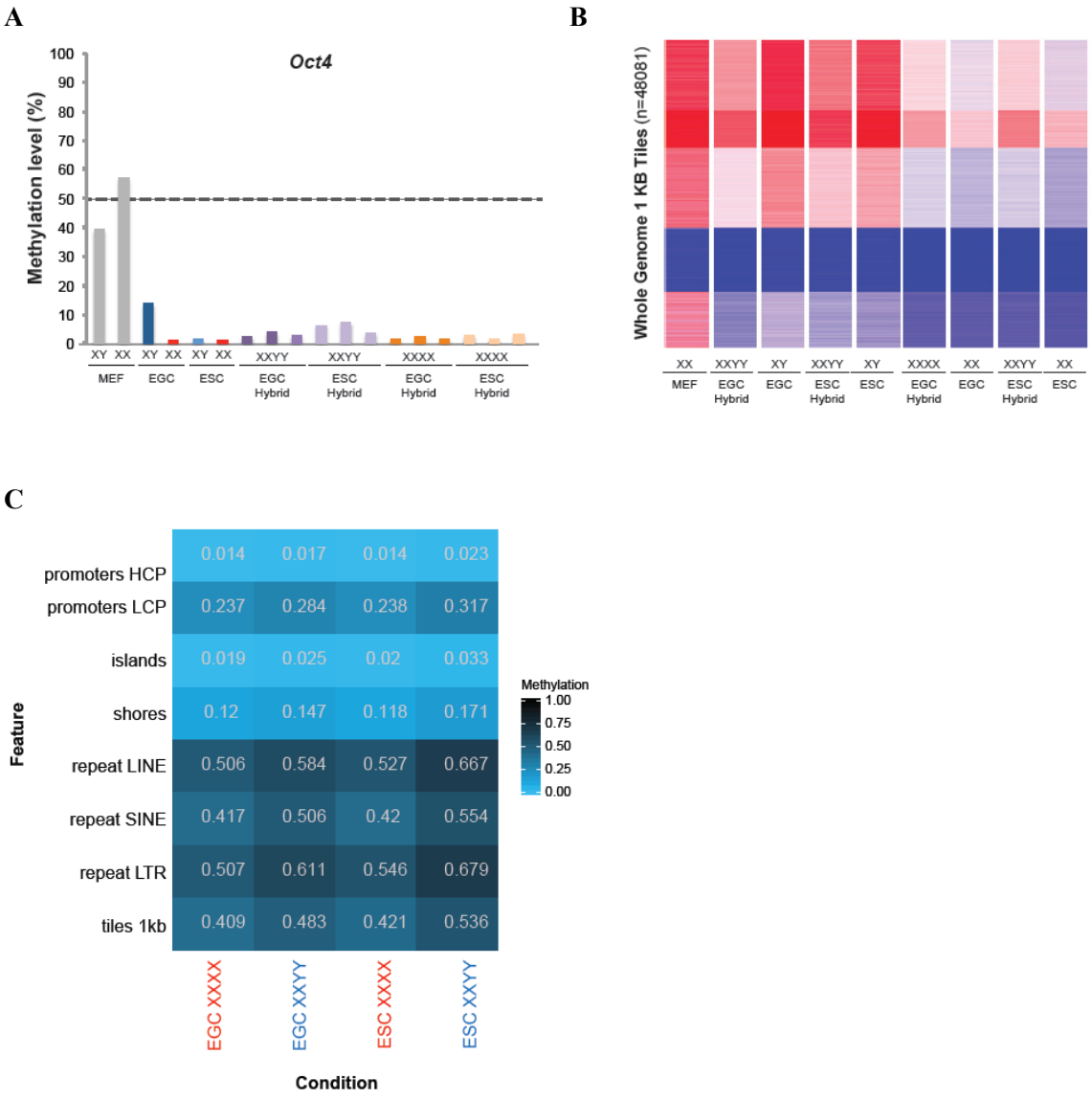


Figure 2.10. Distribution of DNA methylation patterns in female hybrids across the genome. **(A)** Methylation levels of ESC/EGC-somatic hybrids and fusion partners at the *Oct4* locus were measured using bisulfite sequencing. **(B-C)** RRBS was used to examine DNA methylation levels of ESC/EGC-somatic hybrids and fusion partners across the entire genome (B) and for various genetic elements (C). The average values of the representative methylation levels for 2 XXYY ESC/MEF hybrid, 3 XXXX ESC/MEF hybrid, 2 XXYY EGC/MEF hybrid, and 3 XXXX EGC/MEF hybrid lines were shown, respectively for each analysis.

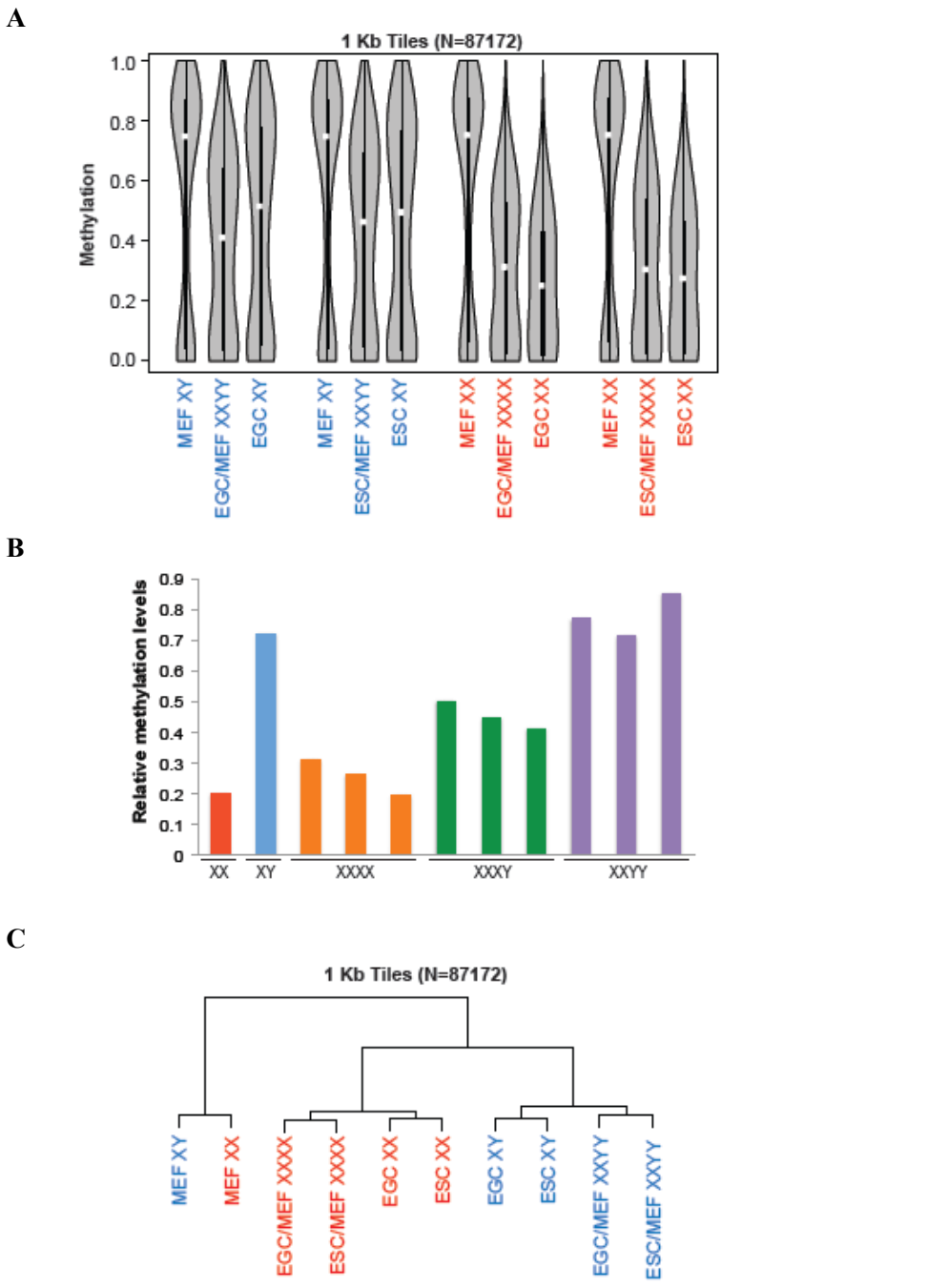


Figure 2.11. Ratio of X chromosomes to autosomes dictates global methylation levels in pluripotent-somatic cell hybrids.
(A) Global methylation levels of 4n ESC-somatic and EGC-somatic hybrids and their respective 2n fusion partners as determined by RRBS analysis. **(B)** Global methylation levels of 2n ESCs (XX and XY) and 4n ESC-somatic hybrids (XXXX, XXXY, and XXYY) as measured dot blot analysis for 5mC. **(C)** Unsupervised clustering of cell lines depicted in (A).

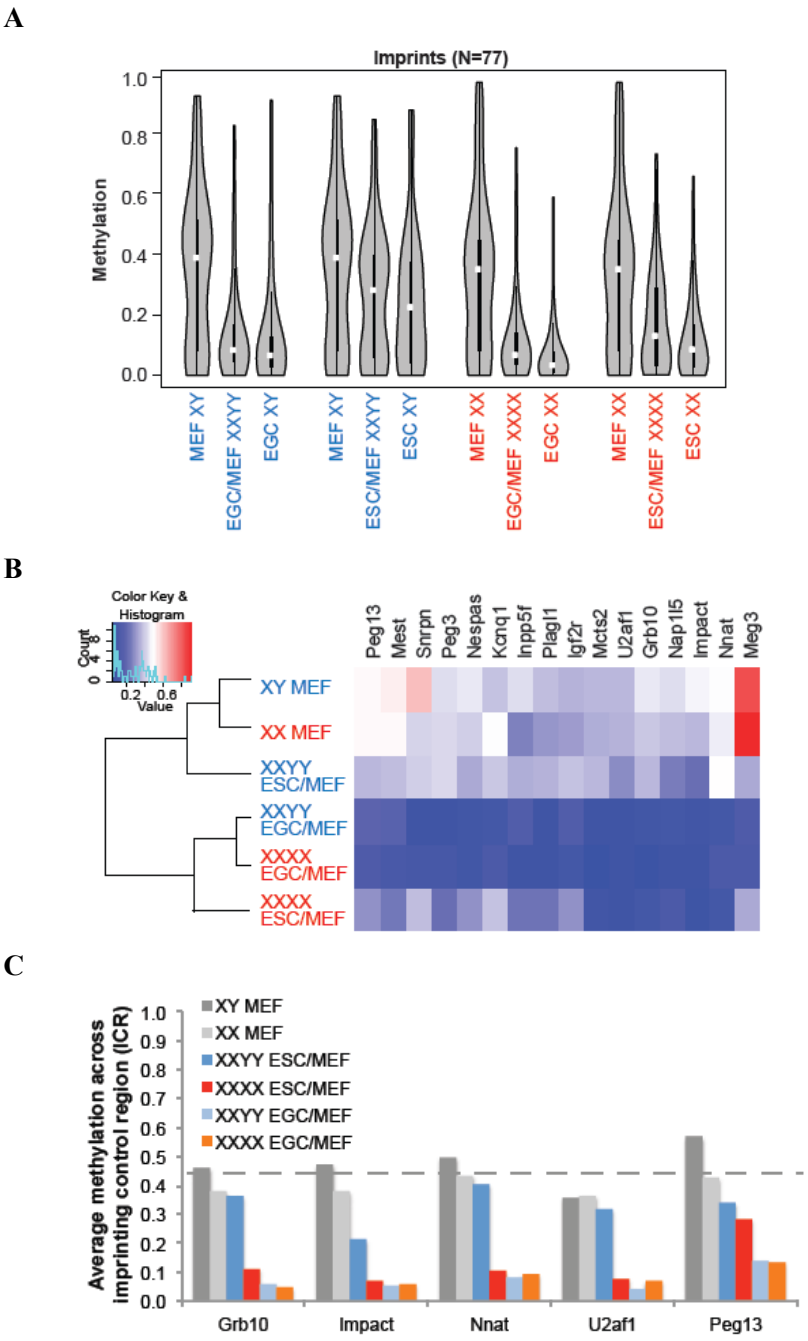


Figure 2.12. . DNA methylation patterns of imprinting control regions (ICR) at imprinted loci in pluripotent-somatic cell hybrids.

(A) Methylation levels of all captured imprinted loci from ESC/EGC-somatic hybrids and their fusion partners as measured by RRBS analysis. **(B)** Dendrogram and heatmap representation of imprinted methylation in ESC/EGC-somatic hybrids and their somatic fusion partners. **(C)** Average methylation levels across imprinted control regions (ICRs) for individual loci. The average values of the representative methylation levels for 2 XXYY ESC/MEF hybrid, 3 XXXX ESC/MEF hybrid, 2 XXYY EGC/MEF hybrid, and 3 XXXX EGC/MEF hybrid lines were shown, respectively for each analysis.

patterns as somatic cells or male parental ESCs, female ESC hybrids exhibited a marked reduction of methylation levels across most imprinted loci, akin to female parental ESCs (Figure 2.12A). Strikingly, EGC hybrids showed an even stronger reduction of imprinted gene methylation than female ESC hybrids, implying erasure of somatic imprints by an EGC-specific, sex-independent demethylation mechanism (Figure 2.12B and C). In agreement, unsupervised clustering based on all imprinted genes separated EGC hybrids and female ESC hybrids into one group and male ESC hybrids into another group (Figure 2.12B). Altogether, these results show that global methylation patterns in ESC-somatic and EGC-somatic hybrids are dictated by the ratio of X chromosomes to autosomes, regardless of the origin of the pluripotent fusion partner, whereas imprinted gene methylation is influenced by both sex and cell type.

Search for candidate genes that drive global hypomethylation in female pluripotent stem cells

To elucidate the mechanisms that may be responsible for global DNA hypomethylation in female stem cells, we compared transcriptional profiles of isogenic male and female ESCs and EGCs by RNA-seq analysis. Unsupervised hierarchical clustering showed that female stem cells are grouped separately from male stem cells, which is consistent with our methylation data but contradicts previous reports[36,80,138] (Figure 2.13); we surmise that some of these previously analyzed female cell lines may have lost one of the two X chromosomes upon passaging, explaining the lack of separation between male and female ESC/EGC lines. Moreover, EGC lines in one of these studies were derived from E8.5 germ cells[36], which are known to exhibit heterogeneous methylation levels[76] and this fact could have contributed to the observed differences.

We next determined mRNA levels of known regulators of DNA methylation in our dataset, assuming that they would be differentially regulated in male and female stem cells. Unexpectedly, expression levels of Dnmt1, Uhrf1, Dnmt3a, Dnmt3b, Tet1 and Tet2 were not

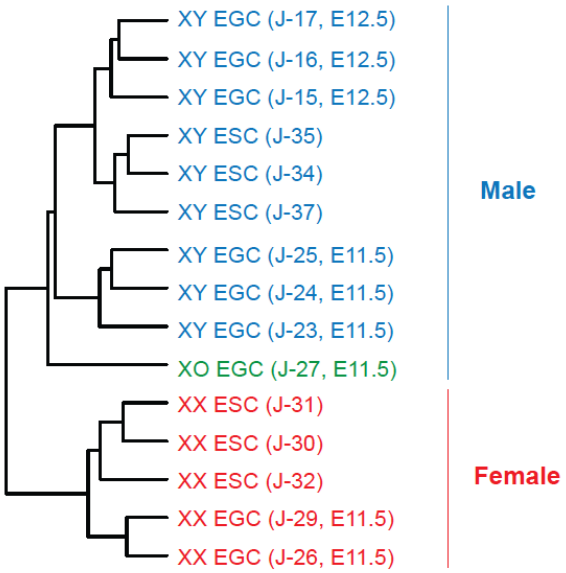


Figure 2.13. Global transcriptional patterns reflect sex rather than cell type.
Unsupervised hierarchical clustering of RNA-sequencing data obtained from isogenic ESCs and EGCs.

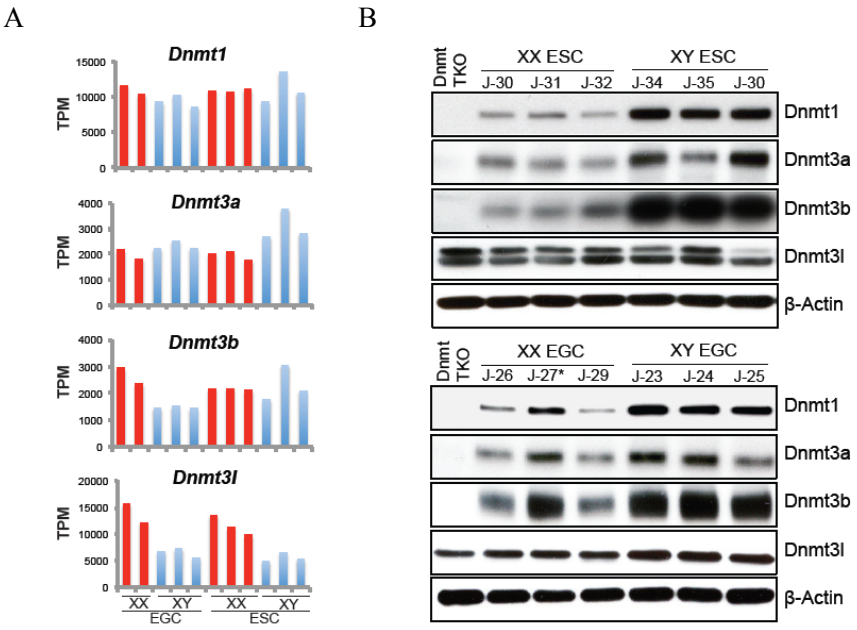
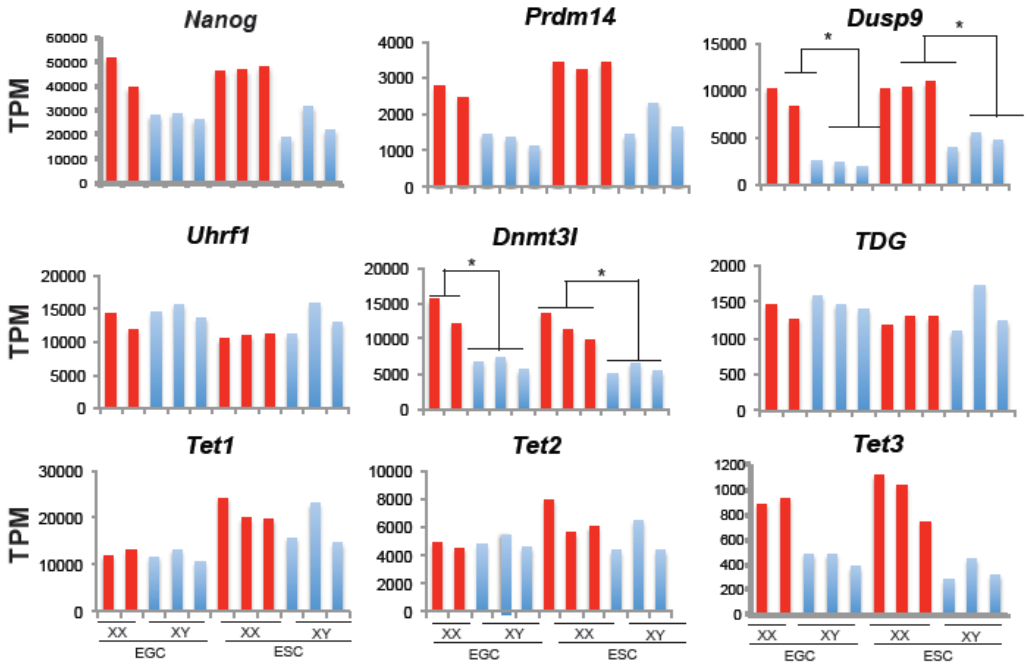


Figure 2.14. Expression levels of Dnmts in isogenic ESCs and EGCs.
(A) mRNA levels for *Dnmt1, Dnmt3a, Dnmt3b,* and *Dnmt3l* in isogenic ESCs and EGCs. **(B)** Western blot analyses for Dnmt1, Dnmt3a, Dnmt3b, and Dnm3l in isogenic ESCs and EGCs. "J-27*" denotes XO EGC lines.

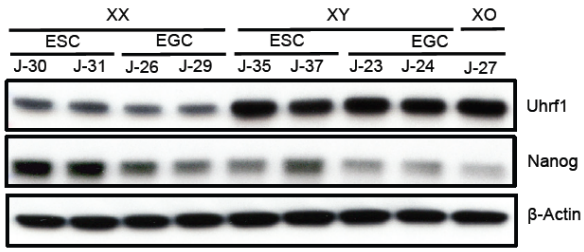
significantly different between male and female cell lines (Figure 2.14A and 2.15A). However, we noticed striking differences at the protein level for these enzymes, with female cell lines showing substantially lower expression of Dnmt1, Dnmt3a, Dnmt3b and Uhrf1 and elevated expression of Tet2 relative to male cell lines (Figure 2.14B, 2.15B and 2.15C). Similarly, *Nanog* transcript abundance was comparable between male and female stem cell lines while Nanog protein levels were roughly two-fold higher in female ESC/EGC lines compared to male ESC/EGC lines (Figure 2.15B). This finding implies that some of the previous studies comparing ESCs and EGCs might have been confounded by this sex effect. For example, EGCs were reported to express Nanog at a higher level compared to ESCs[139]. Given that this study used a female EGC line but a male ESC line, our observation may explain this previous interpretation.

We next searched for candidate upstream regulators of female hypomethylation that are X-linked and dependent on the X chromosome-to-autosome ratio. These search criteria were based on the observation that female XO stem cells regain DNA methylation[84,85] (Figure 2.4A, 2.4B and 2.5A-C) and our finding that pluripotent 2n and 4n cell lines exhibit equivalent 5mC levels (Figure 2.11A and B). We identified 254 X-linked genes that were commonly upregulated in female EGCs and ESCs compared to male EGCs and ESCs, respectively (Figure 2.16A). We tried to further narrow down these candidates by searching for regulators of the MAPK or GSK3 pathways. MAPK signal inhibition was recently associated with a relative resistance to differentiation of female ESCs compared to male ESCs[85], and its suppression in concert with GSK3 inhibition (2i/LIF) keeps murine stem cells in a naïve, hypomethylated state[36,78,79,140]. This effort uncovered the dual specificity phosphatase Dusp9 as a promising candidate gene as it meets the aforementioned criteria; *Dusp9* is X-linked, highly expressed in female ESCs/EGCs (Figure 2.15A and 2.16B) and a known inhibitor of the MAP kinases ERK1/2[141], whose coding regions are located on autosomes. Moreover, DUSP9 overexpression reportedly keeps ESCS in a self-renewing, undifferentiated state in the absence of BMP4 while its knockdown triggers

A



B



C

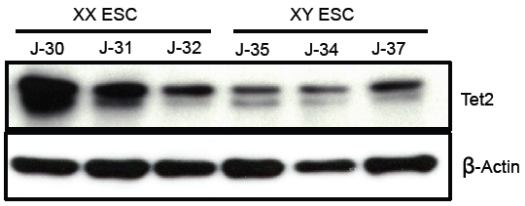


Figure 2.15. Expression levels of candidates implicated in DNA hypomethylation in female stem cells.
(A) mRNA levels of indicated genes associated with DNA methylation in isogenic ESC and EGC lines. **(B)** Western blot analysis for Uhrf1 and Nanog in isogenic ESC and EGC lines. **(C)** Western blot analysis for Tet2 in isogenic ESC and EGC lines.

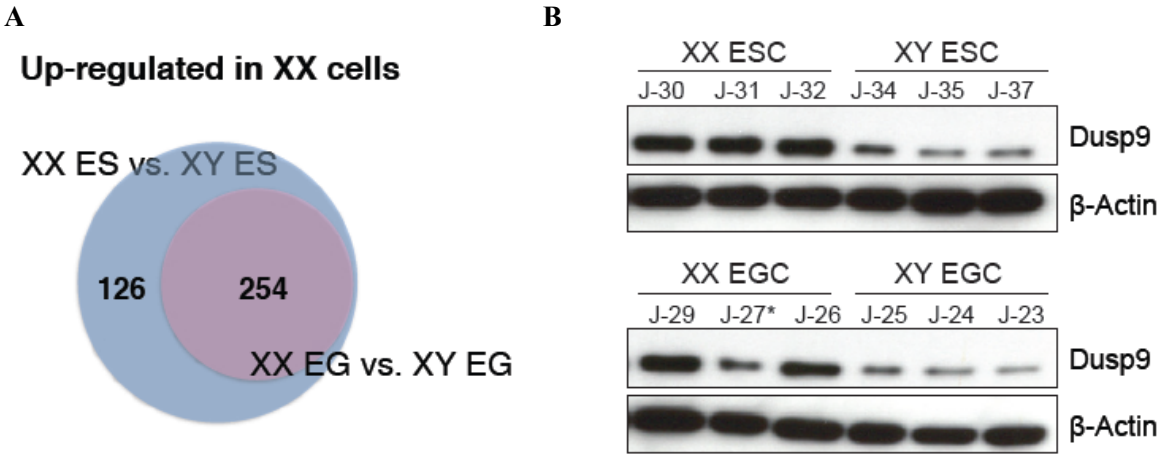


Figure 2.16. Upregulation of X-linked gene, *Dusp9* in female stem cells.
(A) Upregulated X-linked genes in XX EGC and ESC lines. **(B)** Western blot analysis for Dusp9 in isogenic XX and XY ESC/EGC lines.

differentiation[141].

Suppression of MAPK signaling by Dusp9 in male ESCs partially mimics female hypomethylation

We next tested the hypothesis that MAPK inhibition by Dusp9 may be responsible for the observed hypomethylation in female stem cells. It should be noted that recent studies concluded that MAPK inhibition alone was insufficient to explain global hypomethylation in 2i culture[78,85]. Likewise, the less differentiated state of female ESCs compared to male ESCs was suggested to require parallel suppression of multiple signaling pathways including MAPK, GSK and AKT[85]. Thus, the consequences of MAPK inhibition alone on global DNA methylation patterns in ESCs remain unexplored. We therefore determined the extent by which pharmacological suppression of the MAPK pathway reduces DNA methylation levels in male ESCs propagated in serum and LIF. Remarkably, treatment of male ESC lines with the MAPK inhibitor PD0325901 over three passages lowered global methylation levels to the same degree as those in female ESCs, indicating that MAPK inhibition alone can fully explain the hypomethylation phenotype in female stem cells (Figure 2.17A-C). Accordingly, unsupervised clustering showed a striking similarity between MAPK inhibitor-exposed male ESCs and female ESCs (Figure 2.17D). In further support of this notion, the distribution of methylation across CpG islands, shores and repetitive elements was equivalent between PD0325901-treated male cells and female cells (data not shown).

Lastly, we asked if overexpression of *Dusp9* in male stem cells could mimic the PD0325901-induced phenotype by introducing a doxycycline-inducible *Dusp9* transgene into male KH2 ESCs[142] (Figure 2.18A and B). Indeed, exposure of this engineered cell line to doxycycline reduced global methylation levels two- to three-fold (Figure 2.18C, E and F). Western blot analysis further demonstrated that *Dusp9* overexpression causes substantial down-regulation of Dnmt3a and Dnmt3b protein levels in doxycycline-treated male ESCs compared to untreated controls, consistent with our observations in female ESCs (Figure 2.18D).

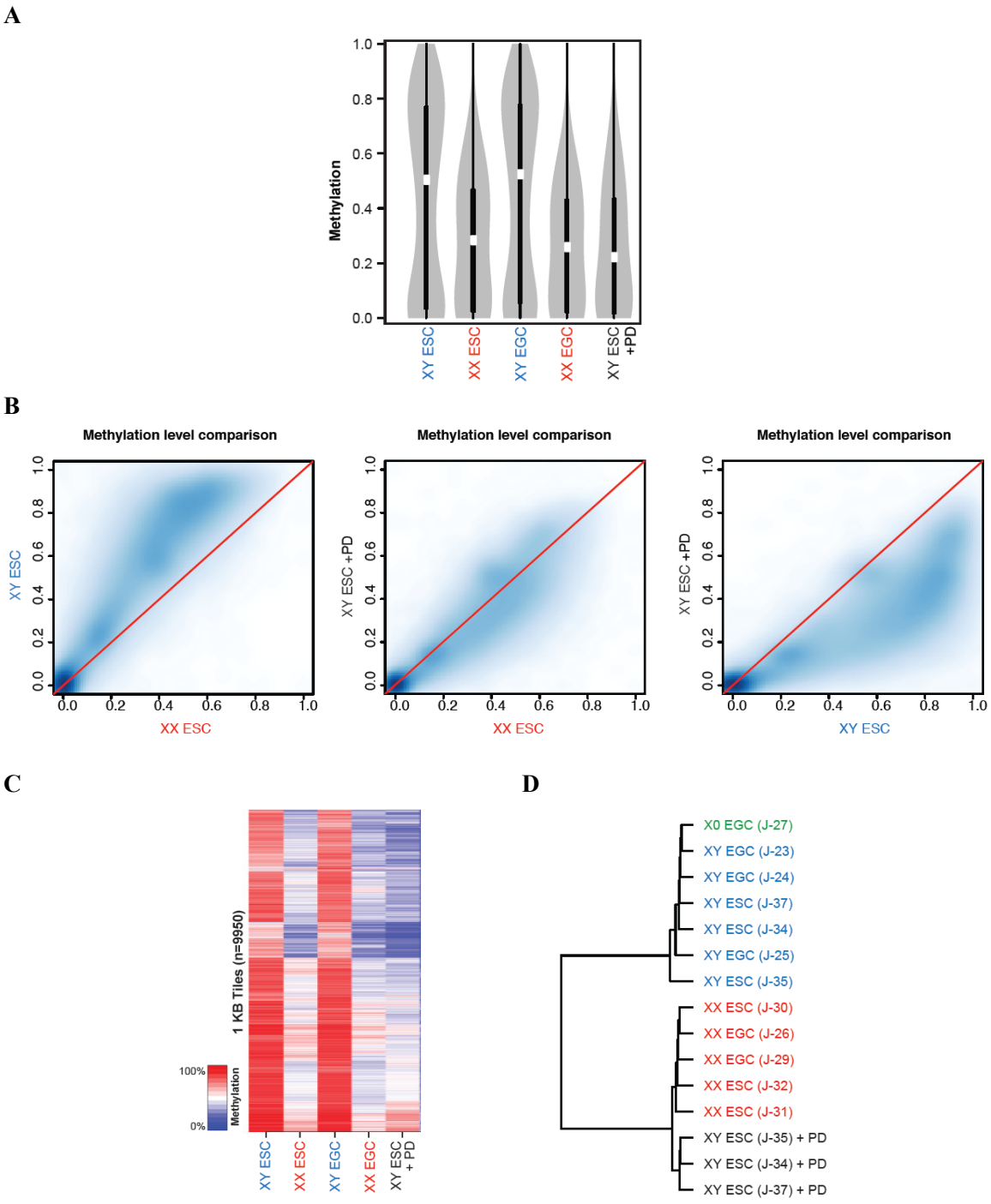


Figure 2.17. MAPK inhibition results in global DNA methylation levels in male stem cells. **(A)** Violin plot of representation of global methylation levels in XY ESCs treated with PD0325901 for three passages and in untreated control cell lines. **(B)** Scatter plots showing correlation of methylation patterns between indicated pairs of cell lines. **(C)** Heatmap representation of methylation levels across 1kb tiles in samples shown in (A). **(D)** Unsupervised clustering of isogenic XX and XY ESC/EGC lines and XY ESC lines treated with PD0325901 based on global methylation levels.

In contrast to female ESCs, Dnmt1 protein levels remained unchanged after *Dusp9* overexpression, possibly explaining the less pronounced demethylation phenotype (Figure 2.18D). We conclude that global DNA hypomethylation in female stem cells can be partially ascribed to Dusp9-mediated inhibition of MAPK signaling, resulting in destabilization of Dnmt3a/b and subsequent passive loss of DNA methylation (Figure 2.19).

Discussion

Here, we show that EGC and ESC lines are highly similar with regards to global transcriptional and DNA methylation patterns when comparing cell lines within the same sex (Figure 2.4B and 2.13). Our results differ from previous reports, which implied that EGCs of either sex are epigenetically equivalent but different from ESCs [75,133,134]. Re-examination of these prior studies reveals that methylation assays were performed with female EGC lines whereas ESC controls were of male origin. Based on this observation and the results presented here, we conclude that the historically observed "demethylation activity" of EGCs may in fact be a consequence of sex rather than cell type. Our data further explain the previous observation that EGCs, in contrast to ESCs, cause global demethylation in fusion hybrids with somatic cells. Indeed, the original report by Tada et al.[133] analyzed hybrids between female (hypomethylated) EGC lines and female (hypermethylated) lymphocytes, resulting in hypomethylated XXXX hybrids. Based on our results, we argue that hypomethylation due to X reactivation of the somatic nucleus rather than EGC-specific demethylation underlies this observation (Figure 2.11A-C). The finding that EGC lines reportedly exhibit demethylation activity only during early passages [135] is also consistent with this notion since female ESCs tend to lose one X chromosome with time in culture, which is expected to cause a shift from hypomethylation back to hypermethylation. Even though we failed to detect methylation differences between ESCs and EGCs when analyzing

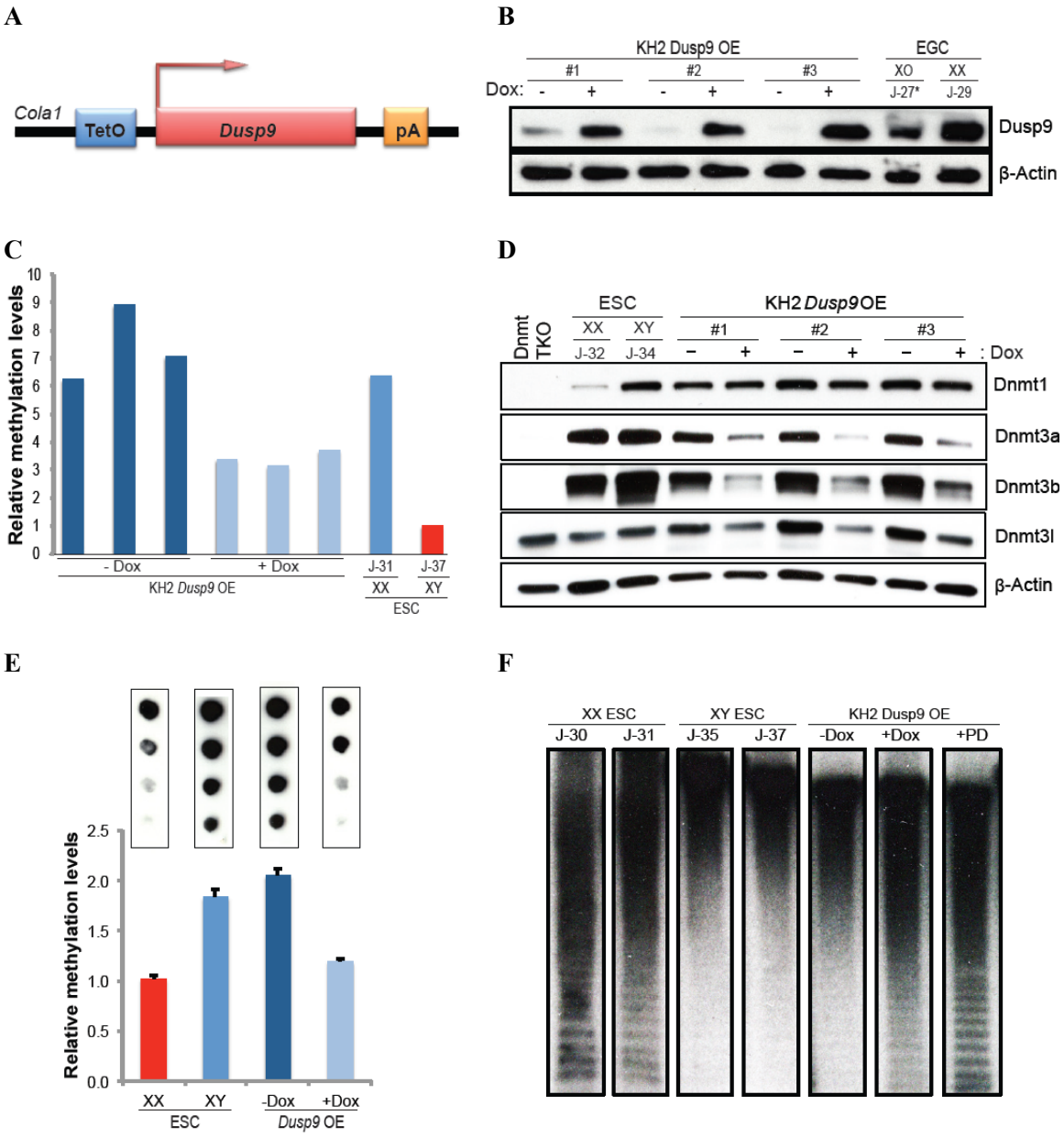


Figure 2.18. Overexpression of X-linked *Dusp9* gene results in global DNA hypomethylation in male stem cells.

(A) Schematic for the overexpression of doxycycline-inducible *Dusp9* in KH2 (male) ESCs; these cells also carry a ROSA26-M2rtTA allele. **(B)** Western blot analysis for DUSP9 levels in transgenic ESC clones exposed to doxycycline. "J-27*" denotes XO EGC line. **(C)** Dot blot analysis for 5mC in Dusp9-transgenic ESC clones exposed to doxycycline for three passages. **(D)** Western blot analysis for Dnmt1, Dnmt3a, Dnmt3b, and Dnmt3l in untreated and *Dusp9*-overexpressing ESCs. **(E)** Dot blot analysis for 5mC levels in cells overexpressing *Dusp9* for 13 passages. Shown on top are dilutions of genomic DNA for each sample, which were used to measure methylation levels below. **(F)** Southern blot analysis for minor satellite repeats using a pMR150-specific probe and HpaII digestion of genomic DNA in XX/XY ESCs, Dusp9-overexpressing KH2 ESCs, and KH2 ESCs treated with PD0325901.

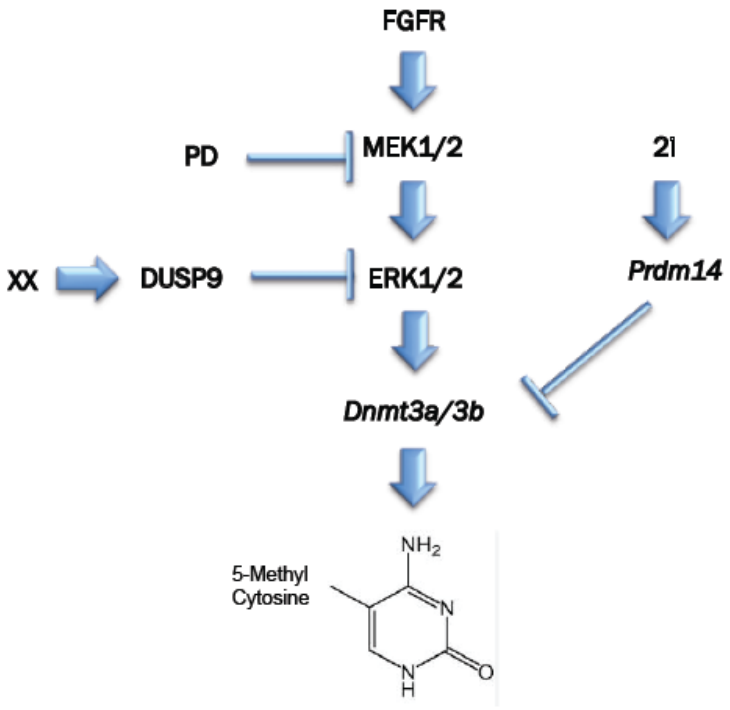


Figure 2.19. Model explaining DNA hypomethylation in female stem cells maintained in serum-LIF.
The 2i-Prdm14-Dnmt3a/3b axis is shown for comparison.

CpG islands and shores, LINEs, SINEs and LTRs, our data support the previous finding that imprinted genes are more susceptible to hypomethylation in EGC hybrids compared to ESC hybrids [133,135] (Figure 2.12A-C). However, we failed to identify candidate genes among the ~100 differentially expressed genes between isogenic ESC and EGC lines that could explain imprint-specific demethylation (data not shown). It will certainly be interesting to further understand the mechanisms by which imprinted genes may be specifically targeted in EGC-somatic hybrids.

We show that pharmacological MAPK inhibition in male ESCs fully recapitulates the observed hypomethylation phenotype of female ESCs/EGCs and generates an epigenetic state that resembles ICM cells (Figure 2.17A-D). Suppression of the MAPK pathway may thus be a key mechanism responsible for female hypomethylation rather than parallel modulation of multiple different pathways (MAPK, GSK3-beta, AKT) as was previously assumed[85]. We further discovered that the X-linked MAPK phosphatase Dusp9 is highly expressed in female ESCs/EGCs, and its forced expression in male ESCs triggered a loss of global methylation levels, providing a molecular link between X chromosome dosage and MAPK inhibition (Figure 2.18A-C). To our knowledge, this is the first X-linked regulator involved in global DNA hypomethylation of female stem cells. Our inability to fully recapitulate female-specific hypomethylation upon *Dusp9* overexpression could be due to (i) inefficient *Dusp9* expression or activity, (ii) insufficient time to (passively) demethylate the genome, or (iii) the potential ability of PD0325901 to block other kinases besides ERK1/ERK2. Alternatively, other X-linked genes may cooperate with *Dusp9* in female hypomethylation. While additional experiments are required to distinguish between these possibilities, our data clearly establish that MAPK inhibition by either PD0325901 treatment or *Dusp9* overexpression causes genomic hypomethylation (Figure 2.19). It is noteworthy that female-specific hypomethylation via Dsup9-MAPK-Dnmt3a/b appears to differ from the proposed 2i-induced process of hypomethylation. While the latter studies proposed a mechanism that involves transcriptional repression of *Dnmt3a/b* and activation of *Tet1/2* by *Prdm14* and Nanog[37,79,80,136], our data support a mechanism that operates

predominantly through post-transcriptional destabilization of Dnmt3a/b. Further experiments on the individual and combined effects of MAPK and GSK3-beta inhibition are warranted to understand these differences.

Our observations in cultured stem cells may be relevant for understanding epigenetic processes in vivo. For example, female primordial germ cells (PGCs), like female pluripotent cell lines, carry two active X chromosomes and exhibit global hypomethylation relative to male PGCs[143]. It should be interesting to ascertain whether female hypomethylation depends on reduced MAPK signaling and influences the maturation state of PGCs as is the case in ESCs[144]. Furthermore, the presence of two active X chromosomes have been linked to oncogenic gene expression patterns and malignant growth[145,146]. These findings raise the possibility that an increase in X-linked gene expression contributes to tumorigenesis by driving global DNA hypomethylation. In support of this notion, global hypomethylation is sufficient to cause genomic instability[147] and cancer[148]. Our data may further have implications for studies interrogating DNA methylation and expression patterns of ESC or EGC lines in other contexts since the sex of cell lines could profoundly influence results. Genetic background and passage number of female ESCs could introduce additional variability because these parameters are known to affect the likelihood of female stem cells to lose one of the X chromosomes. Indeed, a recent study found that methylation patterns of mouse ESCs exposed to 2i media is strongly influenced by the sex of cell lines[80]. Given that DNA methylation patterns are tightly linked with histone marks, it is possible that previous studies examining chromatin modifications may have been influenced by the sex of cell lines[38].

In summary, our data provide new molecular insights into the role of sex chromosomes and associated signaling pathways on the epigenetic state of distinct pluripotent stem cells. Our data further resolve some of the discrepancies surrounding methylation differences between ESCs and EGCs and emphasize the importance of controlling for sex, genetic background and culture conditions when assessing the epigenome and transcriptome of pluripotent stem cell lines.

Mouse ESC generation

E3.5 blastocysts were retrieved from the uterine horn by flushing and collected under a dissection microscope. A whole blastocyst was plated onto MEF feeder (Globalstem) containing 2i-LIF medium[35]. The cells were cultured for 4 days without changing medium. After that, the cells were cultured for additional 3-4 days and fed every two days with fresh 2i-LIF medium until picking. Outgrowth from the culture was picked, trypsinized, and re-plated onto MEF feeders containing KO DMEM (Invitrogen) supplemented with 15% FBS (Hyclone), 1 mM L- glutamine, 100 uM non-essential amino acids, and 0.1 mM beta-mercaptoethanol, and LIF.

Mouse EGC generation

Genital ridges containing primordial germ cells (PGCs) were retrieved from E11.5/12.5 mouse embryos. PGCs were collected from the genital ridges by trypsinization and centrifugation. Subsequently, isolated PGCs were cultured on stem cell factor (SCF)-secreting MEF feeder containing 2i-LIF medium[35] supplemeted with 60 ng/ml exogenous SCF and 10 ng/ml bFGF. The cells were cultured for 3 days without changing medium. After that, the cells were cultured for additional 7-11 days and fed every other day with fresh 2i-LIF medium until picking. Outgrowth from the culture was picked, trypsinized, and re-plated onto MEF feeders containing KO DMEM (Invitrogen) supplemented with 15% FBS (Hyclone), 1 mM L- glutamine, 100 uM non-essential amino acids, and 0.1 mM beta-mercaptoethanol, and LIF.

AP Staining and Immunostaining

Alkaline phosphatase staining was performed with the Vector Red substrate kit (Vector Labs). Immunostaining was done as previously described[149]; anti-Nanog antibody (Abcam; ab80892) and anti-Sox2 antibody (Santa Cruz; sc-17320).

Bisulfite Sequencing

Bisulfite treatment of DNA was performed with the EpiTect Bisulfite Kit (Qiagen) according to manufacturer's instructions. Amplified products were purified by using gel filtration columns, cloned into the pCR2.1-TOPO vector (Invitrogen), and sequenced with M13 forward and reverse primers.

Southern Blot Analysis for Global DNA Methylation

One microgram of genomic DNA was digested with HpaII and fragments were separated on a 0.8% agarose gel. DNA was blotted onto HybondXL membrane (Amersham Biosciences) and hybridized with the pMR150 probe as previously described[150].

Dot blot assay

Genomic DNA samples were prepared with 2-fold serial dilutions in TE buffer and then denatured in 0.4 M NaOH/10 mM EDTA at 95°C for 10 min and followed by adding an equal volume of cold 2 M ammonium acetate (pH 7.0). Denatured DNA samples were spotted on a nitrocellulose membrane in an assembled Bio-Dot apparatus (Bio-Rad) according to manufacturer's instruction. The membrane was washed with 2X SSC buffer and UV-cross linked for 10 min. Then the membrane was blocked with 5% non-fat milk for 1 hr and incubated with anti-5mC (Active Motif; 39649) for HRP-conjugated secondary antibodies and enhanced chemiluminescence detection. The membrane was subsequently stained with methylen blue to confirm corresponding amounts of DNA per each sample. Quantification was performed by image-J software analysis.

Reduced representation bisulfite sequencing (RRBS)

Global, basepair-resolution measurements of DNA methylation was measured by Reduced Representation Bisulfite Sequencing (RRBS) as described in (http://genomebiology.com/2012/13/10/r92). Briefly, genomic DNA from ESCs was digested using the MspI enzyme which cuts at C^CGG sites. Bisulfite treatment of DNA fragments was used to convert unmethylated cytosines to uracil, and this change was observed after sequencing and aligning library reads to the reference genome. Libraries were sequenced using the HiSeq 2500 platform, and aligned to an in-silico MspI-digested mm9 genome using MAQ(http://maq.sourceforge.net/) and the Picard Pipeline (http://broadinstitute.github.io/picard/faq.html).

Cell Fusion

One million EGCs or ESCs were combined with one million MEFs and fused with PEG-1500 (Roche) according to the manufacturer's directions. Selection was initiated 24 hr post-fusion with puromycin (2 mg/mL) and neomycin (300 mg/mL). Cell-cycle analysis was performed on a FACSCalibur (BD) with propidium iodide; signal area was used as a measure of DNA content.

RNA-sequencing

Undifferentiated EGCs/ESCs were sorted for SSEA1+ to control for the homogeneity of cells before RNA extraction. The quality and quantity of total input mRNA was determined on an Agilent BioAnalyzer 2100 using Agilent RNA 6000 Nano kit. One microgram of total RNA from each sample was then used as input for library preparation using Illumina TruSeq RNA Sample Prep Kit, following manufacturer's instructions. Each paired-end library was prepared with an adaptor with unique index sequence. The size profile and quantity of resulting libraries were than determined on the BioAnalyzer 2100 with Agilent High Sensitivity DNA kit. These libraries were then pooled together at equal molar concentration and sequenced on an Illumina HiSeq 2000.

Western blot analysis

In order to control for the homogeneity of the cells, differentiated EGCs/ESCs were removed by pre-plating for 1 hr at 37'C. Whole cell lysates from the cells were loaded to 4-20% gradient SDS-PAGE gels and then transferred to nitrocellulose membranes (BIO-RAD) by using Trans-Blot® TurboTM Transfer System (BIO-RAD). Blocked membranes were incubated with antibodies against the protein of interest; anti-Dnmt1 (Cell Signaling; 5119S), anti-Dnmt3a (Santa Cruz; sc-20703), anti-Dnmt3b (Abcam; ab13604), anti-Dnmt3l (Cell Signaling; 12309S), anti-Dusp9 (Abcam; ab167080), and anti-beta-Actin (Cell Signaling; 5125S).

Gene targeting of ESCs

Dusp9 cDNA was cloned into the shuttle plasmid pBS31 using NotI and MfeI digestion. The resulting plasmid was electroporated into KH2 ESCs together with a plasmid driving expression of Flp recombinase[142]. Correct targeted clones were screened for by hygromycin selection and confirmed by Southern blotting as previously described[142].

Bioinformatistics analyses

The methylation at each covered CpG was calculated as the number of reads in which the CpG was methylated (no bisulfite conversion took place) divided by the number of total reads covering the CpG. Region analysis was performed using methods described in (http://dx.doi.org/10.1038/nmeth.1414) on promoters (2Kb centered at refSeq transcription start sites), 1Kb tiles, and imprinting control regions as defined in (http://www.ncbi.nlm.nih.gov/pubmed/18600261).

Chapter 3

Genetic background drives transcriptional differences between human ESCs and iPSCs

Attributions

I generated all hiPSC lines and *in vitro* differentiated fibroblasts, which are genetically matched to HUES2 or HUES3 lines. Alex Meissner independently generated additional hiPSC lines. I designed all assays. I performed AP staining, immunostaining, lactate production and glucose uptake assays, and Western blot and qPCR assays. Hotae Lim and In Young Choi performed neural differentiation experiments and completed immunostaining, qPCR assays and Western bot assays on the differentiated cells. For the RNA-sequencing, I collected samples and isolated RNA, and the Tufts Genomics Core performed the sequencing. Peter Park, John Rinn, Soohyun Lee, William Mallard, Guidantonio M. Taliazucchi, and Francesco Ferrari performed bioinformatics analysis of RNA-sequencing data.

Abstract

Despite the therapeutic potential of transcription factor-induced pluripotency, it remains controversial whether human induced pluripotent stem cells (hiPSCs) are equivalent to human embryonic stem cells (hESCs). Here, we compared isogenic, transgene-free hESC and hiPSC lines in order to assess the contribution of genetic background and reprogramming method to transcriptional patterns. Surprisingly, transcriptional variations originating from different genetic backgrounds were dominant over variations due to the different cell type of origin of pluripotent cell lines. Moreover, the few differences we detected between hESC and hiPSC lines neither predicted functional outcome, nor could they distinguish an independently derived, larger set of unmatched hESC/hiPSC lines. We conclude that hESCs and hiPSCs are transcriptionally and functionally highly similar and cannot be distinguished by a defined gene expression signature. Our data further imply that genetic background variation is a major confounding factor for transcriptional comparisons, explaining some of the previously observed expression differences between hESCs and hiPSCs.

Introduction

Induced pluripotent stem cells (iPSCs) are derived from somatic cells upon overexpression of defined transcription factors such as *Oct4, Sox2, Klf4* and *c-Myc*[90]. Patient-specific hiPSCs provide a platform to study diseases in a dish and they may be used to treat degenerative diseases in the future[91,92,151]. Given the therapeutic potential of reprogramming technology, it is imperative to understand the extent to which hiPSCs are similar to hESCs, which serve as the gold standard of pluripotent cell lines. Previous studies came to conflicting conclusions regarding this question. While initial reports noted that hESC and hiPSC lines are

fundamentally different at the transcriptional level, subsequent studies concluded that both hESC and hiPSC lines exhibit a similar degree of variability[88,119,152]. More recent reports suggested that hiPSC lines could be distinguished from hESC lines based on small sets of differentially expressed genes (DEGs)[130,153]. However, different reprogramming methods and culture conditions were employed in these studies and their impact on transcriptional patterns remains unclear[120,121]. Moreover, the analyzed hiPSC lines were generally of a different genetic background compared to control hESC lines, which may have influenced transcriptional patterns[88,119-122,128,130,152].

We previously showed that a comparison of genetically matched mouse ESC and iPSC lines eliminates the majority of transcriptional differences that were observed between unmatched cell lines[118]. Although we were unable to identify consistent transcriptional differences between murine ESC and iPSC lines in that study, we discovered a small group of transcripts that was aberrantly silenced in a subset of iPSC lines and adversely affected their developmental potential. The goal of this study was to extend our analyses to the human system to ask whether molecular differences can be identified in hiPSC lines relative to hESC lines when controlling for genetic background and reprogramming methodology.

Results

Approach to generate isogenic hESCs and hiPSCs

To compare genetically matched hESC with hiPSC lines devoid of viral integration, we set out to generate hiPSCs from in vitro-differentiated hESCs using a non-integrating Sendai virus (SeV)-based reprogramming system[97]; SeV is an RNA virus that is diluted from infected cells in a replication-dependent manner, hence leaving no genetic footprint behind[97] (Figure 3.1A and B). We chose two well-characterized hESC lines, HUES2 and HUES3[154], for these experiments. We intentionally selected male hESC lines because of the previous observation that

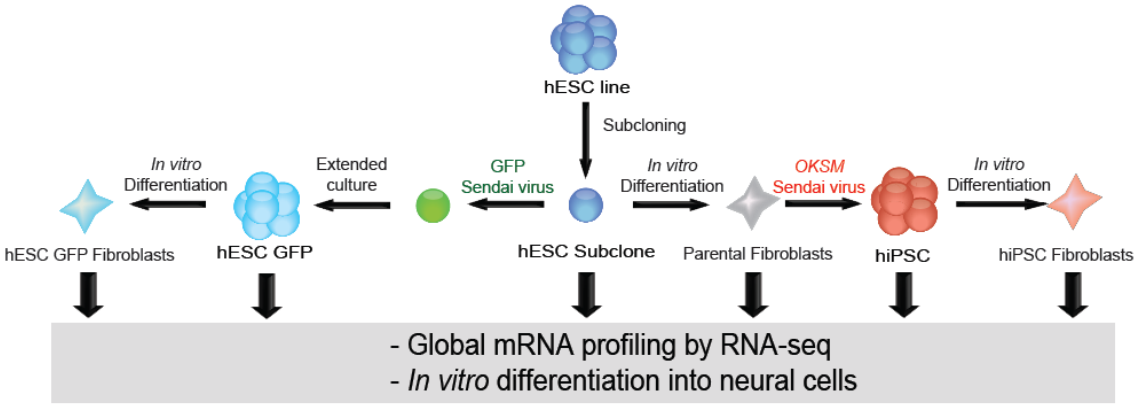
female iPSCs exhibit defects in X chromosome reactivation[155], which might confound subsequent interpretations[153].

First, we subcloned each line in order to ensure genetic and epigenetic homogeneity of cells and to properly control for the clonal origin of hiPSCs. We then in vitro differentiated these subclones by withdrawing growth factors and sorting fibroblast-like cells based on hCD90+/hTRA-1-81– expression (Figure 3.2). Importantly, these fibroblast-like cells were highly similar to primary human fibroblasts by morphological criteria and global transcriptional analysis (Figure 3.3B). Moreover, hCD90+/hTRA-1-81– populations failed to form Alkaline Phosphatase (AP)-positive colonies in hESC media, documenting proper differentiation and the absence of residual pluripotent cells in the culture (Figure 3.3A and B). In a final step, these fibroblast-like cultures were reprogrammed into hiPSCs by infecting the cells with non-integrating SeV vectors expressing *OCT4, KLF4, SOX2*, and *c-MYC* (OKSM), as was previously reported[97] (Figure 3.1A and B). Emerging colonies were isolated after approximately 3 weeks, expanded and confirmed to be positive for AP activity and endogenous OCT4 expression, indicating successful reprogramming (Figure 3.2). Moreover, we ensured loss of SeV expression in all lines, demonstrating reprogramming factor-independent self-renewal (Figure 3.4A and B).

Genetic background accounts for most transcriptional differences between hESCs and hiPSCs

We next determined whether the reprogramming method itself contributes to transcriptional changes in hiPSCs. To this end, the parental hESC subclones were infected with GFP-expressing SeV (SeV-GFP) and passaged until GFP fluorescence was no longer detectable before analyzing cell lines by RNA-sequencing (Figure 3.1A and 3.4C). Surprisingly, we found 63 genes that were differentially expressed between uninfected hESC subclones (hESC SCs) and infected hESC subclones (hESC GFPs), demonstrating that viral infection itself leads to subtle but statistically significant transcriptional changes that persist after viral loss (Figure 3.5A).

A



B

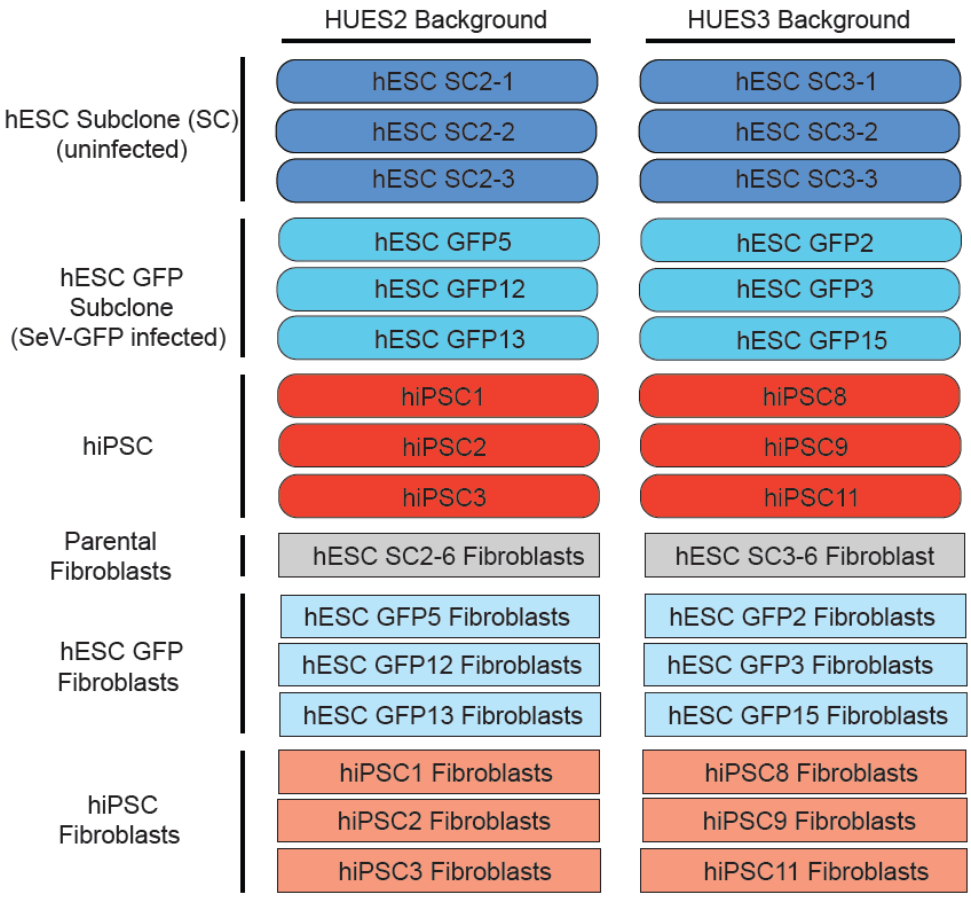


Figure 3.1. Generation of genetically matched hESC and hiPSC lines.
(A) Schematic for the generation of genetically matched hESC and hiPSC lines. **(B)** Genetically matched hESC/hiPSC lines were generated in two backgrounds and both subtypes were subsequently used for RNA-sequencing.

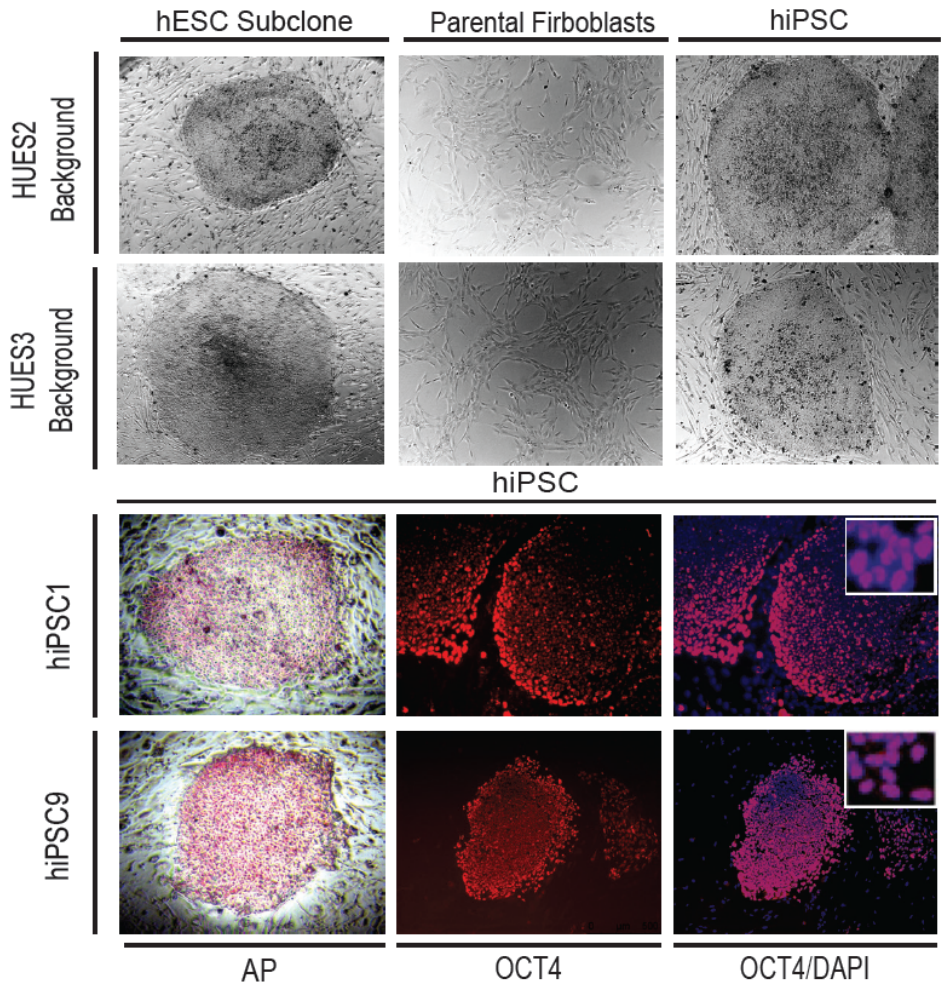
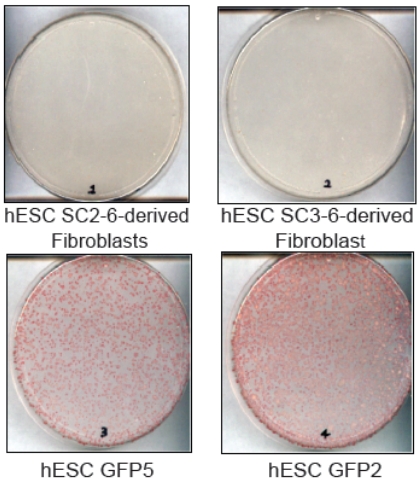


Figure 3.2. Characterization of genetically matched hESC and hiPSC lines.
(A) Top panel shows bright images of hESC subclones, parental fibroblast, whereas the bottom panel shows hiPSC lines stained for alkaline phosphatase activity or OCT4 expression. Co-staining with DAPI confirmed nuclear expression of OCT4 (inset)**.**

A



B

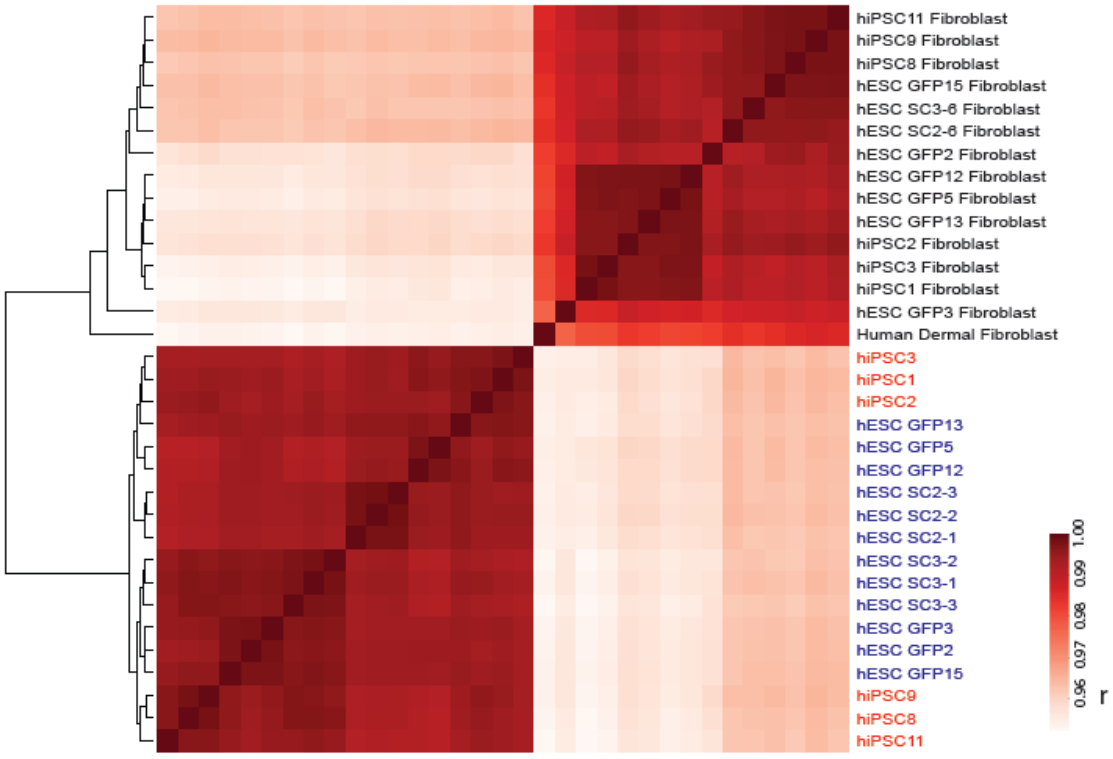


Figure 3.3. *In vitro* differentiation of hESCs into fibroblasts.
(A) Representative AP staining for parental fibroblast of each genetic background (top panel) and control hESC GFP lines from corresponding background (bottom panels) that were cultured in hESC media. Parental fibroblast formed no pluripotent colonies where hESC lines formed multiple pluripotent colonies. **(B)** Heatmap and dendrogram for all isogenic hESC subclones, *in vitro*-differentiated fibroblasts, derivative hiPSCs, and dermal fibroblast based on pairwise Pearson correlation (r) on global gene expression levels. hiPSC lines, red; hESC lines, blue; fibroblasts, black.

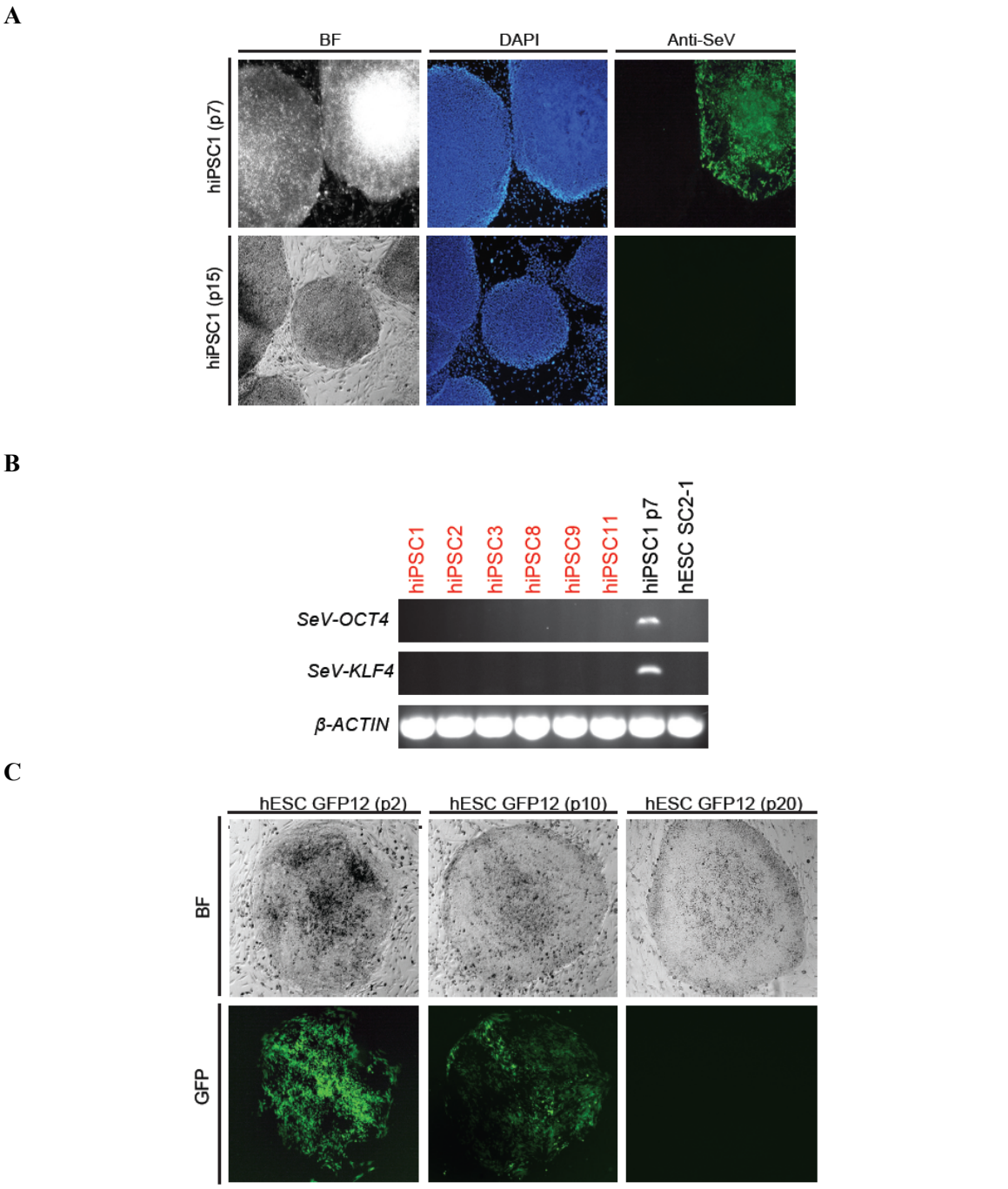


Figure 3.4. Loss of SeV expression in genetically matched hESC and hiPSC lines.
(A) The hiPSC1 lines was stained with DAPI and anti-SeV antibody at passage 7 (top panels) and 15 (bottom panels). Representative images are shown. **(B)** Transgene-specific primers were used to detect the expression of *Oct4* and *Klf4* at passage 15. hiPSC1 at passage 7 was used as a positive control. **(C)** hESCs were infected with SeV-GFP. Representative bright field and GFP expression images from the hESC GFP12 line at passage 2, 10, and 20.

Accordingly, the expression signature of this 63-DEG set could separate hESC SC lines from hESC GFP lines (Figure 3.5B). Based on these observations, we decided to use expression data from hESC GFP lines as the better controls for all subsequent comparisons with SeV-generated hiPSC lines.

When comparing transcriptional profiles of hESC subclones (hESC SCs and hESC GFPs), *in vitro*-differentiated fibroblasts and derivative hiPSCs by unsupervised clustering, the strongest differences were found between pluripotent cell lines and differentiated cell types, consistent with previous observations[93,119,120] (Figure 3.3B). We also observed a clear segregation of pluripotent samples into two transcriptionally related groups, irrespective of whether cell lines were infected with SeV or not (Figure 3.6A expanded from Figure 3.3B). This segregation could not be explained by the cellular origin of cell lines from embryos (hESCs) or somatic cells (hiPSCs) but instead correlated with the genetic background of each line. That is, HUES2 ESC clones clustered with HUES2 iPSC lines whereas HUES3 ESC clones clustered with HUES3 iPSC lines. Consistent with this finding, overall transcriptional variation among the individual groups is significantly lower when comparing hiPSCs to matched hESCs than to unmatched hESCs (Figure 3.6B). We conclude that transcriptional differences between pluripotent cell lines are predominantly driven by genetic background.

Differentially expressed genes between isogenic hESCs and hiPSCs

Although genetic background accounted for most transcriptional differences among the analyzed pluripotent cell lines, we noticed that hESCs clustered with each other and separately from hiPSCs within a given background, suggesting subtle but consistent transcriptional differences that reflect distinct cellular origins (Figure 3.6A). In order to define DEGs that distinguish hESC from hiPSC lines regardless of reprogramming method, we compared transcriptional profiles of hiPSC lines with those of matched hESC GFP lines. This analysis

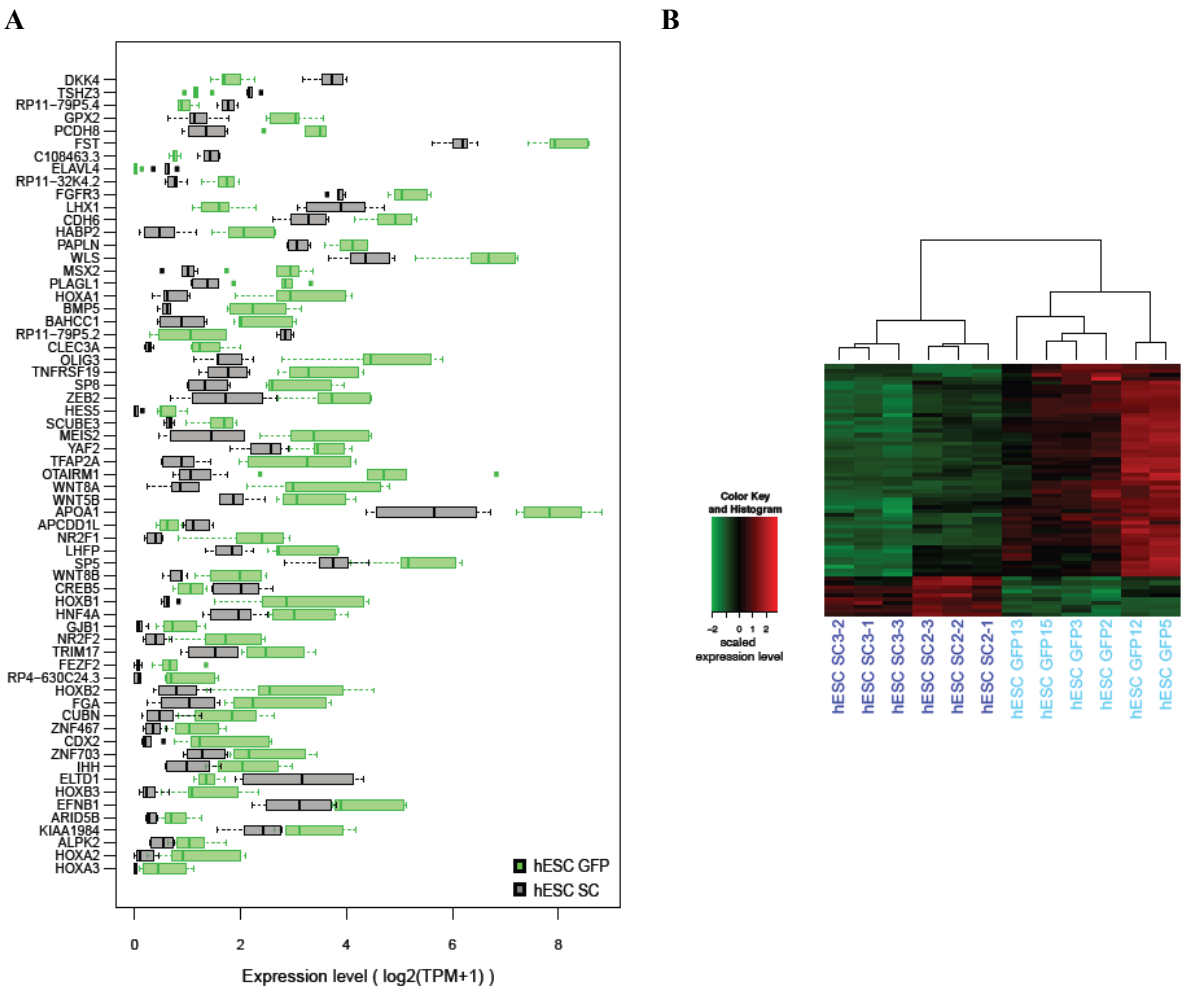


Figure 3.5. Differentially expressed genes (DEGs) between hESC SC and hESC GFP lines. **(A)** Expression levels of 63 genes that were identified to be significantly differentially expressed between 3 biological replicate hESC GFP and 3 biological replicate hESC SC lines. (FDR<0.01 and fold change >2 or <1/2, details in the Methods section) within each of the two genetic backgrounds. Green and grey boxes indicate the expression level for each differentially regulated gene in 6 hESC GFP and 6 hESC SC lines, respectively. TPM; transcripts per million. **(B)** Heatmap and dendrogram for all isogenic hESCs lines using the expression levels of the 63 DEGs from (A). hESC SC lines, blue; hESC GFP lines, light blue.

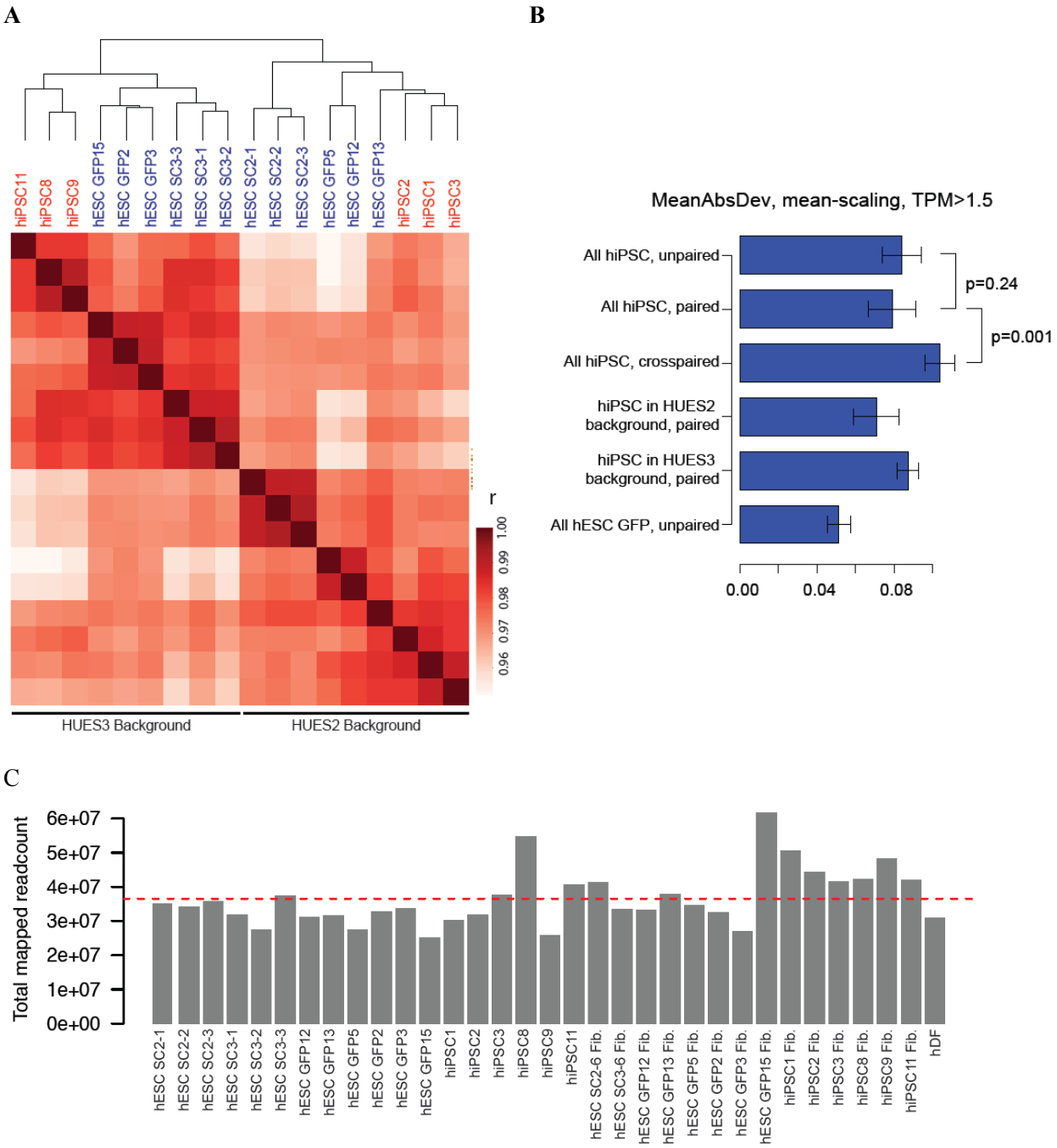


Figure 3.6. Comparison of transcriptional profiles for genetically matched hESCs and hiPSCs.

(A) Heatmap and dendrogram for all isogenic hESC and hiPSC lines based on pairwise Pearson correlation (r) on global gene expression levels (log-scaled). hiPSC lines, red; hESC lines, blue. **(B)** Barplots showing mean absolute deviation of hiPSC or hESC GFP cell lines relative to a set of reference hESC GFP lines in global gene expression level. Unpaired, relative to all hESC GFP lines; paired, relative to genetically matched hESC GFP lines; cross-paired, relative to genetically unmatched hESC GFP lines. **(C)** Total number of mapped reads for individual RNA-seq samples, technical replicates merged. Red dotted line indicates average.

revealed that 52 and 91 genes were up- and down-regulated, respectively, in hiPSC lines derived from the HUES2 background, whereas 77 and 426 genes were up- and down-regulated in hiPSC lines derived from the HUES3 background, respectively. We only focused on DEGs that were commonly dysregulated in both genetic backgrounds, yielding 49 genes in total (Figure 3.7A). The 49-DEG signature was sufficient to reliably separate our hiPSC lines from our hESC lines, as expected (Figure 3.7B). Importantly, this small number of DEGs contrasts with previous studies, which identified hundreds to thousands of DEGs when comparing unmatched hESCs and hiPSCs[88,93,119-122,152].

The relatively good depth of our RNA-seq data (~40 million mapped reads per sample on average) (Figure 3.6C) and the large number of genes identified as differentially expressed between hESC and hiPSC lines in each genetic background (Figure 3.7A) suggest that the small number of DEGs is not likely due to a low sensitivity. Although we failed to detect any Gene Ontology term that was significantly enriched among these 49 DEGs, we noticed that 48 out of 49 DEGs were downregulated in hiPSCs relative to hESCs. This raised the possibility that the associated genes were silenced in fibroblasts and failed to be properly reactivated in derivative hiPSCs. However, examination of the expression levels of these genes in fibroblast-like cells failed to show a consistent pattern, thus excluding incomplete reprogramming (Figure 3.8).

We next asked whether any of the transcriptional differences between hESCs and hiPSCs would manifest at the functional level. We focused on two DEGs, *LDHA* and *SLC2A1* (also known as *GLUT1*), because of their strong basal expression in hESC samples and reduced expression in all hiPSCs (Figure 3.8, red arrows). Both gene products are involved in energy metabolism; LDHA plays an important role in glycolysis by catalyzing the conversion of pyruvate to lactate[156], whereas SLC2A1 facilitates glucose uptake in cells[157]. Accordingly, *LDHA* and *SLC2A1* are expressed at low levels in fibroblasts, which produce energy through oxidative phosphorylation, whereas they are abundantly expressed in pluripotent cells, which produce energy through glycolysis[158] (Figure 3.8). Based on the downregulation of these two genes in all

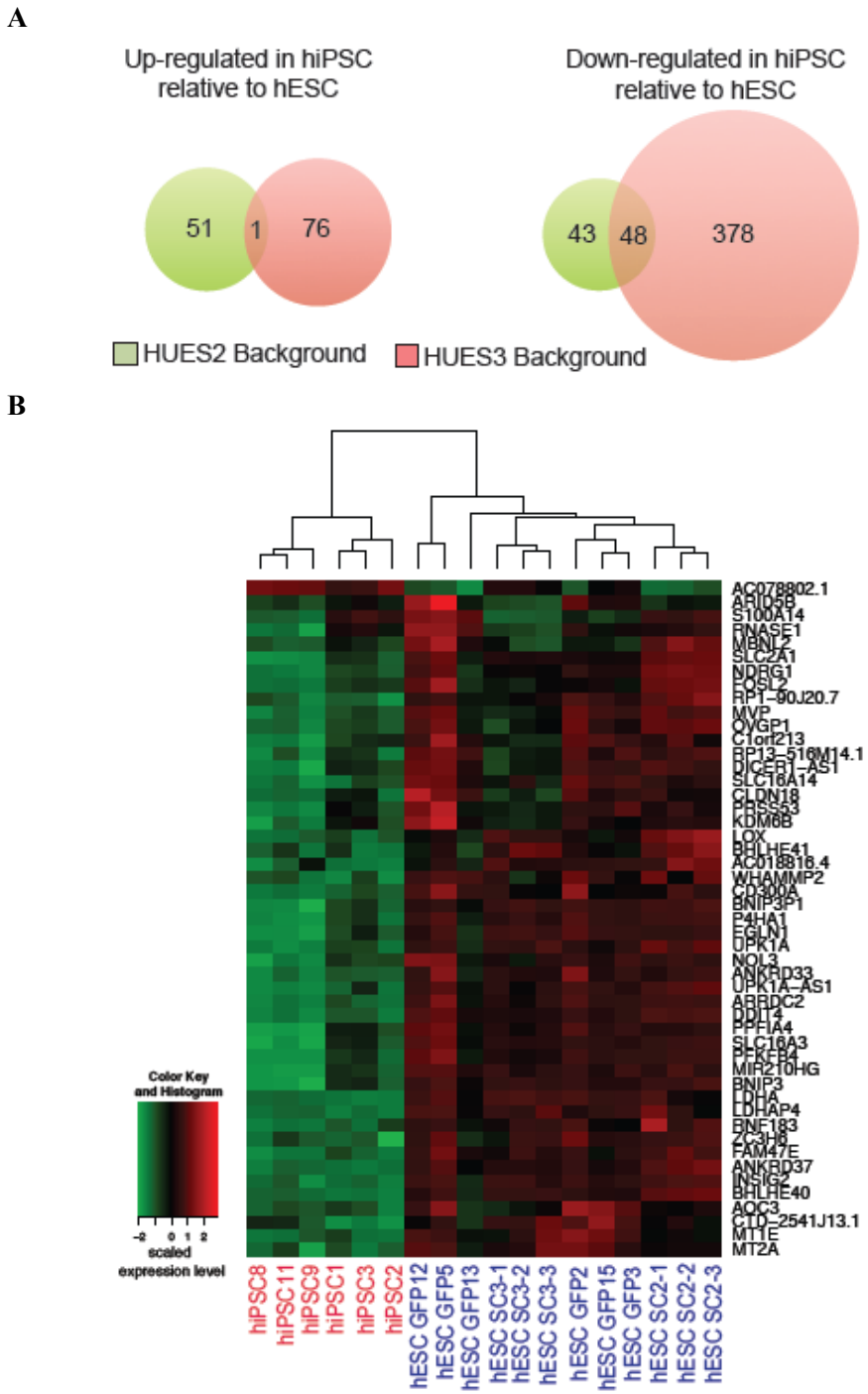


Figure 3.7. Differentially expressed genes (DEGs) between hESC GFP and hiPSC lines.
(A) Venn diagram showing the number of genes consistently up- or down-regulated in 3 biological replicate hiPSC lines when compared to 3 biological replicate hESC GFP lines from the same genetic background. (FDR<0.01 and fold change <2 or <1/2, details in the Methods section). **(B)** Heatmap and dendrogram for all isogenic hESC and hiPSC lines based on the 49 differentially expressed genes (DEGs) that were common between the HUES2 and HUES3 backgrounds, using hierarchical clustering based on row-scaled expression level. hiPSC lines, red; hESC lines, blue.

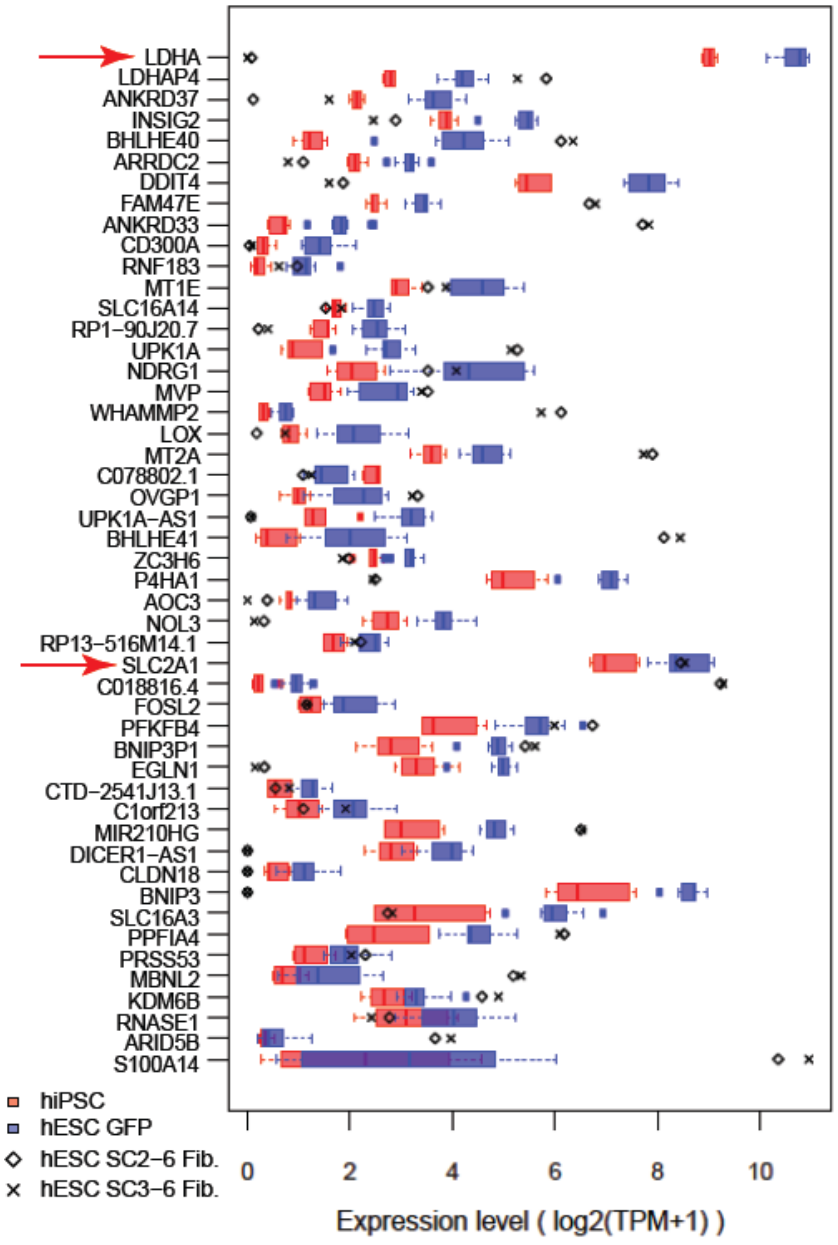


Figure 3.8. Expression levels of differentially expressed genes (DEGs) between hESC GFP and hiPSC lines.

Box plot of 6 hESC GFP lines, 6 hiPSC lines, and parental fibroblasts for the 49 DEGs. Red and blue boxes indicate the expression level of each gene in hiPSC and hESC GFP lines, respectively. Diamonds and crosses indicate the expression level of each corresponding gene in parental fibroblasts derived from HUES2 and HUES3 backgrounds, respectively. Genes are ordered by Student's t-test p-value between the 6 hiPSC and 6 hESC GFP lines. Red arrows depict genes discussed in main text.

examined hiPSC lines compared to hESC lines by RNA-seq and qPCR analyses (Figure 3.8, 3.9A and B), we hypothesized that hiPSC lines might be less glycolytic when compared to hESC GFP lines. To test this experimentally, we measured lactate production and glucose uptake levels between isogenic hiPSC and hESC GFP lines. Unexpectedly, neither lactate production nor glucose uptake levels differed between hiPSC and hESC GFP lines (Figure 3.9C). Further, examination of LDHA protein levels in these cell lines failed to show a difference despite the observed transcriptional changes (Figure 3.9D). Altogether, these results show that two of the tested 49 DEGs do not translate into functional differences, possibly due to posttranscriptional compensatory mechanisms.

The absence of pronounced transcriptional differences between undifferentiated hESCs and hiPSCs does not rule out the existence of iPSC-specific aberrations that become detectable only after differentiation. We therefore also performed RNA-sequencing of fibroblasts derived from all hESC and hiPSC lines following *in vitro* differentiation (Figure 3.1A). This comparison uncovered only two DEGs that were consistently upregulated in hiPSC-derived fibroblasts compared to hESC-derived fibroblasts and there was no overlap with the 49 DEGs identified between undifferentiated hESC and hiPSC lines (Figure 3.10A and B). We conclude that few transcriptional differences persist in differentiated cells derived from isogenic hESC and hiPSC lines.

Dysregulation of genes in a subset of hiPSC lines

Given that most of the DEGs between undifferentiated hESC GFP and hiPSC lines produced low abundance transcripts that were not obviously connected trough a common biological process (Figure 3.8), we next focused on genes that were dysregulated in only a subset of hiPSC lines. To distinguish these transcripts from DEGs that are dysregulated across all hESC and hiPSC lines, we refer to these transcripts as “inconsistently Differentially Expressed Genes” (iDEGs) (Figure 3.11). We have previously shown that iDEGs, identified between isogenic

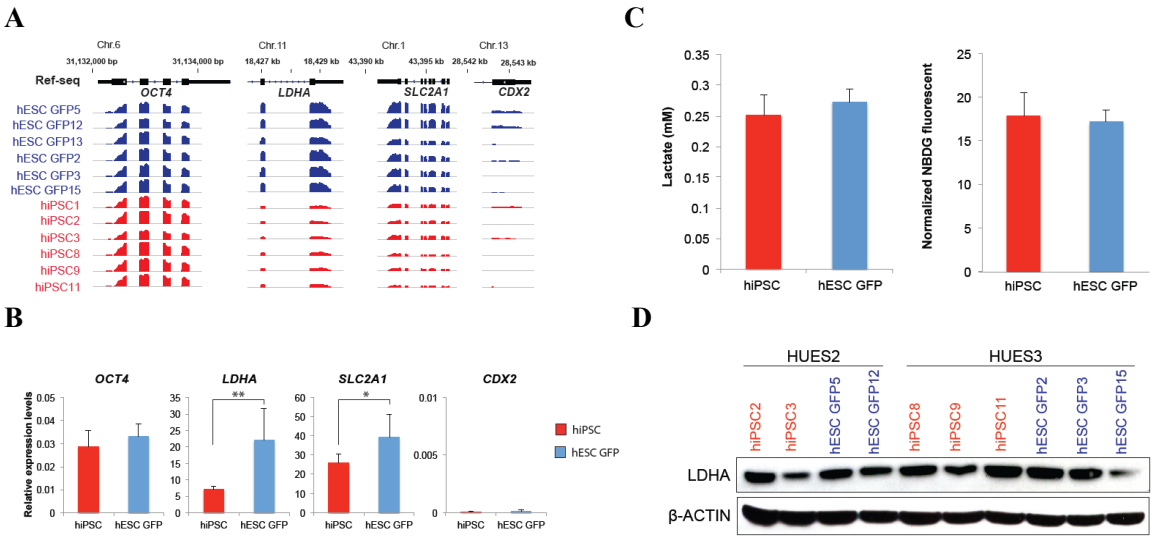


Figure 3.9. Functional tests of differentially expressed genes (DEGs) between hESC GFP and hiPSC lines.
(A) RNA-seq read density of hESC GFP and hiPSC lines for *OCT4, LDHA, SLC2A1,* and *CDX2*. **(B)** Expression levels of *OCT4, LDHA, SLC2A1,* and *CDX2* by qPCR in hESC GFP and hiPSC lines. Normalized to *ACTB*. (n=6). Student's t-test *, p<0.05; **, p<0.01. Mean ± s.d. **(C)** Lactate production levels (left) and glucose uptake levels (right) of hESC GFP and hiPSC lines. (n=6). Mean ± s.d. Red arrows **(D)** Representative Western blot for LDHA levels in hESC GFP (blue) and hiPSC (red) lines.

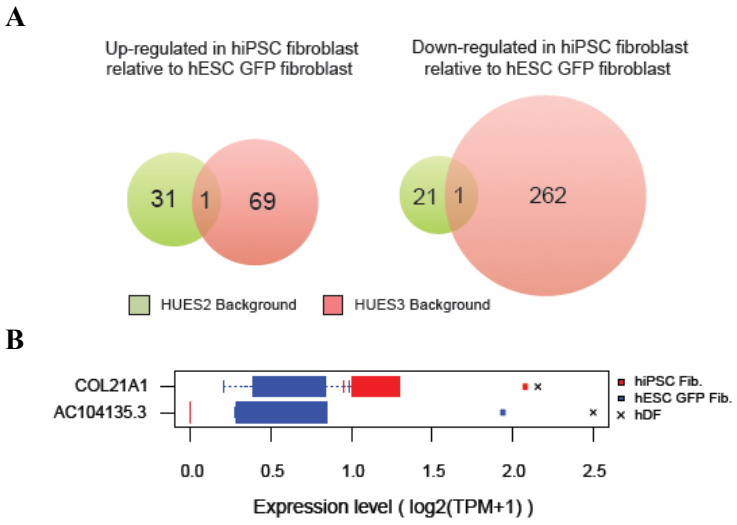


Figure 3.10. Comparison of isogenic hESC GFP and hiPSC lines for differentiation potential into fibroblasts.
(A) Venn diagram showing the number of up- or down-regulated genes in 3 biological replicate hiPSC fibroblasts relative to 3 biological replicate hESC GFP fibroblasts within each genetic background. A total of 32 and 22 genes were up- and down-regulated in hiPSC fibroblasts in HUES2 background (green), respectively. Likewise, 70 and 263 genes were up- and down-regulated in hiPSC fibroblasts in HUES3 background (scarlet), respectively. **(B)** Box plot of 12 *in vitro*-differentiated fibroblast-like cell lines and primary dermal fibroblasts (cross) based on the 2 DEGs (identified in (A)) between hiPSC fibroblasts (red) and hESC GFP fibroblasts (blue).

mouse ESCs and iPSC, could predict the differentiation potential of subsets of iPSC lines[118]. When applying the same principle to our human data set, we found that 34 genes were upregulated, whereas 27 genes were downregulated in some of the HUES2-derived hiPSC lines when compared to genetically matched hESC GFP lines. Similarly, 9 genes were upregulated and 32 genes were downregulated in some of the HUES3-derived hiPSC lines relative to matched hESC GFP controls (Figure 3.11). Only eight iDEGs were commonly dysregulated between the two different genetic backgrounds and were thus selected for further analysis (Figure 3.11 and 3.12A).

Two iDEGs (*IRX2* and *DPP10*) have been previously linked to neural development and psychiatric disease[159-162]. Silencing of *IRX2* and *DPP10* in some hiPSC lines but not in any hESC lines (Figure 3.12B) was confirmed by qPCR (Figure 3.12C). To ask whether *IRX2* and *DPP10* expression status in hiPSC lines correlates with their differentiation potential into the neural lineage, we differentiated hiPSC and control hESC GFP lines into neuroectodermal cells by using a previously published dual SMAD inhibition protocol[163] (Figure 3.12D). Surprisingly, aberrant silencing of *IRX2* and *DPP10* had no effect on the potential of hiPSCs to differentiate into the neural lineage as determined by expression levels of *NESTIN, SOX1, PAX6*, and *FOXG1*, which are well-established markers of neuroectoderm differentiation from human pluripotent stem cells[164] (Figure 3.12E). Consistent with this, Western blot analysis and immunostaining demonstrated that PAX6 and SOX1 were equally expressed during neural differentiation from hiPSC and hESC GFP lines (Figure 3.12F and G). Together these results suggest that hiPSC and hESC lines share equivalent neural differentiation potentials regardless of *IRX2* and *DPP10* expression status.

Genetic background explains previously identified gene expression differences between hESCs and hiPSCs

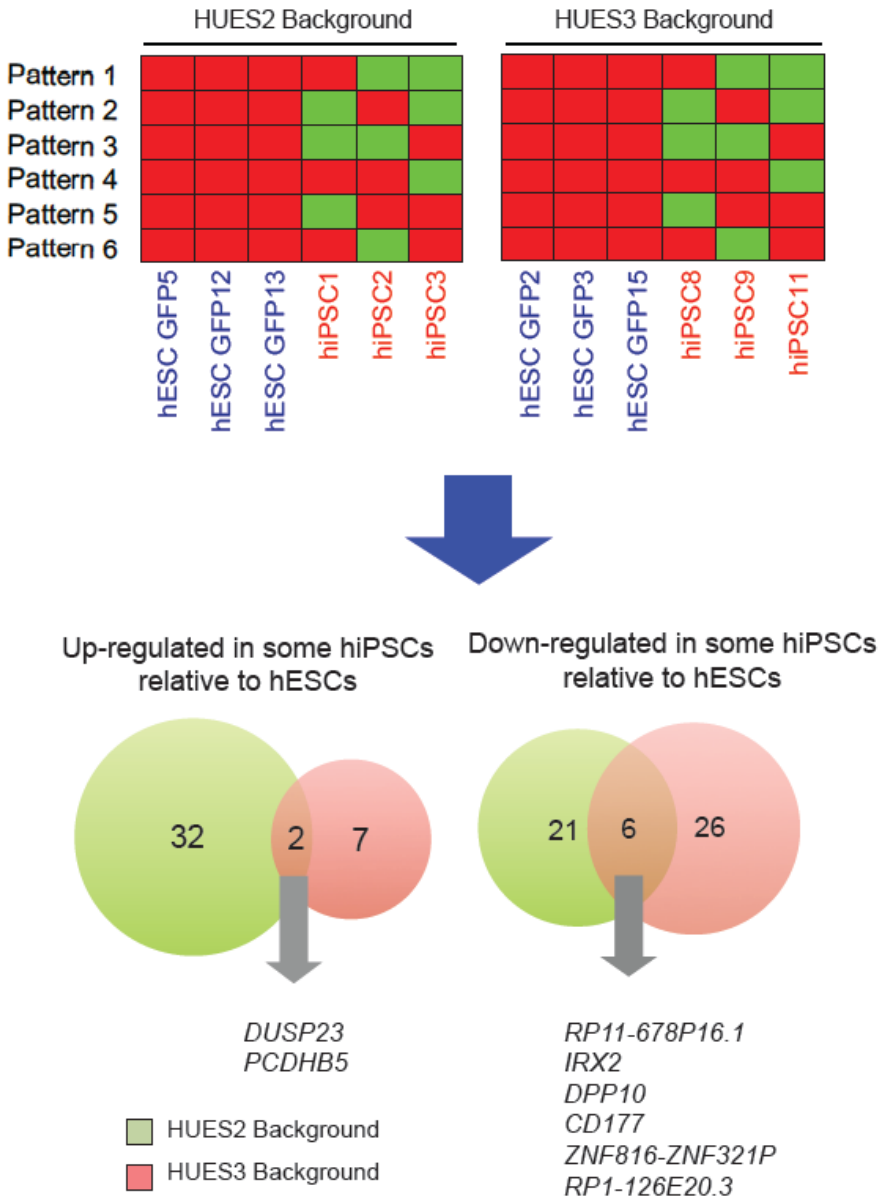


Figure 3.11. Inconsistently differentially expressed genes (iDEGs) between hESC GFP and hiPSC lines.

Schematic for inconsistently differentially expressed genes (iDEGs) that were dysregulated in only a subset of hiPSC lines when compared to hESC GFP lines. Red and green boxes stand for 6 discrete grouping patterns of samples for differential expression analysis, where one or two hiPSC lines are pretended to be a replicate of the hESC lines of the same genetic background. Differentially expressed genes for each pattern were identified and merged within each genetic background and the intersection was taken for the two genetic backgrounds (Venn diagram, green: HUES2, scarlet: HUES3). 8 iDEGs that were common between the two backgrounds are indicated.

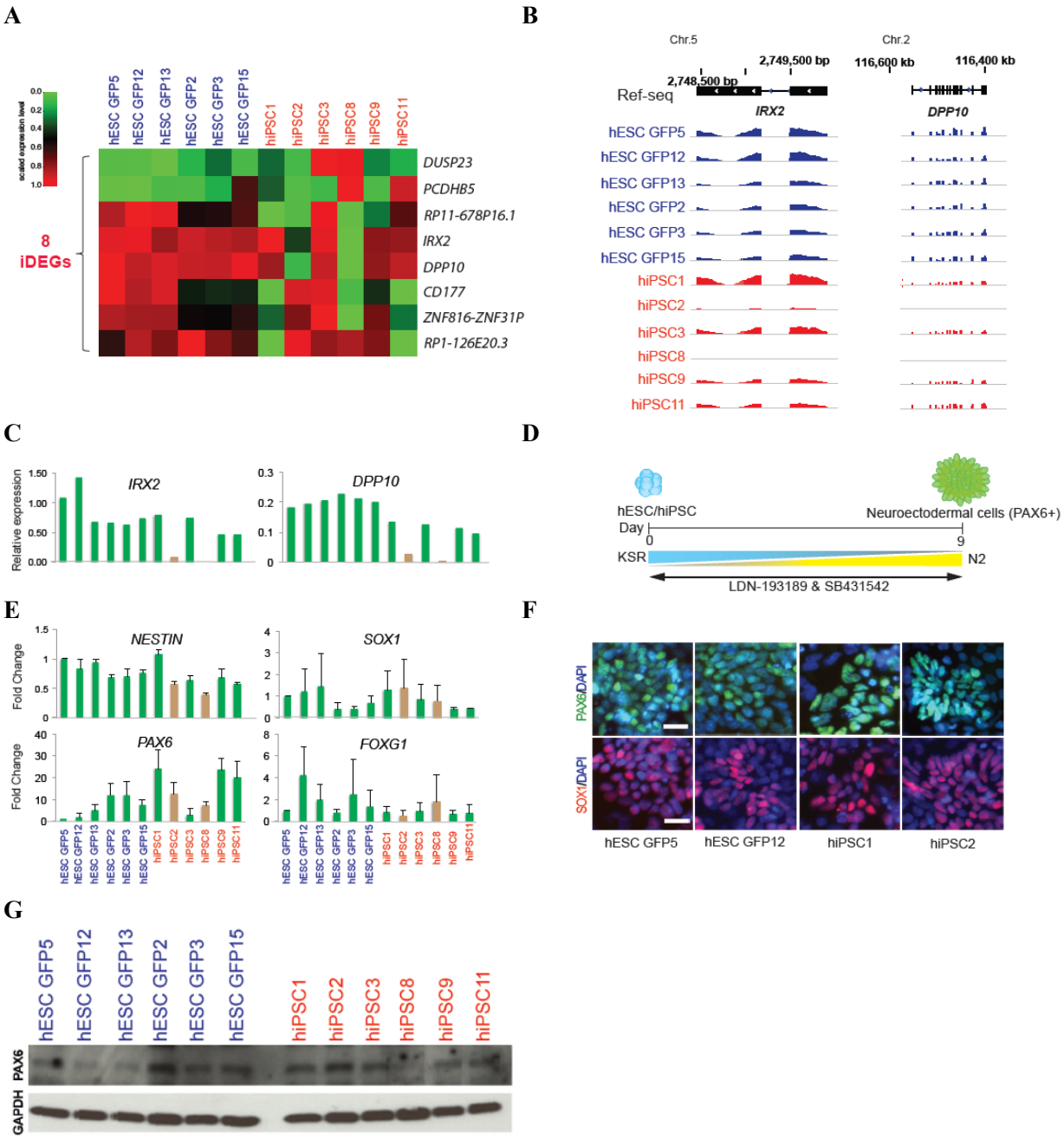


Figure 3.12. Functional tests of inconsistently differentially expressed genes (iDEGs) between hESC and hiPSC lines.

(A) Heatmap of the 8 inconsistently differentially expressed genes (iDEGs) for all isogenic hESCs and hiPSC lines (as defined in Figure 3.11) within each of the two genetic backgrounds at FDR<0.01 and fold change <2 or <1/2. hiPSC lines, red; hESC lines, blue. **(B)** Genome browser images of *IRX2* and *DPP10* RNA-seq reads in hESC GFP and hiPSC lines. **(C)** Expression levels of *IRX2* and *DPP10* by qPCR in each hESC GFP and hiPSC line. Normalized to *ACTB*. Brown bars indicate the hiPSC lines that have undergone aberrant silencing of *IRX2* and *DPP10*. **(D)** Schematic for neural induction using the combination of SB431542, an ALK inhibitor, and LDN-193189, a BMP inhibitor. **(E)** Fold change of the neural markers *NESTIN, SOX1, PAX6*, and *FOXG1* by qPCR in hESC GFP and hiPSC lines with respect to the hESC GFP5 line. Brown bars

Figure 3.12. (Continued) indicate the hiPSC lines that have undergone aberrant silencing of *IRX2* and *DPP10*. Results are shown from three independent experiments. Mean ± s.d. **(F)** Immunofluorescence staining of PAX6 (green) and SOX1 (red) indicates neural differentiation at day6 in hESC GFP and hiPSC lines. DAPI (blue). **(G)** hESC GFP and hiPSC lines were differentiated into neuroectodermal cells and Western blot analysis was used to detect neural differentiation by PAX6 expression at day 6 in each cell line. GAPDH was used as a control.
expression status.

We reasoned that any robust differences identified with our isogenic system should also be seen in hiPSC lines derived from primary somatic cells as well as in datasets published by other investigators. We therefore profiled an additional 18 hESC and 12 hiPSC lines derived from primary fibroblasts by RNA-sequencing; these cell lines have been previously analyzed by Bock et al.[88] through microarray analysis. Although 16 genes were identified as differentially expressed between these non-isogenic hESC and hiPSC lines (Table 3.1), unsupervised clustering of these samples was unable to segregate hiPSCs and hESCs, which is consistent with the aforementioned study (Figure 3.13A). We next asked whether the gene expression signature that distinguishes isogenic hESCs and hiPSCs (Figure 3.8) could separate this extended set of hESCs and hiPSCs. Our stringently defined 49-DEG signature was unable to distinguish the transcriptomes of this extended set of primary fibroblast-derived hiPSC lines (Figure 3.13B).

Finally, we determined the degree of overlap between our DEGs identified within isogenic and non-isogenic hESC/hiPSC lines and two previously reported sets of DEGs[122,130]. Even though we failed to detect consistent gene expression signatures among the various groups of DEGs (Figure 3.13C), two of our 49 DEGs (*MT1E, S100A14*) were shared with Phanstiel et al.[122] whereas two of our 8 iDEGs (*IRX2* and *DPP10*) overlapped with Phanstiel et al.[122] (data not shown). Importantly, there was little to no overlap among DEGs discovered by independent laboratories and these DEGs could not distinguish hiPSC and hESC lines from other data sets (Table 3.2 and Figure 3.14A-G). This supports the view that genetic background and reprogramming method account for the majority of previously reported transcriptional differences between hESCs and hiPSCs. In support of this notion, we found that the DEGs (hESCs vs. hiPSCs) published by Phanstiel et al.[122] were able to distinguish our isogenic cell lines by genetic background rather than cell type (Figure 3.13D). This observation may stem from the fact that hiPSC lines from only one individual were compared to genetically distinct hESC lines in that study. Collectively, these meta-analyses show that no common set of DEGs can be identified when cross-comparing available gene expression datasets. Moreover, some of the published

Table 3.1. Significantly differentially expressed genes between non-isogenic hESC and hiPSC lines.
16 genes were identified as differentially expressed between non-isogenic hESC and hiPSC lines in this study. (FDR<0.01 and fold change >2 or <1/2, details in the Methods section).

POTEC	*APOL1*
EN2	*CFH*
CYP2E1	*PTPRC*
AC004538.3	*GPNMB*
RP11-114H21.2	*ZNF676*
PRAC	*ZNF257*
GPRC5A	*MEG3*
IL1RAP	*AL132709.7*

Table 3.2. Significantly differentially expressed genes (DEGs) from multiple studies.
Definition and number of significantly differentially expressed genes between hESCs and hiPSCs from multiple studies.

Studies	Definition of DEGs	No. of DEGs	Reprog. Method
Choi et al.	Isogenic hESC vs. hiPSC	49	Sendai virus
Choi et al.	Non-isogenic hESC vs. hiPSC	16	Retrovirus
Choi et al.	Non-infected hESC vs. SeV-GFP-infected hESC	63	SeV-GFP virus
Phanstiel et al.	Non-isogenic hESC vs. hiPSC	154	Episomal vector

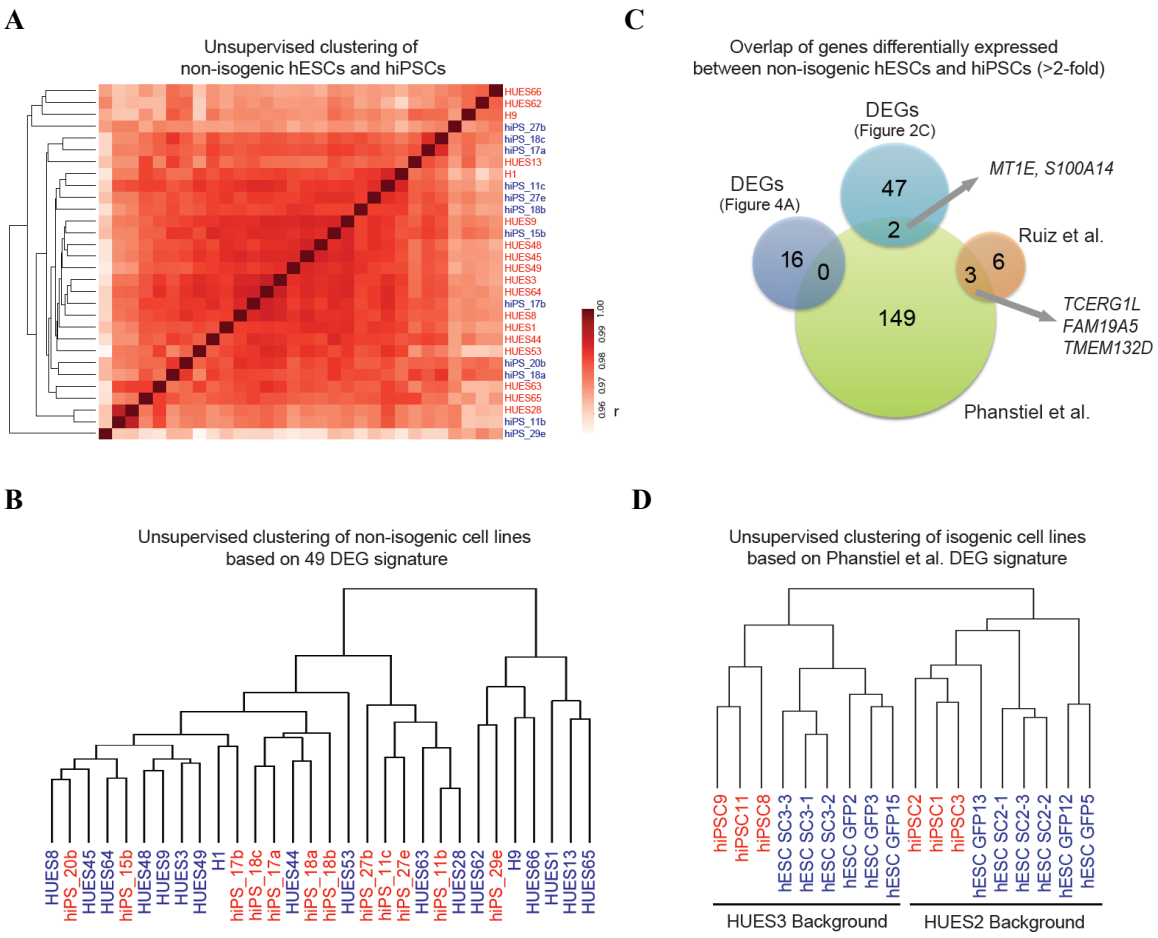


Figure 3.13. Transcriptional analysis separates human pluripotent stem cell genes based on genetic background.

(A) Dendrogram and heatmap for all non-isogenic hESC (blue) and hiPSC (red) lines based on pairwise Pearson correlation (r) on global gene expression levels (log-scaled). **(B)** Dendrogram based on the 49 DEGs identified using isogenic lines in Figure 3.8 for all non-isogenic hESC (blue) and hiPSC (red) lines **(C)** Venn diagram of differentially expressed genes between hESCs and hiPSCs from this study and previously reported reprogramming studies comparing non-isogenic hESCs and hiPSCs. Overlapping genes between DEGs from independent reprogramming experiments are indicated by arrows. **(D)** Dendogram for all isogenic hESC and hiPSC lines using the differentially expressed genes identified in Phanstiel *et al.*[122]hiPSC lines, red; hESC lines, blue.

DEGs may, in fact, segregate pluripotent cell lines based on background rather than cell type of origin.

Discussion

Here, we show that isogenic hESC and hiPSC lines are transcriptionally highly similar to one another, suggesting that genetic background variability accounts for most of the previously reported gene expression differences between hESCs and hiPSCs. This conclusion is particularly relevant in studies where only a limited number of hESC lines or a single iPSC donor individual was used as this may further inflate transcriptional differences[122,153,165,166]. Our finding that a previously reported set of DEGs between 4 hESC lines and hiPSC lines derived from a single individual separated our hESC and hiPSC lines by genetic background rather than cell type further supports this conclusion (Figure 3.13D).

Our study also documents for the first time that a commonly used non-integrating reprogramming method can subtly but stably alter transcriptional patterns in iPSCs (Figure 3.5A and B). However, transcriptional changes introduced by the SeV-based reprogramming approach were not consistent with those of either retroviral or episomal vector-based reprogramming methods, suggesting that each reprogramming system introduces unique transcriptional alterations into iPSCs (Figure 3.14D and G). Whereas the molecular mechanisms of this observation remain to be elucidated, our findings highlight the importance of controlling for the process of iPSC induction when studying transcriptional patterns in iPSCs. Indeed, a recent comparison of hiPSCs generated with different methods showed that hiPSCs derived with integrating vectors (e.g., retroviral transgenes) more often exhibit expression, methylation and differentiation defects compared to hiPSCs produced with non-integrating approaches[128].

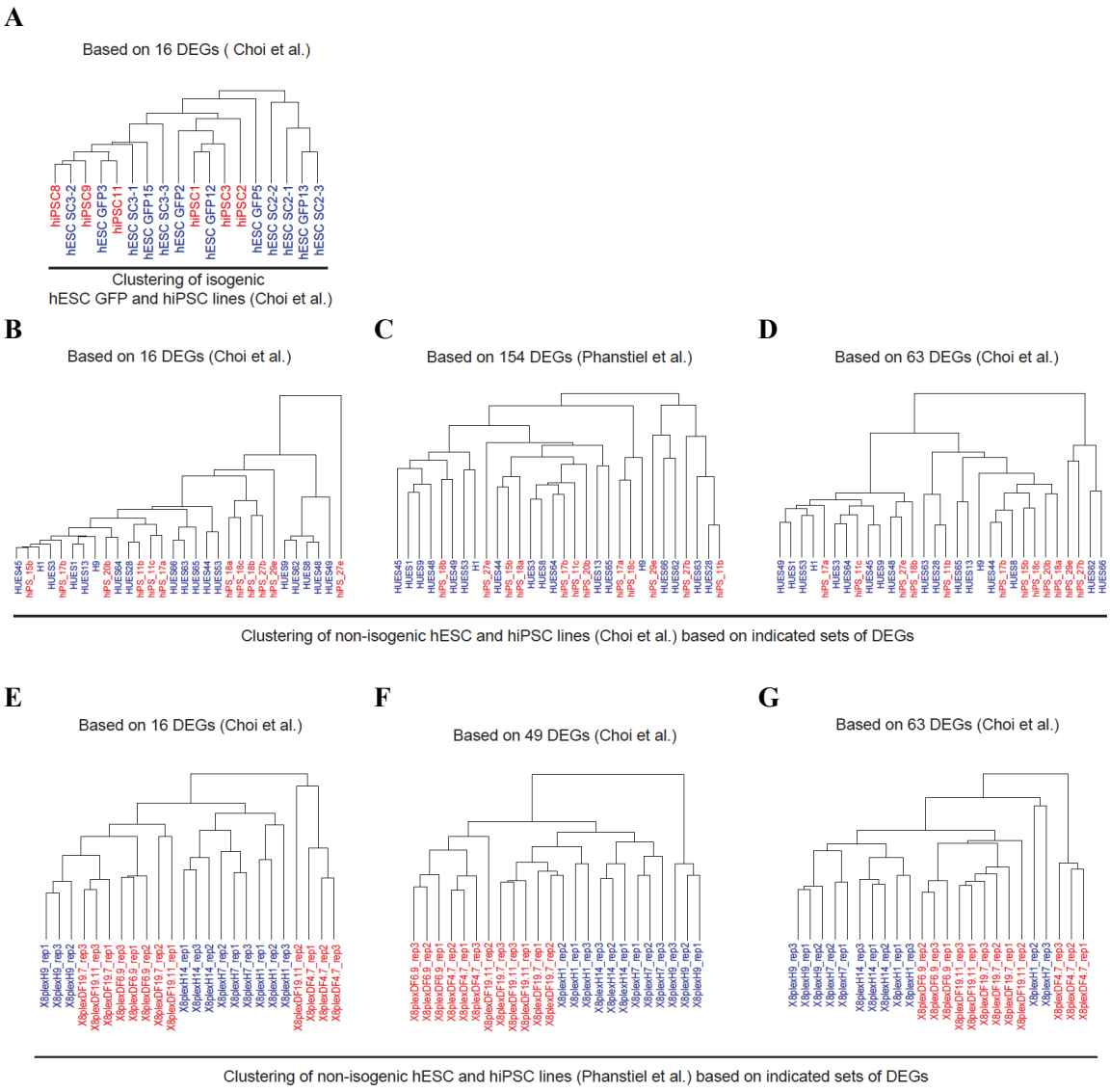


Figure 3.14. Analyses of differentially expressed genes between hESC and hiPSC lines using independent reprogramming data sets.
(A) Dendogram for all isogenic hESC (blue) and hiPSC (red) lines from Choi et al. based on expression levels of the 16 DEGs identified in Table 3.1. **(B-D)** Dendrograms for all non-isogenic hESC (blue) and hiPSC (red) lines (Choi et al.) based on DEGs defined in other studies (see Table 3.2). **(E-G)** Dendrograms for all non-isogenic hESC (blue) and hiPSC (red) lines (Phanstiel et al.) based on DEGs defined in other studies (see Table 3.2).

Although we identified 49 DEGs that could consistently distinguish our hESCs and hiPSCs and 8 iDEGs that were dysregulated in a subset of hiPSC lines, we found no evidence that these genes predicted functional differences (energy production or differentiation potential into the neural lineage). With the exception of two transcripts, *LDHA* and *SLC2A1*, most DEGs were expressed at relatively low levels in hESCs and hiPSCs. The lack of an obvious phenotype could be due to sufficient expression of the analyzed genes in hiPSCs or compensation by posttranscriptional mechanisms, as appears to be the case with LDHA (Figure 3.9D). Alternatively, our metabolic and in vitro-differentiation assays may not have been sensitive enough to detect possible differences. Another possibility is that hiPSCs are distinguished from hESCs by epigenetic differences that do not manifest in the pluripotent state. However, our finding that fibroblast-like cells derived from all examined hESC and hiPSC lines showed no discernable transcriptional differences argues against this explanation (Figure 3.10A and B).

Our results may also have implications for the use of iPSC technology in disease modeling approaches where hiPSC lines from healthy individuals are usually compared to hiPSC lines from affected individuals. Because of the apparent influence of genetic background on gene expression and possibly differentiation patterns of hiPSCs, it will be critical to study a sufficient number of hiPSC lines to detect robust phenotypes; this is particularly relevant in complex diseases where the causal mutation(s) are not known.

Materials and Methods

Cell culture

hESC lines and hiPSC lines were cultured with mouse embryonic fibroblasts (MEFs, Globalstem) pre-plated at 12-15,000 cells/cm2. Medium containing DMEM/F12, 20% knockout serum replacement, 1mM L-glutamine, 100 uM MEM non-essential amino acids, and 0.1 mM

beta-mercaptoethanol was used. 10 ng/ml of FGF-2 was added after sterile filtration and cells were fed daily and passaged weekly using 6U/mL dispase or mechanically.

hiPSC generation

hESC lines were cultured in fibroblast medium without FGF-2 containing DMEM, 10% FBS, 1 mM L-glutamine, 100 uM MEM non-essential amino acids, and 0.1 mM beta-mercaptoethanol, for a week. Cells were passaged three times using 0.25% trypsin and then sorted for hThy1+/hTRA-1-81– populations. Sorted fibroblast-like cells were plated, passaged one more time, and then reprogrammed by using CytoTune®-iPS Sendai Reprogramming Kit (Invitrogen) following manufacturer's instructions.

RNA-sequencing

Undifferentiated hESC/hiPSC cells were sorted for hTRA-1-81+ to control for the homogeneity of cells before RNA extraction. The quality and quantity of total input mRNA was determined on an Agilent BioAnalyzer 2100 using Agilent RNA 6000 Nano kit. One microgram of total RNA from each sample was then used as input for library preparation using Illumina TruSeq RNA Sample Prep Kit, following manufacturer's instructions. Each paired-end library was prepared with an adaptor with unique index sequence. The size profile and quantity of resulting libraries were than determined on the BioAnalyzer 2100 with Agilent High Sensitivity DNA kit. These libraries were then pooled together at equal molar concentration and sequenced on an Illumina HiSeq 2000.

Immunostaining

Immunostaining was performed using the following antibodies: α-hTRA-1-81 (330704, BioLegend), Streptavidin APC (17-4317-82, eBioscience) α-hCD90 (328118, BioLegend), α-

Sendai viral protein (PD029, MBL International), and α-OCT4 (ASK-3006, Applied StemCell), α-PAX6 (Cat. no. PAX6, DSHB), and α-SOX1 (Cat. no. 4194, Cell Signaling).

Lactate production assay

Lactate production assay was done according to Zhong et at.[167] Lactate concentration was determined with the Lactate Assay Kit (BioVision). O.D. was measured at 570nm, 30 min. after addition of substrate.

Glucose uptake assay

The glucose uptake assay was done according to Sebastián et al.[168]. Cells were grown under normal conditions for 24 hr and 100 mM 2-NBDG (Invitrogen) was added to the media for 2 hr. Fluorescence was measured in a FACSCalibur Analyzer (BD).

Neural differentiation

Neural induction was performed as previously reported[163]. Briefly, cells were dissociated to single cells using Accutase and plated on gelatin for 10 minutes to remove MEFs. Non-adherent cells were collected and plated on Geltrex-treated dishes at a density of 150-200k cells per well of a 24-well plate in the presence of MEF-conditioned hESC media containing 10 ng/ml of FGF-2 (Life Tech) and 10 uM of Y-27632 (Tocris). Neural differentiation was initiated when cells were confluent using KSR media containing 820 ml of Knockout DMEM (Life Tech), 150 ml Knockout Serum Replacement (Life Tech), 1 mM L-glutamine (Life Tech), 100 uM MEM non-essential amino acids (Life Tech), and 0.1 mM beta-mercaptoethanol (Life Tech) to inhibit SMAD signaling, 100 nM of LDN-193189 (Cat. no. ab142186, Abcam) and 5 uM of SB431542 (Cat. No. 13031, Cayman Chemical) were added on Days 0 through 9. Cells were fed daily, and N2 media (Life Tech) was added in increasing 25% increments every other day starting on Day 4 (100% N2 on Day 10).

Western blot analysis

For Western blot analysis of PAX6, 10 ug of whole cell lysates was loaded to 4-20% gradient SDS-PAGE gels and then transferred to nitrocellulose membranes (BIO-RAD) by using Trans-Blot® TurboTM Transfer System (BIO-RAD). Blocked membranes were incubated with antibodies against PAX6 (Cat. no. 5790, Abcam) or GAPDH (Cat. no. 2118, Cell Signaling), respectively. For Western blot analysis of LDHA, undifferentiated hESC/hiPSC cells were sorted for hTRA-1-81+ in order to control for the homogeneity of the cells, and then the rest of the procedure ensued as above. LDHA (Cat. no. 2012S, Cell Signaling), β-ACTIN (Cat. no. MA5-15739-HRP, Thermo Scientific).

RNA extraction and qPCR

Total RNA was extracted from differentiating hESC/hiPSC lines using the TRIzol Reagent (Life Tech), and 0.51 ug of RNA was reverse transcribed by High Capacity cDNA Reverse Transcription Kit RT2 first strand kit (ABIQiagen). Primer sequences are provided below. qRT-PCR mixtures were prepared with SYBR Green PCR Master Mix Universal (Applied BiosystemsKapabiosystem) and reactions were done with the Eppendorf Realplex2.

Bioinformatics Methods

RNA-seq data processing

RNA-seq reads were mapped using Bowtie v0.12.7[169] allowing up to 2 mismatches, to the library of human transcriptome sequences obtained from ENSEMBL (GRCh37.67) reference chromosomes, then entries with identical gene symbols were merged. The transcriptome includes both protein-coding genes and non-coding genes such as lincRNAs. EMSAR v1.0 was used to

quantify the expression levels in TPM (transcripts per million) and to infer read counts for individual genes. This EMSAR (Estimation by Mappability-based Segmentation And Reclustering) algorithm we developed improves upon our previous algorithm called NEUMA (Normalization by Expected Uniquely Mappable Area)[170], which performs gene length normalization with adjustment for mappability within each gene. EMSAR utilizes multiply-aligned reads more efficiently, resulting in superb accuracy in quantification of transcript levels (Lee et al, manuscript in submission).

Read coverage plots were generated using the Integrative Genome Browser[171], after rescaling each track by a factor proportional to the total number of uniquely mapped reads, for best visualization.

Identification of differentially expressed genes (DEGs)

To identify differentially expressed genes (DEGs), edgeR[172] v3.4.2 was used after removing genes with a zero total count across all samples. Tag-wise dispersion was computed with the 'movingave' trend option. Final list was obtained by filtering Benjamini-Hochberg-corrected p-values (false discovery rate, FDR) < 0.01 and | log2(fold change) | ≥ 1. HUES2 samples and HUES3 samples were analyzed separately, then commonly up- or down-regulated genes were collected. The disproportionate enrichment of down-regulated genes in hiPSCs was reproduced when another DEG calling method, DESeq[173] v1.8.3, was used.

To identify inconsistent DEGs, one or two of the hiPSCs were pretended to be a replicate of hESC-GFP and a conventional DEG analysis was performed within each genetic background, using DESeq[173] v1.8.3. because of the software's ability to handle comparisons without replicates, a situation that arises when two of three iPSCs are pretended to be hESCs. FDR correction was performed on the entire set of p-values from the six different runs and the final list of genes was selected based on FDR < 0.01 and | log2(fold change) | ≥ 1. Lastly, iDEGs commonly identified in the two genetic backgrounds were collected.

For nonisogenic data, FDR < 0.15 was used instead of FDR < 0.01 to identify sufficient number of DEGs for downstream analysis, since the latter lead to only three DEGs. Total 220 out of 1000 permutations of sample labels resulted in three or more DEGs given FDR < 0.01, suggesting that the even the three DEGs picked up at FDR < 0.01 are likely false positives.

Comparison of distance between cell lines using mean absolute deviation

First, the mean absolute deviation in expression level for each gene was measured between a hiPSC sample and a reference hESC represented by the average across all hESC GFPs for that gene. Then, the average of the deviation over all genes over all hiPSCs was taken. Mean absolute deviation for sample xi is | xi - mean(Y) |, where the random variable Y is over the set of reference values, and the average absolute deviation is a more robust measure than standard deviation. We applied the same procedure but using either genetically matched or unmatched hESC GFPs as a reference hESC set. Changing technical details such as using median absolute deviation or mean squared deviation instead of mean absolute deviation, or scaling gene expression level versus not scaling, did not alter the conclusion.

Clustering of samples using various DEGs

Unsupervised hierarchical clustering was performed using the complete linkage method, after z-score transformation of log(TPM+1) across samples for each gene. Genes with missing or ambiguously matching IDs between published and our gene lists were excluded. The row scaling for Fig. 3a was based on (x-min(x))/(max(x)-min(x)) normalization instead of z-score, for best visualization.

Clustering of samples using global gene expression

Unsupervised clustering for all samples was done using all genes with TPM > 1.5 in at least one of the 33 samples. Pearson correlation (r) between pairs of samples were computed on log(TPM+1) and 1-r was used as the distance metric.

Functional enrichment analysis

Enrichment for Gene Ontology terms in the 49 DEGs was tested using DAVID version 6.7[174,175], with all the non-zero genes used as input for DEG analysis as the background set.

Chapter 4

Discussion and Future Directions

Summary of Thesis Work

The overall goal of my thesis project was to examine the molecular similarities and differences between pluripotent stem cell lines in order to further our understanding of the pluripotent state, as well as examine their potential use in regenerative medicine. By generating low passage, isogenic mouse pluripotent cells, we were able to determine that female pluripotent cell types are globally hypomethylated when compared to their male counterparts due to the increased dosage of the X-linked phosphatase, *Dusp9*. Additionally, by comparing isogenic hESC and hiPSC lines, we found that these two cell types are highly similar to one another at the transcriptional and functional levels, contrary to many previous reports.

Future Studies

Global DNA Hypomethylation In Female Pluripotent Stem Cells

Previous studies have demonstrated that DNA in female ESCs is globally hypomethylated when compared to male ESCs and these differential DNA methylation levels and this phenotype is associated with the number of X chromosomes[83,84]. In this thesis, we identified the Dusp9 phosphatase as an X-linked candidate regulator of DNA methylation and showed that its overexpression in male ESCs results in a decrease of overall DNA methylation levels towards a female ESC-like state. These data suggest that the dosage of *Dusp9* is responsible for the methylation differences observed between male and female ESCs. In order to determine whether *Dusp9* is solely responsible for the methylation differences identified between male and female ESCs, it would be valuable to generate female ESCs that are heterozygous for *Dusp9* (*Dusp9*$^{+/-}$). Removing one *Dusp9* allele from female ESCs would create female ESCs that are genetically similar to male wild-type ESCs with respect to *Dusp9* dosage. Once these cells have been

passaged, the amount of DNA methylation in female *Dusp9*$^{+/-}$ ESCs would be measured and compared to male ESCs. This analysis would complement the overexpression experiments and allow us to determine whether *Dusp9* is not only sufficient but also required for female hypomethylation.

Dusp9 was recently suggested to inhibit MAPK signaling by dephosphorylating one of its downstream kinases, ERK1/2[141]. ERK1/2 is constitutively expressed in many cell types and plays a key role in the regulation of cell proliferation and differentiation, depending on cellular context. When MAPK signaling is active, the kinase MEK phosphorylates ERK1/2 (p-ERK1/2), leading to its ability to phosphorylate downstream effectors. Dusp9 expression was recently shown to decrease the level of p-ERK1/2, thus attenuating overall MAPK signaling[141]. We demonstrated that treating male ESCs with a MEK inhibitor results in a dramatic reduction of DNA methylation levels, similar to those found in female ESC, supporting the link between MAPK signaling, *Dusp9* dosage, and DNA methylation. Despite this intriguing connection, it is still important to understand precisely how this signaling cascade results in a reduction of DNA methylation levels. First, in order to determine whether Dusp9 works directly on ERK1/2 to reduce DNA methylation levels, *ERK1/2* double knockout ESCs (*ERK1/2* DKO) would need to be analyzed. We would expect to see that the dosage of *Dusp9* has no effect on DNA methylation levels in male or female *ERK1/2* DKO ESCs, which are viable (S. Meloche, pers. communication). However, if the levels of DNA methylation remain unchanged in *ERK1/2* DKO ESCs when compared to wild-type, this would suggest that Dusp9 is acting independently of ERK1/2 and further studies would need to be performed in order to determine Dusp9's downstream effector.

In order to influence DNA methylation levels, MAPK signaling must ultimately affect the levels of the Dnmts, the family of proteins that are responsible for methylating DNA. Consistent with this, we saw a decrease in Dnmt3a/b at the protein level when *Dusp9* was overexpressed in male ESCs. These results suggest that MAPK signaling can decrease overall Dnmt levels, however how this may occur is poorly understood. Since MAPK signaling works

via phosphorylation of downstream effectors, it would be worthwhile to perform phospho-proteomic analyses on Dnmt3a/b in wild-type and *Dusp9* overexpression male ESCs. First, in order to determine whether Dnmts can be regulated at the posttranscriptional level via phosphorylation, Dnmt3a/b in wild-type and *Dusp9* overexpressing male ESCs would need to be analyzed for their phosphorylation status. We would expect to identify phosphorylation sites on Dnmt3a/b that are differentially phosphorylated between wild-type and *Dusp9* overexpression male ESCs. This could help in identifying the residues, whose phosphorylation is associated with Dnmt3a/b stability. Inhibition of phosphorylation by introducing mutations at the differentially phosphorylated residues might affect the stability of Dnmt3a/3b, leading to the change in cellular levels of Dnmt3a/b. This would also help in determining the potential kinase responsible for Dnmt3a/b stability. However, if phosphorylation states of Dnmt3a/b remain unchanged between wild-type and *Dusp9* overexpressing male ESCs, it would suggest that the post-transcriptional regulation of Dnmt3a/b is not mediated by phosphorylation and further investigation on the regulation of Dnmt3a/b with respect to sex of cell lines will be warranted.

Dusp9-null female ESCs^{-} undergo spontaneous differentiation (data not shown), which is consistent with the previous study that knock-down of *Dusp9* in ESCs resulted in differentiation[141]. Interestingly, heterozygous female ESCs formed both undifferentiated and differentiated colonies (data not shown). Since female ESCs with two active X chromosomes spontaneously lose one X chromosome upon extended culture, it is possible that undifferentiated and differentiated colonies of heterozygous female ESCs are caused by the loss of *Dusp9* knock-out allele and wild-type allele, respectively. This observation further suggests that the differentiation phenotype might be due to aberrant accumulation of global DNA methylation. In order to test this hypothesis, it would be necessary to determine whether *Dusp9*-null male ESCs exhibit higher global methylation level compared to wild-type ESCs and suppression of global methylation level by depleting Dnmt1 rescues the differentiation phenotype in *Dusp9*-null male ESCs. However, if the global methylation level in *Dusp9*-null male ESCs remains unchanged or

Dnmt1 knock-down cannot rescue the the differentiation phenotype of *Dusp9*, it would suggest that DNA methylation is not the reason why *Dusp9*-null male ESCs exit pluripotency. Alternatively, it is possible that upregulation of MAPK pathway in *Dusp9*-null male ESCs causes the cells to differentiate. To test that hypothesis, it would be needed to determine if MAPK pathway inhibition by small molecules can rescue the differentiation phenotype.

During female development, both X chromosomes become transiently active in the ICM cells of the preimplantation blastocyst. Likewise, female ESCs generated from the ICM cells also maintain two active X chromosomes in culture. Recent data has suggested that this increase in X dosage results in a transient developmental delay[85], however the biological implications of differential X gene dosage between male and female embryos remain unknown. Thus, it would be important to determine whether the differences in DNA methylation seen *in vitro* between male and female ESCs are also present during *in vivo* development.

Comparison of hiPSCs and hESCs

The question of whether hESCs and hiPSCs are equivalent has been an actively studied topic since hiPSCs were first generated. Many groups have reported on the potential transcriptional and functional differences between hESCs and hiPSCs, arriving at different conclusions regarding their equivalency. Our lab previously showed that the genetic background of mouse ESCs can have a substantial impact on transcriptional patterns in ESCs, indicating that it is important to control for genetic background when comparing different cell lines. Based on these observations, we hypothesized that the variation previously reported between hESCs and hiPSCs could also be due, in part, to the genetic background of the cell lines as opposed to inherent differences between hESCs and hiPSCs. Therefore, we generated and compared genetically matched hESC and hiPSC lines by differentiating hESC lines to fibroblast-like cells and generating hiPSCs from these differentiated cells. By using this approach to compare and contrast hESCs and hiPSCs, we concluded that that integration-free reprogramming generates

hiPSCs that are transcriptionally equivalent to hESCs. This finding suggests that the previously reported differences between hESCs and hiPSCs may be due to the different genetic backgrounds of the cell lines used. While our approach was able to control for genetic background when comparing hESCs and hiPSCs, we did not fully characterize the fibroblast-like cells generated from hESCs. Although our transcriptional analysis showed that the fibroblast-like cells are highly similar to dermal fibroblasts, the fibroblast-like cells might not be epigenetically equivalent to adult fibroblasts. It is possible that the fibroblast-like cells we generated retain some of the epigenetic marks of their parental hESCs, which might mask some differences between hiPSCs and hESCs. Therefore, in-depth epigenetic characterization of the fibroblast-like cells is warranted to rule out this possibility. For example, DNA methylation analyses on pluripotency-associated genes and fibroblast-specific genes will be informative to confirm that some epigenetic marks were fully erased in the differentiated fibroblast-like cells that were used to generate hiPSCs.

We sought to evaluate the functionality of hiPSC lines by comparing hiPSC lines to genetically matched hESC lines because hESCs are considered to be the gold standard for human pluripotent stem cells. By *in vitro* differentiating genetically matched hESC and hiPSC lines into fibroblasts and neuroectodermal precursor cells, we found that these two pluripotent cell types share the same degree of differentiation capabilities. However, the inability to generate chimeras from hESCs prevents us from confirming whether hESCs can give rise to all cell types of the body *in vivo*, which constitutes the strictest definition of pluripotency. In contrast, mouse ESCs have been demonstrated to be pluripotent using chimera formation and tetraploid complementation assays. Each of these assays demonstrates that ESCs in question are competent to form every cell type of the body, thus unequivocally demonstrating full pluripotentiality. Recently, one group has reported that hESCs can contribute to the formation of interspecies chimera between human and mouse by injecting a mouse blastocyst with hESCs that were coaxed to more closely resemble mouse ESCs at the transcriptional level[113]. However, this result has not

yet been reproduced because of technical and ethical issues. Therefore, it would be valuable to develop universal standards that can precisely assess human pluripotency so we can confidently assess the quality of hiPSCs and transition reprogramming technology into the clinic.

Outlook

The ability of pluripotent stem cells that can self-renew indefinitely *in vitro* and be differentiated into any lineage promises enormous potential for basic research and regenerative medicine. Although the field has moved closer to the realization of that promise, lots of open questions have yet to be answered. The work presented in this thesis will lay the groundwork for future research to answer those questions.

Chapter 5

References

1. Stevens, L. C. & Little, C. C. Spontaneous Testicular Teratomas in an Inbred Strain of Mice. *Proc. Natl. Acad. Sci. U.S.A.* **40,** 1080–1087 (1954).

2. KLEINSMITH, L. J. & PIERCE, G. B. MULTIPOTENTIALITY OF SINGLE EMBRYONAL CARCINOMA CELLS. *Cancer Res.* **24,** 1544–1551 (1964).

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

4. Mintz, B. & Illmensee, K. Normal genetically mosaic mice produced from malignant teratocarcinoma cells. *Proc. Natl. Acad. Sci. U.S.A.* **72,** 3585–3589 (1975).

5. Papaioannou, V. E., McBurney, M. W., Gardner, R. L. & Evans, M. J. Fate of teratocarcinoma cells injected into early mouse embryos. *Nature* **258,** 70–73 (1975).

6. Martin, G. R. 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 (1981).

7. Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309,** 255–256 (1984).

8. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 8424–8428 (1993).

9. COLLINS, D. H. & PUGH, R. C. CLASSIFICATION AND FREQUENCY OF TESTICULAR TUMOURS. *Br J Urol* **36,** SUPPL:1–11 (1964).

10. Stevens, L. C. Origin of testicular teratomas from primordial germ cells in mice. *J. Natl. Cancer Inst.* **38,** 549–552 (1967).

11. Resnick, J., Bixler, L., Cheng, L. & Donovan, P. J. Long-term proliferation of mouse primordial germ cells in culture. (1992).

12. Matsui, Y., Zsebo, K. & Hogan, B. L. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70,** 841–847 (1992).

13. Labosky, P. A., Barlow, D. P. & Hogan, B. L. Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. *Development* **120,** 3197–3204 (1994).

14. Stewart, C. L., Gadi, I. & Bhatt, H. Stem cells from primordial germ cells can reenter the germ line. *Dev. Biol.* **161,** 626–628 (1994).

15. Evans, M. J. The isolation and properties of a clonal tissue culture strain of pluripotent mouse teratoma cells. *J Embryol Exp Morphol* **28,** 163–176 (1972).

16. Smith, T. A. & Hooper, M. L. Medium conditioned by feeder cells inhibits the differentiation of embryonal carcinoma cultures. *Exp. Cell Res.* **145,** 458–462 (1983).

17. Smith, A. G. & Hooper, M. L. Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Dev. Biol.* **121,** 1–9 (1987).

18. Smith, A. G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336,** 688–690 (1988).

19. Williams, R. L. *et al.* Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336,** 684–687 (1988).

20. Davis, S. *et al.* LIFR beta and gp130 as heterodimerizing signal transducers of the tripartite CNTF receptor. *Science* **260,** 1805–1808 (1993).

21. Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes & Development* **12,** 2048–2060 (1998).

22. Hall, J. *et al.* Oct4 and LIF/Stat3 additively induce Krüppel factors to sustain embryonic stem cell self-renewal. *Cell Stem Cell* **5,** 597–609 (2009).

23. Niwa, H., Ogawa, K., Shimosato, D. & Adachi, K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* **460,** 118–122 (2009).

24. Tai, C.-I. & Ying, Q.-L. Gbx2, a LIF/Stat3 target, promotes reprogramming to and retention of the pluripotent ground state. *J. Cell. Sci.* **126,** 1093–1098 (2013).

25. Martello, G., Bertone, P. & Smith, A. Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *EMBO J.* **32,** 2561–2574 (2013).

26. Ye, S., Li, P., Tong, C. & Ying, Q.-L. Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. *EMBO J.* **32,** 2548–2560 (2013).

27. Ying, Q.-L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115,** 281–292 (2003).

28. Feng, X.-H. & Derynck, R. Specificity and versatility in tgf-beta signaling through Smads. *Annu. Rev. Cell Dev. Biol.* **21,** 659–693 (2005).

29. Burdon, T., Stracey, C., Chambers, I., Nichols, J. & Smith, A. Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Dev. Biol.* **210,** 30–43 (1999).

30. Fukada, T. *et al.* Two signals are necessary for cell proliferation induced by a cytokine receptor gp130: involvement of STAT3 in anti-apoptosis. *Immunity* **5,** 449–460 (1996).

31. MacDonald, B. T., Tamai, K. & He, X. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Developmental Cell* **17,** 9–26 (2009).

32. Yi, F. *et al.* Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. *Nature Cell Biology* **13,** 762–770 (2011).

33. Wray, J. *et al.* Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nature Cell Biology* **13,** 838–845 (2011).

34. Faunes, F. *et al.* A membrane-associated β-catenin/Oct4 complex correlates with ground-state pluripotency in mouse embryonic stem cells. *Development* **140,** 1171–1183 (2013).

35. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).

36. Leitch, H. G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nature Structural and Molecular Biology* **20,** 311–316 (2013).

37. Yamaji, M. *et al.* PRDM14 Ensures Naive Pluripotency through Dual Regulation of Signaling and Epigenetic Pathways in Mouse Embryonic Stem Cells. *Cell Stem Cell* **12,** 368–382 (2013).

38. Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149,** 590–604 (2012).

39. Wray, J., Kalkan, T. & Smith, A. G. The ground state of pluripotency. *Biochem. Soc. Trans* **38,** 1027 (2010).

40. Kanda, A., Sotomaru, Y., Shiozawa, S. & Hiyama, E. Establishment of ES cells from inbred strain mice by dual inhibition (2i). *J. Reprod. Dev.* **58,** 77–83 (2012).

41. Fuda, N. J., Ardehali, M. B. & Lis, J. T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461,** 186–192 (2009).

42. Chen, X. *et al.* Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell* **133,** 1106–1117 (2008).

43. Loh, Y.-H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38,** 431–440 (2006).

44. Young, R. A. Control of the Embryonic Stem Cell State. *Cell* **144,** 940–954 (2011).

45. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441,** 349–353 (2006).

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

47. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141,** 432–445

(2010).

48. Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biology* **9,** 625–635 (2007).

49. Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138,** 722–737 (2009).

50. Schwarz, B. A., Bar-Nur, O., Silva, J. C. R. & Hochedlinger, K. Nanog is dispensable for the generation of induced pluripotent stem cells. *Curr. Biol.* **24,** 347–350 (2014).

51. Carter, A. C., Davis-Dusenbery, B. N., Koszka, K., Ichida, J. K. & Eggan, K. Nanog-independent reprogramming to iPSCs with canonical factors. *Stem Cell Reports* **2,** 119–126 (2014).

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

53. Martello, G. *et al.* Esrrb Is a Pivotal Target of the Gsk3/Tcf3 Axis Regulating Embryonic Stem Cell Self-Renewal. *Cell Stem Cell* **11,** 491–504 (2012).

54. Cartwright, P. *et al.* LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development* **132,** 885–896 (2005).

55. Kim, J. *et al.* A Myc Network Accounts for Similarities between Embryonic Stem and Cancer Cell Transcription Programs. *Cell* **143,** 313–324 (2010).

56. Wernig, M., Meissner, A., Cassady, J. P. & Jaenisch, R. c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell Stem Cell* **2,** 10–12 (2008).

57. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4,** 487–492 (2009).

58. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230–1234 (2007).

59. Hayashi, K., Lopes, S. M. C. de S., Tang, F. & Surani, M. A. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic

states. *Cell Stem Cell* **3,** 391–401 (2008).

60. Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135,** 909–918 (2008).

61. Hackett, J. A. & Surani, M. A. Regulatory Principles of Pluripotency: From the Ground State Up. *Cell Stem Cell* **15,** 416–430 (2014).

62. Hackett, J. A. & Surani, M. A. DNA methylation dynamics during the mammalian life cycle. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **368,** 20110328 (2013).

63. Smith, Z. D. *et al.* DNA methylation dynamics of the human preimplantation embryo. *Nature* 1–18 (2014). doi:10.1038/nature13581

64. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484,** 339–344 (2012).

65. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447,** 425–432 (2007).

66. Surani, M. A., Hayashi, K. & Hajkova, P. Genetic and epigenetic regulators of pluripotency. *Cell* **128,** 747–762 (2007).

67. Siegfried, Z. & Simon, I. DNA methylation and gene expression. *Wiley Interdiscip Rev Syst Biol Med* **2,** 362–371 (2010).

68. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14,** 204–220 (2013).

69. Denis, H., Ndlovu, '. N. & Fuks, F. Regulation of mammalian DNA methyltransferases: a route to new mechanisms. *EMBO reports* **12,** 647–656 (2011).

70. Goll, M. G. & Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74,** 481–514 (2005).

71. Avvakumov, G. V. *et al.* Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* **455,** 822–825 (2008).

72. Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A. & Cheng, X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* **449,** 248–251 (2007).

73. Hajkova, P. *et al.* Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329,** 78–82 (2010).

74. Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* **117,** 15–23 (2002).

75. Tada, T. *et al.* Epigenotype switching of imprintable loci in embryonic germ cells. *Dev. Genes Evol.* **207,** 551–561 (1998).

76. Shovlin, T. C., Durcova-Hills, G., Surani, A. & McLaren, A. Heterogeneity in imprinted methylation patterns of pluripotent embryonic germ cells derived from pre-migratory mouse germ cells. *Dev. Biol.* **313,** 674–681 (2008).

77. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454,** 766–770 (2008).

78. Habibi, E. *et al.* Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells. *Cell Stem Cell* 1–10 (2013). doi:10.1016/j.stem.2013.06.002

79. Ficz, G. *et al.* FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency. *Cell Stem Cell* 1–9 (2013). doi:10.1016/j.stem.2013.06.004

80. Hackett, J. A. *et al.* Synergistic Mechanisms of DNA Demethylation during Transition to Ground-State Pluripotency. *Stem Cell Reports* **1,** 518–531 (2013).

81. Mak, W. *et al.* Reactivation of the paternal X chromosome in early mouse embryos. *Science* **303,** 666–669 (2004).

82. Augui, S., Nora, E. P. & Heard, E. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat. Rev. Genet.* **12,** 429–442 (2011).

83. Ooi, S. *et al.* Dynamic instability of genomic methylation patterns in pluripotent stem cells. *Epigenetics Chromatin* **3,** 17 (2010).

84. Zvetkova, I. *et al.* Global hypomethylation of the genome in XX embryonic stem cells. *Nat Genet* **37,** 1274–1279 (2005).

85. Schulz, E. G. *et al.* The Two Active X Chromosomes in Female ESCs Block Exit from the Pluripotent State by Modulating the ESC Signaling Network. *Cell Stem Cell* **14,** 203–216 (2014).

86. Thomson, J., Itskovitz-Eldor, J. & Shapiro, S. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science* (1998).

87. Robinton, D. A. & Daley, G. Q. The promise of induced pluripotent stem cells in research and therapy. *Nature* **481,** 295–305 (2012).

88. Bock, C. *et al.* Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144,** 439–452 (2011).

89. Tachibana, M. *et al.* Human Embryonic Stem Cells Derived by Somatic Cell Nuclear Transfer. *Cell* 1–11 (2013). doi:10.1016/j.cell.2013.05.006

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

91. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).

92. Yu, J. *et al.* Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science* **318,** 1917–1920 (2007).

93. Maherali, N. *et al.* A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell Stem Cell* **3,** 340–345 (2008).

94. Loh, Y.-H. *et al.* Generation of induced pluripotent stem cells from human blood. *Blood* **113,** 5476–5479 (2009).

95. Woltjen, K. *et al.* piggyBac transposition reprograms fibroblasts to induced pluripotent

stem cells. *Nature* **458,** 766–770 (2009).

96. Yu, J. *et al.* Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences. *Science* **324,** 797–801 (2009).

97. Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. & Hasegawa, M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci.* **85,** 348–362 (2009).

98. Daheron, L. *et al.* LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. *Stem Cells* **22,** 770–778 (2004).

99. Xu, R.-H. *et al.* BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nature Biotechnology* **20,** 1261–1264 (2002).

100. Zhang, P. *et al.* Short-term BMP-4 treatment initiates mesoderm induction in human embryonic stem cells. *Blood* **111,** 1933–1941 (2008).

101. Xu, R.-H. *et al.* Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat Methods* **2,** 185–190 (2005).

102. Levine, A. J. & Brivanlou, A. H. GDF3, a BMP inhibitor, regulates cell fate in stem cells and early embryos. *Development* **133,** 209–216 (2006).

103. James, D., Levine, A. J., Besser, D. & Hemmati-Brivanlou, A. TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132,** 1273–1282 (2005).

104. Beattie, G. M. *et al.* Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers. *Stem Cells* **23,** 489–495 (2005).

105. Vallier, L., Alexander, M. & Pedersen, R. A. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J. Cell. Sci.* **118,** 4495–4509 (2005).

106. Amit, M. *et al.* Clonally derived human embryonic stem cell lines maintain pluripotency

and proliferative potential for prolonged periods of culture. *Dev. Biol.* **227,** 271–278 (2000).

107. Xu, C. *et al.* Feeder-free growth of undifferentiated human embryonic stem cells. *Nature Biotechnology* **19,** 971–974 (2001).

108. Eiselleova, L. *et al.* A complex role for FGF-2 in self-renewal, survival, and adhesion of human embryonic stem cells. *Stem Cells* **27,** 1847–1857 (2009).

109. Buecker, C. *et al.* A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. *Cell Stem Cell* **6,** 535–546 (2010).

110. Hanna, J. *et al.* Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 9222–9227 (2010).

111. Wang, W. *et al.* Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 18283–18288 (2011).

112. Chan, Y.-S. *et al.* Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell* **13,** 663–675 (2013).

113. Gafni, O. *et al.* nature12745. *Nature* **504,** 282–286 (2013).

114. Takashima, Y. *et al.* Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. *Cell* **158,** 1254–1269 (2014).

115. Ware, C. B. *et al.* Derivation of naive human embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 4484–4489 (2014).

116. Theunissen, T. W. *et al.* Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15,** 471–487 (2014).

117. Huang, K., Maruyama, T. & Fan, G. The naive state of human pluripotent stem cells: a synthesis of stem cell and preimplantation embryo transcriptome analyses. *Cell Stem Cell* **15,** 410–415 (2014).

118. Stadtfeld, M. *et al.* Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* **465,** 175–181 (2010).

119. Chin, M. H. *et al.* Induced Pluripotent Stem Cells and Embryonic Stem Cells Are Distinguished by Gene Expression Signatures. *Cell Stem Cell* **5,** 111–123 (2009).

120. Guenther, M. G. *et al.* Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell* **7,** 249–257 (2010).

121. Newman, A. M. & Cooper, J. B. Lab-Specific Gene Expression Signatures in Pluripotent Stem Cells. *Cell Stem Cell* **7,** 258–262 (2010).

122. Phanstiel, D. H. *et al.* Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat Methods* **8,** 821–827 (2011).

123. Doi, A. *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* **41,** 1350–1353 (2009).

124. Ohi, Y. *et al.* Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nature Cell Biology* **13,** 541–549 (2011).

125. Sommer, C. A. *et al.* Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells* **27,** 543–549 (2009).

126. Stadtfeld, M. *et al.* Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. *Nat Genet* 1–10 (2012). doi:10.1038/ng.1110

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

128. Koyanagi-Aoi, M. *et al.* Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. *Proc. Natl. Acad. Sci. U.S.A.* (2013). doi:10.1073/pnas.1319061110

129. Schlaeger, T. M. *et al.* A comparison of non-integrating reprogramming methods. *Nature Biotechnology* 1–8 (2014). doi:10.1038/nbt.3070

130. Ruiz, S. *et al.* Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 16196–16201 (2012).

131. Huang, K. *et al.* A Panel of CpG Methylation Sites Distinguishes Human Embryonic Stem Cells and Induced Pluripotent Stem Cells. *Stem Cell Reports* **2,** 36–43 (2014).

132. Hayashi, K. & Surani, M. A. Resetting the epigenome beyond pluripotency in the germline. *Cell Stem Cell* **4,** 493–498 (2009).

133. Tada, M., Tada, T., Lefebvre, L., Barton, S. C. & Surani, M. A. Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *EMBO J.* **16,** 6510–6520 (1997).

134. Tada, M., Takahama, Y., Abe, K. & Nakatsuji, N. Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Curr. Biol.* (2001).

135. Piccolo, F. M. *et al.* Different Roles for Tet1 and Tet2 Proteins in Reprogramming-Mediated Erasure of Imprints Induced by EGC Fusion. *Molecular Cell* 1–11 (2013). doi:10.1016/j.molcel.2013.01.032

136. Okashita, N. *et al.* PRDM14 promotes active DNA demethylation through the Ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells. *Development* **141,** 269–280 (2014).

137. Sharov, A. A. *et al.* Transcriptome analysis of mouse stem cells and early embryos. *PLoS Biol.* **1,** E74 (2003).

138. Sharova, L. V. *et al.* Global gene expression profiling reveals similarities and differences among mouse pluripotent stem cells of different origins and strains. *Dev. Biol.* **307,** 446–459 (2007).

139. Hatano, S.-Y. *et al.* Pluripotential competence of cells associated with Nanog activity.

Mech. Dev. **122,** 67–79 (2005).

140. Grabole, N. *et al.* embor201367a. *EMBO reports* 1–9 (2013). doi:10.1038/embor.2013.67

141. Li, Z. *et al.* BMP4 Signaling Acts via dual-specificity phosphatase 9 to control ERK activity in mouse embryonic stem cells. *Cell Stem Cell* **10,** 171–182 (2012).

142. Beard, C., Hochedlinger, K., Plath, K., Wutz, A. & Jaenisch, R. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* **44,** 23–28 (2006).

143. Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463,** 1101–1105 (2010).

144. Schulz, E. G. & Heard, E. Role and control of X chromosome dosage in mammalian development. *Current Opinion in Genetics & Development* **23,** 109–115 (2013).

145. Anguera, M. C. *et al.* Molecular signatures of human induced pluripotent stem cells highlight sex differences and cancer genes. *Cell Stem Cell* **11,** 75–90 (2012).

146. Yildirim, E. *et al.* Xist RNA Is a Potent Suppressor of Hematologic Cancer in Mice. *Cell* **152,** 727–742 (2013).

147. Chen, R. Z., Pettersson, U., Beard, C., Jackson-Grusby, L. & Jaenisch, R. DNA hypomethylation leads to elevated mutation rates. *Nature* **395,** 89–93 (1998).

148. Gaudet, F. *et al.* Induction of tumors in mice by genomic hypomethylation. *Science* **300,** 489–492 (2003).

149. Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300,** 131–135 (2003).

150. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research* **33,** 5868–5877 (2005).

151. Park, I.-H. *et al.* Disease-specific induced pluripotent stem cells. *Cell* **134,** 877–886 (2008).

152. Chin, M. H., Pellegrini, M., Plath, K. & Lowry, W. E. Molecular Analyses of Human Induced Pluripotent Stem Cells and Embryonic Stem Cells. *Cell Stem Cell* **7,** 263–269 (2010).

153. Teichroeb, J. H., Betts, D. H. & Vaziri, H. Suppression of the Imprinted Gene NNAT and X-Chromosome Gene Activation in Isogenic Human iPS Cells. *PLoS ONE* **6,** e23436 (2011).

154. Cowan, C. A. *et al.* Derivation of embryonic stem-cell lines from human blastocysts. *N. Engl. J. Med.* **350,** 1353–1356 (2004).

155. Tchieu, J. *et al.* Female Human iPSCs Retain an Inactive X Chromosome. *Cell Stem Cell* **7,** 329–342 (2010).

156. Everse, J. & Kaplan, N. O. Lactate dehydrogenases: structure and function. *Adv. Enzymol. Relat. Areas Mol. Biol.* **37,** 61–133 (1973).

157. Mueckler, M. *et al.* Sequence and structure of a human glucose transporter. *Science* **229,** 941–945 (1985).

158. Folmes, C. D. L. *et al.* Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. *Cell Metab.* **14,** 264–271 (2011).

159. Cohen, D. R., Cheng, C. W., Cheng, S. H. & Hui, C. C. Expression of two novel mouse Iroquois homeobox genes during neurogenesis. *Mech. Dev.* **91,** 317–321 (2000).

160. Girirajan, S. *et al.* Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *Am. J. Hum. Genet.* **92,** 221–237 (2013).

161. Marshall, C. R. *et al.* Structural Variation of Chromosomes in Autism Spectrum Disorder. *The American Journal of Human Genetics* **82,** 477–488 (2008).

162. Matsumoto, K. *et al.* The prepattern transcription factor Irx2, a target of the FGF8/MAP kinase cascade, is involved in cerebellum formation. *Nat Neurosci* **7,** 605–612 (2004).

163. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by

dual inhibition of SMAD signaling. *Nature Biotechnology* **27,** 275–280 (2009).

164. Zhang, X. *et al.* Pax6 Is a Human Neuroectoderm Cell Fate Determinant. *Cell Stem Cell* **7,** 90–100 (2010).

165. Abyzov, A. *et al.* Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* **492,** 438–442 (2012).

166. Loewer, S. *et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet* **42,** 1113–1117 (2010).

167. Zhong, L. *et al.* The Histone Deacetylase Sirt6 Regulates Glucose Homeostasis via Hif1a. *Cell* **140,** 280–293 (2010).

168. Sebastián, C. *et al.* The Histone Deacetylase SIRT6 Is a Tumor Suppressor that Controls Cancer Metabolism. *Cell* **151,** 1185–1199 (2012).

169. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10,** R25 (2009).

170. Lee, S. *et al.* Accurate quantification of transcriptome from RNA-Seq data by effective length normalization. *Nucleic Acids Research* **39,** e9 (2011).

171. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics* **14,** 178–192 (2013).

172. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2009).

173. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11,** R106 (2010).

174. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4,** 44–57 (2009).

175. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37,** 1–13 (2009).