Transposable Elements and Human Pluripotency

Dissertation

zur Erlangung des akademischen Grades des
Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie
der Freie Universität Berlin

Vorgelegt von
Jichang Wang
王继厂

aus Jiangsu, China

April 2015

Die vorliegende Arbeit wurde von Dezember 2009 bis März 2015 unter der wissenschaftlichen Leitung von Dr. Zsuzsanna Izsvák am Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft (MDC), Berlin angefertigt.

1. Gutachter: Dr. Zsuzsanna Izsvák

Max-Delbrück-Centrum für molekulare Medizin in der Helmholtz-Gemeinschaft (MDC)
Robert-Rössle-Str. 10
13125 Berlin
Telefon: 030-9406 3510
E-Mail: zizsvak@mdc-berlin.de

2. Gutachter: Prof. Dr. Udo Heinemann

Max-Delbrück-Centrum für molekulare Medizin in der Helmholtz-Gemeinschaft (MDC)
Robert-Rössle-Str. 10
13125 Berlin
Telefon: 030-9406 3420
E-Mail: Heinemann@mdc-berlin.de

Datum der Disputation: 28.07.2015

Acknowledgment

Most of all I would like to thank my supervisor Dr. Zsuzsanna Izsvák for giving me the opportunity to study on the exciting projects in her group. I appreciated a lot her patience with me when I joined in her group with few of knowledge about molecular biology and poor English；I appreciated a lot her tolerance and open-minded when I argued with her about scientific questions. I appreciated a lot her guidance on how to do research; I appreciated a lot her support and care when I suffered from the disaster in my life.
I would like to thank Dr. Zoltán Ivics for his crucial advice and critical comments throughout my PhD study.

I would like to thank Prof. Dr. Udo Heinemann for the supervision at Freie Universität Berlin.

I would like to thank all my current and former labmates in the Mobile DNA group for the wonderful work discussion and help in my research and life. Especially, I would like to thank Drs. Yongming Wang and Mingbing Zhou for kindness and patience when I came to Berlin with loneliness and helplessness；I would like to thank Drs. Ismahen Ammar, Sanam Bashir, Anantharam Devaraj, Ivana Grabundzija, Esther Grueso, Dawid Grzela, Nina Hein-Fuchs, Lajos Mates, Judit Menyhert, Csaba Miskey, Eniko-Eva Nagy, Suneel Narayanavari, Ana Osiak, Tamas Rasko, Andrea Solf, Attila Szvetnik, and Ilija Bilic, Huqiang Cai, Malgorzata Anna Dalda, Helena Escobar, Angelica Garcia-Perez, Sandra Neuendorf, Ana Jimenez Orgaz, Vaishnavi Raghunathan, Julia Rugor, Manvendra Singh, Marta Swierczek for the advice and technical support as well as generous encouragement.

I would like to thank all of my collaborators for fantastic collaborations in the projects of "*Sleeping Beauty* transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells" and "Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells". I would like to thank Dr. Zoltan Cseresnyes for the technical assistance in the confocal imaging; I would like to thank the team of the FACS Core Facility in the MDC for their technical supports in FACS.

I would like to thank Sandra Neuendorf for translating the summary of my thesis into German.

I'm very grateful to all the Chinese community members of the MDC for coming together to share the experiences in overseas study and life. Especially, I appreciated a lot the help and support from Drs Maolian Gong, Na Liu, Qiu Jiang, Yu Shi, Kun Song, and

Zisong Chang and other friends.

Finally, I want to express my biggest gratitude to my parents and my wife. If without their support and love, I could not finish my PhD study with some achievement. Of course, the birth of my son has been motivating me to work harder and harder on science.

Summary

Ancient, transposable element (TE) - derived sequences, occupying around 60% of the human genome were considered as functionless junk until very recently. Curiously, TE-derived sequences could gain novel cellular functions in evolutionary time. Importantly, TEs, as natural gene delivery tools can be developed for many applications in translational research, including generating induced pluripotent stem cells (iPSCs). Derivation of pluripotent stem cells (PSCs) from embryos or somatic cells using transcription factor-mediated reprogramming strategies holds the promise in regenerative medicine. Deciphering the process of cell fate decision of human PSCs (hPSCs) will facilitate to optimize protocols of maintaining and differentiating these cells. Curiously, regardless of their ability to transpose, TE-derived sequences are activated during embryogenesis. To explore if TE-derived sequences have any role during the early steps of development, human induced pluripotent stem cells (hiPSCs) were generated from human foreskin fibroblasts, using the cutting-of-edge *Sleeping Beauty* TE-based reprogramming system. Using hiPSCs as the platform, I employed the RNA-seq technique to perform genome-wide transcription profiling of hiPSCs, intermediately differentiated cells (embryoid bodies) and their parental somatic cells (fibroblasts). By comparative analysis of their transcriptome, I observed that an ancient, primate-specific endogenous retrovirus family, HERVH was highly expressed in hPSCs. By means of ChIP-seq, gain/loss-of-function assays, I revealed that a set of core pluripotency regulators, OCT4, NANOG, KLF4 and LBP9 modulate HERVH expression in hPSCs. The expression of HERVH contributes to several novel primate-specific transcripts, including lncRNAs and chimeric gene products, with the potential of modulating pluripotency. Depletion of HERVH-derived transcription compromises self-renewal of hPSCs, while its induction promotes pluripotency from somatic cells. Using a HERVH-based reporter system, I have observed that a sub-population of conventional hPSCs shares some of the key features with naïve mouse embryonic stem cells and the human inner cell mass. Importantly, recruited to the regulatory circuitry of primate pluripotency, the HERVH-derived transcription redefines pluripotency, as being species specific. My work contributes to decipher unexpected roles of ancient TEs in fate decision of PSCs.

Zusammenfassung

Bis vor kurzem wurden die abgeleiteten Sequenzen der alten, transposablen Elemente (TE), welche rund 60% des menschlichen Genoms ausmachen, als funktionsloser Schrott betrachtet. Seltsamerweise konnten TE-abgeleitete Sequenzen während der Evolution neuartige, zelluläre Funktionen erwerben. Wichtig ist, dass TEs als natürliches Gen-Übertragungswerkzeug für viele Anwendungen in der translationalen Forschung entwickelt werden können, einschließlich der Erzeugung induzierter pluripotenter Stammzellen (iPS-Zellen). Die Herstellung pluripotenter Stammzellen (PSCs) aus Embryonen oder somatischen Zellen mit Hilfe transkriptionsfaktor-vermittelten Re-programmierungs-Strategien ist vielversprechend für die regenerative Medizin. Die Entschlüsselung des Zellschicksals der menschlichen PSCs (hPSCs) wird es ermöglichen, Protokolle für die Aufrechterhaltung und Differenzierung dieser Zellen zu optimieren. Merkwürdigerweise werden TE-abgeleitete Sequenzen, unabhängig von ihrer Fähigkeit zu transponieren, während der Embryogenese aktiviert. Um zu untersuchen, ob TE-abgeleitete Sequenzen eine Rolle in den frühen Stufen der Entwicklung haben, wurden menschliche, induzierte, pluripotente Stammzellen (iPS-Zellen) aus menschlichen Vorhaut-Fibroblasten unter Verwendung der Spitzentechnik des Sleeping Beauty TE-basierten Reprogrammierungsystems erzeugt. Mit hiPS als Plattform, nutzte ich die RNA-Sequenzierungstechnik, um genomweite Transkriptionsprofilierung von iPS-Zellen, intermediär differenzierten Zellen (embryonale Körperchen) und deren parentale somatische Zellen (Fibroblasten) durchzuführen. Durch vergleichende Analyse der Transkriptome beobachtete ich, dass eine alte, primaten-spezifische, endogene Retrovirus-Familie, HERVH, stark in hPSCs exprimiert war. Mittels ChIP-seq, „gain/loss-of-function-Assays“ konnte ich aufzeigen, dass eine Reihe von Kern-Pluripotenz-Regulatoren (OCT4, NANOG, Klf4 und LBP9) die HERVH-Expression in hPSCs modulieren. Die Expression von HERVH steuert mehrere neuartige primaten-spezifische Transkripte bei, einschließlich lncRNAs und chimäre Genprodukte, welche das Potential zur Modulation der Pluripotenz haben. Die Verminderung HERVH-abgeleiteter Transkription beeinträchtigt die Selbsterneuerung von hPSCs, während die Induktion die Pluripotenz der Körperzellen fördert. Mit Hilfe eines HERVH-basierten Reportersystems habe ich festgestellt, dass eine Subpopulation von konventionellen hPSCs einige der wichtigsten Merkmale mit naïven, embryonalen Stammzellen der Maus und des menschlichen Embryoblasten gemeinsam hat. Wichtig ist, einhergehend mit den regulatorischen Kreisläufen der Primaten-Pluripotenz, dass die HERVH-abgeleitete Transkription die Pluripotenz als artspezifisch neu definiert. Meine Arbeit trägt dazu bei,

unerwartete Funktionen der alten TEs in Bezug auf das Zellschicksal der PSCs zu entschlüsseln.

Table of Contents

1 Introduction

1.1 Naive pluripotency

1.1.1 The birth of naive pluripotency

Mammalian development starts from a fertilized egg (referred as zygote), and concludes with the formation of the fetus. The cells from zygotes, 2-cell, 4-cell, 8-cell and morula embryos are called “totipotent”, as they have the ability to form all the embryonic and extraembryonic tissues. While, the epiblast cells in the inner cell mass (ICM) of blastocysts can give rise to all somatic lineages and the germline, and are hence called “pluripotent” (Stadtfeld and Hochedlinger, 2010). During the developmental process, cells gradually lose potency, and are progressively differentiated to fulfill the specialized functions of somatic tissues. At the late blastocyst stage the ICM establishes the pluripotent epiblast and an overlying extra-embryonic layer of Gata6-expressing primitive endoderm (also known as hypoblast) (Silva et al., 2009). At this point the epiblast cells enter the developmental “ground state,” and considered as naïve pluripotent stem cells. Naivety reflects the ability of a cell to self-renew, while retaining the potential for unbiased differentiation and germline contribution (Boroviak and Nichols, 2014). The naïve ground state in the epiblast is characterized by uniform expression of core pluripotency transcription factors (TFs) (Dunn et al., 2014). Though the naïve pluripotency in the embryo is a transient condition, its nature can be captured indefinitely *in vitro,* through derivation of embryonic stem cells (ESCs) from the ICM (Evans and Kaufman, 1981; Martin, 1981). Naïve ESCs possess the same features as the ground state epiblast. For example, ESCs can give rise to mice when they are injected to the tetraploid hosts, reflecting their unbiased differential potential (Nagy et al., 1990). Similarly to the epiblast cells, the silent paternal X chromosome gets reactivated in the naïve ESCs derived from female mouse embryos (Mak et al., 2004; Okamoto et al., 2004).

In contrast, the “primed” state, which corresponds to the postimplantation embryos, is poised to initiate lineage-specification program. The primed state is epigenetically restricted, as one of X chromosomes has been randomly silenced (Bao et al., 2009; Tan et al., 1993). Recently, the so-called epiblast stem cells (EpiSCs) have been derived from the post-implantation embryos, providing a model to mimic the primed pluripotency *in vitro* (Brons et al., 2007; Tesar et al., 2007). Though EpiSCs exhibit pluripotent features similar to ESCs, such as the capacity to differentiate into three germ layers both *in vitro* and *in vivo*, they perform poorly in the tetraploid complementation assay (Brons et al., 2007; Tesar et al., 2007). Interestingly, the ESCs and EpiSCs states can transit between

each other under certain conditions, indicating that there is state-specific circuitry for their maintenance (Bao et al., 2009).

1.1.2 Naïve pluripotency definition by the transcription factor network

Regulatory networks of TFs determine the gene expression programs that stabilize naive pluripotency (Dunn et al., 2014). Two TFs, *Oct4* and *Sox2*, are widely recognized as being fundamental for both the acquisition and maintenance of pluripotency (Avilion et al., 2003; Masui et al., 2007). In the absence of *Oct4* and *Sox2* expression, ESCs differentiate progressively toward trophectoderm (Niwa et al., 2000; Thomson et al., 2011).

In addition to Oct4 and Sox2, Nanog is also considered as a core pluripotency TF. Indeed, Nanog-null epiblast cells cannot be found in embryos, reflecting its crucial role in the acquisition of naive pluripotency (Chambers et al., 2007; Silva et al., 2009). However, Nanog is not required for maintenance of naive pluripotency, as Nanog-null ESCs can self-renew infinitely, and retain the features for naïve pluripotency (Chambers et al., 2007; Silva et al., 2009). Nonetheless, the binding sites of Oct4, Sox2 and Nanog frequently overlap in the whole genome of ESCs, suggesting these three core TFs acts cooperatively to form a robust regulatory network for the naïve pluripotency (Boyer et al., 2005).

The downstream TFs, regulated directly by these three core pluripotency factors, usually act co-activators of *Oct4*, *Sox2* and *Nanog*. These downstream TFs are referred as “ancillary” pluripotency regulators, including *Klf2*, *Esrrb*, *Klf4*, *Prdm14*, *Sall4*, *Tfcp2l1*, and *Tbx3.* These TFs are individually dispensable for pluripotent identity, but can reinforce and buffer the pluripotency network against pro-differentiation influences from microenvironment (Hackett and Surani, 2014). The expression of all of above TFs is much lower or undetectable in EpiSCs compared to ESCs, reflecting their specificities for naïve pluripotency. Interestingly, *Nanog* and the ancillary” pluripotency regulators are uniformly expressed in ESCs under serum-free condition, but exhibit heterogeneous expression under serum condition (Kumar et al., 2014). Their expression fluctuates between *on* and *off* states at the single-ESC level, suggesting that ESCs show the metastable naïve pluripotency under serum condition, which may be affected by extrinsic signals. To obtain the stabilized naïve pluripotency *in vitro*, research is dedicated to identify signal pathway(s) that may impact on the stability of naïve pluripotency.

1.1.3 Signaling pathways for naïve pluripotency

Mouse ESCs (mESCs) are derived from the blastocyst and cultured on feeder cells (mitotically inactivated mouse embryonic fibroblasts, MEFs). In serum-containing medium MEFs secrete leukemia inhibitory factor (LIF). LIF is essential for maintaining the self-renewal of mESCs, by activating Jak/Stat3 signaling pathway via binding the gp130/LIF-R cell-surface receptor complex (Yoshida et al., 1994). This is consistent with the finding that Stat3 is required for sustaining the pluripotent ICM *in vivo* (Do et al., 2013). The downstream targets of LIF/Stat3 are mainly naïve pluripotency TFs, such as *Klf4*, *Gbx2*, *Tfcp2l1* (Cartwright et al., 2005; Martello et al., 2013; Niwa et al., 2009; Ye et al., 2013). In conjunction with LIF, serum sustains self-renewal of mESCs. The crucial component of the serum is BMP4, which may, at least in part, enhance E-cadherin expression via downstream SMAD signaling pathways. The *Inhibitor of Differentiation* (*Id*) genes activated by BMP4 can substitute for serum to maintain mESC self-renewal (Ying et al., 2003). In the absence of BMP4, mESCs progressively differentiate toward neuroectoderm (Ying et al., 2003). In summary, LIF/Stat3 and BMP4/SMAD pathways act cooperatively to maintain the self-renewal of mESCs under serum condition by suppressing differentiation towards specific lineages.

Although mESCs exhibit the features for naïve pluripotency under serum condition, they also express some lineage markers at low levels. This may be associated with the Fgf/MAPK signaling pathway that is also activated by LIF. In fact, the Fgf/MAPK signaling pathway is detrimental to the acquisition and maintenance of naïve epiblast (Nichols et al., 2009; Yamanaka et al., 2010). In mESCs, Erk1/2, the downstream effector of Fgf/MAPK signaling pathway is confirmed to degrade Klf2, and destabilize the naïve pluripotency network (Yeo et al., 2014). In parallel, Erk1/2 promote establishment of the primed state by depositing PRC2 (polycomb repressive complex 2) proteins on developmental genes (Tee et al., 2014). Indeed, mESCs can be transited into EpiSCs, a more “primed” state, by adding Fgf or Activin A, TGFβ to the medium (Vallier et al., 2009). Collectively, these findings reveal that Fgf/MAPK and Activin/TGFβ/SMAD signaling pathways are essential to sustain the self-renewal of mouse EpiSCs.

Therefore, it was hypothesized that the inhibition of the Fgf/MAPK signaling pathway might robustly stabilize the ground state of mESCs (Silva and Smith, 2008). Still, inhibiting the Fgf/MAPK signaling pathway alone by PD032901 is insufficient to support ESC viability in the absence of LIF. However, when Austin Smith and his colleagues combined PD032901 with CHIR9901 (one inhibitor of GSK3β) (termed “2i condition”) the propagation of mESCs was successfully stabilized and clonogenicity was improved even without LIF or serum (Sato et al., 2004; Ying et al., 2008). Thus, the 2i condition is

capable of robustly capturing the naïve mESCs *in vitro*. Using the 2i condition to culture mESCs diminishes the probability of differentiation towards specific lineages, and improves the generation efficacy of germline transmitted chimeras from mESCs (Ying et al., 2008).

The effects of GSK3β inhibition on self-renewal are principally mediated through the stimulation of canonical WNT signaling. Inhibition of GSK3β stabilizes β-catenin that is translocated into the nucleus, and enhances the expression of pluripotency factors (Kelly et al., 2011; Yi et al., 2011). In the nucleus, β-catenin can stabilize the naïve pluripotency by antagonistically interacting with TCF3, a transcriptional repressor of naïve pluripotency TFs (e.g. *Nanog*, *Nr0b1*, *Tfcp2l1* and *Esrrb*) (Faunes et al., 2013; Martello et al., 2013; Martello et al., 2012; Wray et al., 2011).

How the Fgf/MAPK inhibitor induces and maintains naïve pluripotency is still elusive. Inhibition of Fgf/MAPK signaling in ESCs drives rapid genome-wide demethylation via reducing expression of the *de novo* methyltransferase genes (*Dnmt3a* and *Dnmt3b),* a pattern like the epigenome of the ICM in the blastocyst (Ficz et al., 2013). Prdm14 was shown to directly suppress *Dnmt3b* expression, and the depletion of *Prdm14* caused increased methylation in 2i/LIF-treated ESCs (Ficz et al., 2013). However, depletion of *Prdm14* did not significantly compromise ESC self-renewal in 2i/LIF condition, indicating that Prdm14 is unlikely the primary mediator of the Fgf/MAPK inhibitor (Yamaji et al., 2013).

1.1.4 Epigenetic pattern of naïve pluripotency

DNA methylation is a heritable epigenetic mark present at cytosine residues in the mammalian genomes. More than 98% of DNA methylation occurs at CpG dinucleotides in somatic cells (Lister et al., 2009). Once established, DNA methylation is faithfully propagated during mitosis in somatic tissues. In contrast, DNA methylation is highly dynamic during embryogenesis. After fertilization, global DNA methylation is gradually reduced and reaches its minimum at the blastocyst stage (Guo et al., 2014; Smith et al., 2014; Smith et al., 2012). After implantation, a major wave of DNA re-methylation occurs resulting in lineage restriction and the loss of cellular potency (Guo et al., 2014; Smith et al., 2012).

Though mESCs are derived from the globally hypomethylated ICM, mESCs cultured in serum condition are hypermethylated with an average CpG methylation level of 70-80%, comparable to the postimplantation embryo (E6.5) (Smith et al., 2012; Stadler et al., 2011). The enhanced global DNA methylation is associated with primed or lineage-restricted cells (Meissner et al., 2008). In contrast, mESCs cultured in 2i/LIF condition

exhibit a globally hypomethylated DNA methylome, resembling the profiling of the ICM in the blastocyst (E3.5-E4.5) (Ficz et al., 2013; Habibi et al., 2013; Hackett et al., 2013; Leitch et al., 2013).

Surprisingly, the global demethylation in mESCs happens quickly during the switch from serum to 2i/LIF condition (Leitch et al., 2013; Shipony et al., 2014). However, some regions (mainly imprinted loci and Intracisternal A-type particle (IAP) retroelements) are still resistant to demethylation, and are marked with H3K9me3 (Ficz et al., 2013; Habibi et al., 2013). Dnmt1 and Uhrf1 were implicated in the interplay between DNA methylation and H3K9me3 in 2i/LIF-treated ESCs (Liu et al., 2013).

Switching from 2i/LIF to serum condition rapidly upregulates, while transferring mESCs from serum to 2i/LIF condition dramatically downregulates the *de novo* methyltransferases (*Dnmt3a*, *Dnmt3b*) (Ficz et al., 2013). DNA demethylases *Tet1/2* exhibit an opposite pattern, implying their roles in global CpG erasure (Ficz et al., 2013; Hackett et al., 2013).

In mESCs, the promoters of developmental genes are marked with both active (H3K4me3) and repressive (H3K27me3) histone marks, a phenomenon called “bivalency” (Heintzman et al., 2009). This bivalent signature is thought to maintain a flexible poised state that can be either rapidly reactivated through removal of methylation at H3K27 or repressed through demethylation of H3K4me3 during lineage commitment (Vastenhouw and Schier, 2012). The genome-wide distribution of H3K27me3 is dramatically lower in naive mESCs as compared to EpiSCs, and consequently the number of bivalent promoters is significantly reduced (Marks et al., 2012). The reduction of H3K27me3 may be a direct effect of Erk inhibition by PD032901, since Erk is necessary for EED activity, a member of the PRC2 complex at target promoters (Tee et al., 2014).

Other repressive marks, such as H3K9me2 and H3K9me3 are also shown a global reduction in the naïve mESCs (Habibi et al., 2013). In summary, multiple repressive modifications (DNA methylation, H3K27me3, H3K9me2, H3K9me3) are apparently reduced or redistributed in the ground state, while several epigenetic mechanisms linked with open chromatin are active. Such epigenetic state makes the cells at the ground state most plastic, which is conducive to the onset of all developmental programs upon appropriate cues.

1.1.5 Metabolic features for naïve pluripotency

In mammals, cells produce ATP by consuming glucose either by glycolysis or oxidative phosphorylation (OXPHOS). Glycolysis is preferred when energy generation is relatively

inefficient, e.g. in hypoxic environment in the uterus during embryo development. By contrast, cells prefer OXPHOS to support more efficient energy production, e.g. in oxygen-rich environment during cell proliferation. During preimplantation, ATP is produced mainly by OXPHOS (Brinster and Troike, 1979; Leese, 2012; Martin and Leese, 1995) and then shifted to a balanced glycolysis and OXPHOS in the low O_2 microenvironment of postimplantation embryos (Houghton et al., 1996; Leese and Barton, 1984). Similarly, naïve mESCs exhibit a bivalent metabolism that can switch between glycolysis and OXPHOS. In contrast, mESCs in serum condition and primed EpiSCs show the glycolytic metabolism (Zhou et al., 2012). Accordingly, comparative transcriptional analysis between naïve mESCs and serum mESCs/EpiSCs revealed that highly expressed genes in naïve ESCs were selectively associated with OXPHOS (Marks et al., 2012; Takashima et al., 2014; Ware et al., 2014). Such a transition of metabolism between naïve and primed state reflects the higher demand for energy, since naïve mESCs proliferate much faster than the primed EpiSCs.

1.1.6 Naïve human pluripotency

Like mESCs, human ESCs (hESCs) are also derived from the ICM of the human blastocyst and can be differentiated into three germ layers both *in vitro* and *in vivo* (Thomson et al., 1998). However, mouse and human ESCs have several divergent properties. First, hESCs are dependent on FGF and Activin A rather than LIF for self-renewal. Second, proliferation rate for hESCs is much slower when compared to naïve mESCs (Hanna et al., 2010). Furthermore, morphologically, hESCs are more similar to mouse EpiSCs than to mESCs. Indeed, expression of primed/lineage-associated genes (e.g. *LEFTY1/2, MYC, GATA4/6*, *T*) is readily detected in hESCs. In addition, similarly to mouse EpiSCs, *OCT4* expression in hESCs is controlled by its proximal enhancer (Brons et al., 2007; Hanna et al., 2010; Tesar et al., 2007; Theunissen et al., 2014). Thus, hESCs may occupy a phase of pluripotency that is similar to the murine primed rather than naive state. This assumption is further supported by comparative analyses of transcriptome and epigenome between hESCs and epiblast cells within the human blastocyst (Guo et al., 2014; Smith et al., 2014; Yan et al., 2013).

Why are hESCs so different from mESCs, when both cell types are derived from the ICM within the blastocyst? One plausible explanation is that the species-specific genetic background determines if the outgrowth of the embryonic cells is maintained in the blastocyst stage (ground state), or continues to progress towards an advanced and stable stage (e.g. postimplantation epiblast status where EpiSCs are derived).

The concept about ground state is very important for elucidating the mechanism of early development, facilitating genetic manipulation, and potentially enhancing the cells differentiated into all lineages *in vitro* (Gafni et al., 2013; Morgani et al., 2013). So far, the ground state has been successfully reproduced in mice and rats (Chen et al., 2013b; Meek et al., 2013). However, the difficulties to derive and stably maintain naïve hESCs *in vitro* have not been completely resolved.

Several studies have reported that hESCs/iPSCs with naïve features can be derived and maintained for limited passages. These strategies included ectopic expression of *NANOG* and *KLF2/4* (Buecker et al., 2010; Hanna et al., 2010; Takashima et al., 2014). However, genetic manipulation of stem cells may induce undesirable artifacts, and limit their clinical application. Therefore, the strategy of improving culture conditions that can support naïve human pluripotency could be a more acceptable approach.

In order to establish optimal conditions to support human naive pluripotency, several large-scale chemical screenings have been performed (Chan et al., 2013; Theunissen et al., 2014; Ware et al., 2014). As a result, several human-specific naïve culture conditions have been established. Some of them were also successfully used to generate naïve-like primate iPSCs from fibroblasts (Fang et al., 2014). Importantly, these novel naïve-like hESCs show the morphology, the global gene expression profile resembling of naïve mESCs and have improved probability of clonogenicity. The different protocols seem to induce alternative pluripotent states, which may functionally mimic the different *in vivo* phases of pluripotency. However, the global DNA methylation pattern of these naïve-like hESCs resembles rather the conventional hESCs than mESCs or human blastocysts, indicating that the genome-wide DNA methylation has not been completely erased (Gafni et al., 2013). As a consequence, the transcription factor regulatory network for ground state has not been successfully reactivated in these naïve-like hESCs, since several core naïve pluripotency TFs are either expressed in a lower level than mESCs, or undetectable at all (Chan et al., 2013; Gafni et al., 2013; Ware et al., 2014). The naïve 5iL/A hESCs, generated by the Jaenisch group seem to have most features similar to naïve mESCs. Still, the silent X chromosomes is not reactivated in these cells, indicating that the chromatin is not thoroughly remodeled to the ground state (Theunissen et al., 2014), Surprisingly, bFGF, TGFβ and Activin A, which are required for derivation and maintenance of mouse EpiSCs and conventional hESCs, are included in these human-specific naive culture conditions, suggesting that FGF and TGFβ/Activin signaling pathways may have positive effects on self-renewal of naïve hESCs. Although these attempts to derive naive hESCs appear to be somewhat controversial, they definitely extend our understanding of the naïve pluripotency in human.

Notably, cross-species comparative analyses of global gene expression profiles reveal, that human and mouse naïve ESCs resemble rather their respective epiblast cells within the blastocysts than each other (Huang et al., 2014a). This observation would strongly indicate that the species-specific genetic background has a crucial influence on pluripotency. Thus, human epiblast cells within the blastocyst rather than the naïve mESCs should be taken as the "golden standard" to derive authentic human naïve pluripotent stem cells. Notably, transposable elements, especially endogenous retroviruses (ERVs) are recently reported to have unexpected regulatory functions during the early development and epigenetic reprogramming in mammals (Fort et al., 2014; Kunarso et al., 2010; Macfarlan et al., 2012; Macia et al., 2014; Ohnuki et al., 2014; Rowe et al., 2013b). Thus, in principle the primate/human-specific transposable elements might also affect pluripotency. Therefore, I will focus on this point for my study.

1.2 Induced pluripotency

1.2.1 History of induced pluripotency from somatic cells

Decades of research on cell fate determination lead to the view that, *in vivo*, differentiated cells are irreversibly committed to their fate. However, the fantastic finding that pluripotency can be induced in somatic cells revealed a remarkable plasticity of the differentiated state. So far, different strategies, such as (i) somatic cell nuclear transfer, (ii) cell fusion and (iii) ectopic expression of defined transcription factors has been developed to reset the epigenome of somatic cell to a pluripotent state.

Reprogramming by somatic cell nuclear transfer

The first studies on induced pluripotency started over 60 years ago, when Gurdon has successfully obtained normal adult frogs from ultraviolet-light-irradiated oocytes transferred with nuclei from highly specialized tadpole intestinal cells (Gurdon, 1962a, b). This method, called as the somatic cell nuclear transfer (SCNT), has been reproduced in mammals (Byrne et al., 2007; Eggan et al., 2004; Hochedlinger and Jaenisch, 2002; Wakayama et al., 1998; Wilmut et al., 1997). From the time when Dolly, the sheep was successfully cloned in 1997, these types of experiments were referred as "cloning" to the public (Wilmut et al., 1997). SCNT is thought to be one of the best ways to mimic the natural fertilization (Gurdon and Wilmut, 2011; Wilmut et al., 1997). The cloned embryonic cells can be cultured *in vitro*, and finally give rise to stabilized embryonic stem cell lines (NT-ESCs). These NT-ESCs are considered to be the closest to the ESCs from the fertilized embryo (Tachibana et al., 2013; Wakayama et al., 2006). Though SCNT is a powerful tool to track the developmental potential of a cell, it is technically challenging. More importantly, the development defects of the cloned animals indicate that this

technique may not completely erase the epigenetic memory from somatic cells (Chan et al., 2012; Gurdon and Wilmut, 2011; Simonsson and Gurdon, 2004; Thuan et al., 2010). A better understanding of the gene/epigenetic regulation during nuclear reprogramming was necessary to develop more reproducible methods to induce somatic cells into a pluripotent state.

Reprogramming by cell fusion

Cell fusion involves fusing two or more cell types from the same or different species, to form a single entity. The fused cells can be hybrids (which can proliferate, causing the nuclei of the original cell to fuse) or heterokaryons (which do not divide and therefore contain multiple distinct nuclei in cytoplasm). Tada, et al. were the first to demonstrate nuclear reprogramming of somatic cells in hybrids. These cells were generated from fusion between female embryonic germ cells and thymocytes from adult mice (Tada et al., 1997). Their tetraploid cells could be differentiated into three germ layers in chimeric embryos, indicating their pluripotency. In their follow-up studies they further showed that somatic cells could be reprogrammed into a pluripotent state after being fused with mESCs (Tada et al., 2001). This technique has been successfully reproduced in human (Cowan et al., 2005; Yu et al., 2006). In these pluripotent cells the pluripotency-associated genes (e.g. *OCT4*, *SOX2* and *NANOG*) were activated, the somatic genes were silenced and the genome went through the genome-wide demethylation process (Cowan et al., 2005; Yu et al., 2006). Heterokaryons are considered to be an ideal platform to elucidate the molecular mechanism of reprogramming to a pluripotent state. Their reprogramming rate is much faster in heterokaryons than in hybrids (Gridina and Serov, 2010). The shortcoming of this method is that the reprogrammed cells induced by cell-fusion are tetraploid, limiting its application for cell therapy.

Reprogramming with transcription factors

The fate of a cell can be changed by ectopic expression of certain TFs involved in establishing and maintaining cellular identity. Lassar *et al*. were the first to demonstrate the *proof of principle*. They reported on the formation of myofibers in fibroblast cell lines transduced with retroviral vectors expressing the skeletal muscle factor MyoD (Lassar et al., 1989). Later, Graf et al. showed that overexpression of C/EBPα could efficiently convert primary B and T cells into functional macrophages (Xie et al., 2004). These findings inspired investigators to ask whether somatic cells can be reprogrammed into a pluripotent state through ectopic expression of multiple pluripotency-associated TFs. Many investigators devoted research to identify the master regulators of the pluripotency network (Avilion et al., 2003; Benvenisty et al., 1992; Bowles et al., 2003; Boyer et al., 2005; Cartwright et al., 2005; Chew et al., 2005; Elling et al., 2006; Jiang et al., 2008; Li

et al., 2005; Loh et al., 2008; Masui et al., 2007; Mitsui et al., 2003; Nakatake et al., 2006; Niwa et al., 2000; Wu et al., 2006; Zhang et al., 2006). In 2006, Yamanaka and his colleague identified four transcription factors (*Oct4*, *Sox2*, *Klf4*, and *c-Myc*, called as OSKM briefly) from a pool of 24 regulators, and successfully generated mESC-like pluripotent cells (called induced pluripotent stem cells (iPSCs)) from mouse fibroblasts (Takahashi and Yamanaka, 2006). Due to its simplicity and reproducibility (Hanna et al., 2007; Maherali et al., 2007; Meissner et al., 2007; Okita et al., 2007; Wernig et al., 2007), the iPSC approach becomes the prevalent method in the whole world, instead of SCNT and cell-fusion. In the last couple of years, iPSCs have been successfully derived from various differentiated cell types in a wide range of species, including humans (Aoi et al., 2008; Gianotti-Sommer et al., 2008; Haase et al., 2009; Kim et al., 2008; Li et al., 2009; Liu et al., 2008; Takahashi et al., 2007; Wu et al., 2009; Yu et al., 2007). These scientific breakthroughs generated a platform to explore the mechanism of epigenetic reprogramming and development of embryos. In addition, iPSCs became feasible to generate custom-tailored cells for modeling and possibly treating human diseases.

The success of iPSCs inspires and boosts the development of other orientations of the stem cell research, such as trans-differentiation. So far, researchers have successfully induced the conversion of fibroblast into neurons by overexpressing the neural factors *Ascl1*, *Brn2*, and *Myt1l* (Vierbuchen et al., 2010); into cardiomyocytes by overexpression of the cardiac factors *Gata4*, *Mef2c*, and *Tbx5* (Ieda et al., 2010); into blood cells (Szabo et al., 2010); into liver cells (Huang et al., 2011; Huang et al., 2014b; Yu et al., 2013; Zhu et al., 2014) or pancreatic acinar cells into insulin-producing β cells by overexpressing the pancreatic factors *MafA*, *Pdx1*, and *Ngn3* (Zhou et al., 2008).

1.2.2 Progress of iPSC generation by defined transcription factors

1.2.2.1 Methods for the delivery of reprogramming transcription factors

The delivery of the reprogramming TFs into mouse or human fibroblast was originally achieved using the retrovirus-based vectors, which are usually silenced in ESCs (Cherry et al., 2000). Silencing the exogenous genes is important. Only the “real” iPSCs that have the reactivated endogenous pluripotency network, and are independent from exogenous genes, are fully reprogrammed. By contrast, the partially reprogrammed iPSCs still depend on the expression of exogenous genes (Hotta and Ellis, 2008). The efficacy of iPSC generation, using retroviral vectors expressing the OSKM set is ~0.1% in mouse embryonic fibroblasts and ~0.01% in human fibroblasts (Takahashi et al., 2007; Takahashi and Yamanaka, 2006). Interestingly, retrovirus-induced innate immunity improves the reprogramming efficiency (Lee et al., 2012).

Lentiviral vectors have also been faithfully used to express different cocktails of reprogramming factors in somatic cells. The efficacy of reprogramming by lentiviral vectors is comparable to the retroviral method, but the lentivirus preparation is bit more complicated and time-consuming. Notably, the combination of TET-inducible reprogramming system with lentiviral vectors renders the reprogramming process controllable (Chang et al., 2009; Welstead et al., 2008).

However, there are serious safety concerns regarding viral vectors. These include insertional mutagenesis, reactivation of endogenous oncogenes or repression of tumor suppressor genes (Stein et al., 2010). Therefore, it is important to develop alternative non-viral or non-integrative strategies to generate iPSCs for therapeutic purposes. So far, there are several different non-viral and non-integrative approaches available for iPSC generation.

The transfection of the linear polycistronic vector, providing almost equivalent expression of multiple-genes from the same promoter, was employed to successfully reprogram mouse fibroblasts into iPSCs. However, the efficiency of this method was quite low compared to the integrating viral vector-based strategies (Hasegawa et al., 2007). As the polycistronic vectors allow reprogramming somatic cells via a single insertion, researchers used transposons (e.g. *PiggyBac*) to improve the transfection and reprogramming frequencies (Woltjen et al., 2009; Yusa et al., 2009). As the reactivation or constitutive expression of the reprogramming factors have oncogenic potential and may inhibit the iPSC differentiation, it might be important to remove them. Flanking the polycistronic reprogramming factors by loxP, the reprogramming cassette can be removed from the matured iPSCs (Sommer et al., 2010). Notably, re-expression of the *PiggyBac* transposase in the matured miPSCs has successfully excised the reprogramming cassette that delivered by the *PiggyBac* transposon without leaving a footprint, though the frequency is quite low (Yusa et al., 2009).

Alternatively to the integrating strategies to generate iPSCs, several non-integrative methods have been developed, These approaches are based on using replication-defective adenoviral vectors (Stadtfeld et al., 2008), F-deficient Sendai viral vectors (Ban et al., 2011; Daheron and D'Souza, 2008; Fusaki et al., 2009; Seki et al., 2010), episomal vectors (Yu et al., 2009), or directly introducing mRNA (Yakubov et al., 2010) or proteins (Cho et al., 2010; Zhou et al., 2009) into the somatic cells. Still, compared to the integrating viral vectors, the reprogramming efficiency is lower, especially in human somatic cells. Furthermore, the preparation and delivery of non-integrative viral vectors, mRNA and proteins are much more complicated, limiting their use in large-scale human iPSC generation for clinical application and drug screening.

Taken together, it is quite important to explore the mechanisms by which somatic cells are reprogrammed into iPSCs, as well as to look for potential barriers to reprogramming. Elucidating the mechanisms underlying the epigenetic reprogramming is conductive to developing novel and efficient reprogramming methods.

1.2.2.2 Reprogramming by the alternative factors

Notably, the defined TFs that Yamanaka used were not the same as that Yu et al (Yu et al., 2007) used in the human system, suggesting that there is remarkable flexibility in the choice of reprogramming factors. As Yamanaka's cocktail contains oncogenic factors (*c-Myc*), and the reprogramming efficiency is low, researchers are constantly trying to modify the original cocktail with alternative factors.

One of the first modifications was to generate iPSCs from fibroblasts using only three of the four factors: *Oct4*, *Sox2*, and *Klf4* (OSK) (Nakagawa et al., 2008; Wernig et al., 2008). However, the three-factor reprogramming process was significantly slower and the efficiency of iPSC generation was poor, indicating the important roles of *c-Myc* in boosting the reprogramming process.

By now, each gene of the original Yamanaka's cocktail has been shown to be replaceable by other transcription factors (Buganim et al., 2012). This attribute has revealed a high degree of redundancy among the genetic factors capable of inducing pluripotency. Furthermore, there is some inter-specific redundancy as well, as mouse *Oct4*, *Klf4*, *Sox2*, and *c-Myc* have reprogramming activities in human fibroblasts, although with reduced efficiency (Grabundzija et al., 2013; Nakagawa et al., 2008). *Klf4* and *c-Myc* could be replaced by *Esrrb* in mouse fibroblasts (Feng et al., 2009) or by the combination of *NANOG* and *LIN28* in human fibroblasts (Yu et al., 2007). *Oct4* could be substituted by *Nr5a1*, *Nr5a2* (Heng et al., 2010) or *Cdh1* (Redmer et al., 2011). *Oct4* can also be replaced by *Tet1*, one of the key regulators of DNA methylation (Gao et al., 2013). By analyzing the single-cell expression data of 48 genes at different time points of the reprogramming process, the Jaenisch lab identified two combinations of four factors that could replace OSKM entirely: *Sall4*, *Esrrb*, and *Lin28* combined with either *Dppa2* or *Nanog*, all of which are downstream targets reactivated by *Sox2* and *Oct4* (Buganim et al., 2012). Importantly, the Jaenisch's cocktail could improve the quality of iPSCs that more efficiently gave rise to "all-iPSC" mice compared with Yamanaka's cocktail (Buganim et al., 2014). Most of the 'pluripotency inducers' are highly expressed in ESCs. Surprisingly, two recent studies reported lineage specifiers that are not enriched in ESCs, but still could substitute for *Oct4* and *Sox2* during the reprogramming process (Montserrat et al., 2013; Shu et al., 2013). Deng and his colleagues found that mesendodermal (ME) lineage specifiers (e.g. *Gata3*, *Gata6*, *Sox7*, *Pax1*, *Gata4*, *C/EBPa*, *HNF4a*, and *Grb2*)

could be alternative to *Oct4*. These TFs inhibited the expression of a set of ectodermal (ECT) specification-related genes that are elevated by *Sox2*, *Klf4*, and *c-Myc*. Similarly, the ectodermal lineage specifiers (*Sox1*, *Sox3*, and *Gmn*) could replace *Sox2* by attenuating the induction of ME genes in a combination with *Oct4*, *Klf4*, and *c-Myc*. Remarkably, *Oct4* and *Sox2* could be replaced by multiple combinations of ECT and ME specifiers simultaneously (Shu et al., 2013). This phenomenon can be explained by a ‘‘seesaw’’ model proposing that a somatic cell has greater potential of reaching pluripotency when it is balanced by two opposing differentiation potentials (Shu et al., 2013).

Researchers also identified several activators that in combination with OSKM promoted the programming process. These activators improve the efficiency and/or the quality of iPSC generation. One example is *Nanog*, which reduces the time for appearance of iPSCs from B cells by half, when it was co-transduced with OSKM (Hanna et al., 2009b). Another example is *UTF1*, which increases the number of iPSCs when overexpressed in human fibroblasts in combination with the OSKM (Zhao et al., 2008). Likewise, combination of *Tbx3* with OSKM improves the quality of iPSCs (Han et al., 2010). Most of these activators are abundant in ESCs and are associated with the maintenance of pluripotency. However, Yamanaka and his colleagues identified one oocyte-specific transcription factor *Glis1*, not expressed in the blastocyst or ESCs, as capable of enhancing the reprogramming efficiency of both mouse and human somatic cells (Maekawa et al., 2011). *Glis1* may represent a link between TF- and SCNT-regulated reprogramming.

There are also many inhibitors of the reprogramming process, and several of them have been identified in genome-wide screens using the RNA interference (RNAi) method (Qin et al., 2014). The most widely known factor is the tumor-suppressor gene, *p53*. Depletion of *p53* by transient RNAi or knockout, dramatically accelerates the reprogramming process, and increases the number of authentic iPSC colonies (Kawamura et al., 2009).

Recently, non-coding RNAs were shown to play important roles in maintaining the self-renewal and differentiation of ESCs (Guttman et al., 2011; Marson et al., 2008), indicating their potential roles in the reprogramming process. For example, *Lin28*, a negative regulator of the *Let-7* microRNA (miRNA) family was shown to accelerate the iPSC generation (Viswanathan et al., 2008). A subset of the mir-290 cluster improved the efficiency of reprogramming induced by OSK. mir-290 was assumed to be a downstream target of *c-Myc* (Judson et al., 2009). Surprisingly, iPSCs could derive from either mouse or human fibroblasts by simply overexpressing certain miRNAs (Anokye-Danso et al., 2011; Miyoshi et al., 2011). miRNAs were implicated to promote reprogramming via

multiple mechanisms, including the regulation of the mesenchymal-to-epithelial transition (MET) (Li et al., 2010; Liao et al., 2011; Samavarchi-Tehrani et al., 2010; Subramanyam et al., 2011). Yamanaka and his colleagues reported that some miRNAs (e.g. *Let-7* miRNAs) have a repressive effect on the reprogramming process (Worringer et al., 2014). Daley et al. identified a *de novo* long non-coding RNA (lncRNA) in human iPSCs, named as LINC-ROR (**L**ong **I**ntergenic **N**on-protein **C**oding RNA, **R**egulator **O**f **R**eprogramming), which promoted reprogramming process of human somatic cells, by inhibiting *p53* (Loewer et al., 2010).

1.2.2.3 Reprogramming by chemical compounds

To make iPSCs more suitable for therapeutic applications, an important aim is to identify chemicals that can replace transgenes used for the epigenetic reprogramming. Chemical screens have identified compounds that facilitate the reprogramming process or can replace individual Yamanaka factors during the generation of iPSC. For example, Pei and his colleagues showed that Vitamin C could improve the reprogramming efficiency and iPSC quality via its role in histone modification (Esteban et al., 2010; Stadtfeld et al., 2012; Wang et al., 2011). Melton et al. have reported that valproic acid (VPA), a histone deacetylase (HDAC) inhibitor, enabled efficient reprogramming of either mouse fibroblasts induced by OSK (Huangfu et al., 2008a) or human fibroblasts induced by only OS (Huangfu et al., 2008b). Lyssiotis et al. found the GSK3β and CDK inhibitor kenpaullone (KP) could replace *Klf4* for *iPSC* generation (Lyssiotis et al., 2009). Inhibition of TGFβ signaling was capable of activating endogenous *Nanog* and replacing *Sox2* and *c-Myc* (Ichida et al., 2009; Maherali and Hochedlinger, 2009). The combination of MEK inhibitor PD0325901, GSK3β inhibitor CHIR99021, TGFβ inhibitor A-83-01, and ROCK inhibitor Y-29632 significantly improved the episomal reprogramming efficiency from human fibroblasts (Yu et al., 2011).

In 2013, Deng and his colleagues successfully identified several *Oct4* replacers (e.g. the cyclic AMP agonist forskolin (FSK)) by the high-throughput chemical screening. Strikingly, the Deng lab obtained germline-competent iPSCs, reprogrammed from mouse fibroblasts using a combination of FSK and a cocktail of inhibitors (CHIR990291, 616452, VPA, Tranylcypromine, and DZNep) that were previously reported to support Oct4-induced reprogramming (Hou et al., 2013). These small molecules are shown to activate several inducers of pluripotency, and also reduce DNA methylation and H3K9 methylation levels at the *Oct4* locus (Hou et al., 2013). Generation of human iPSCs using chemical compounds alone is expected to be the next major step in the field of induced pluripotency. The combination of TGFβ, HDAC and MEK inhibitors enabled the reprogramming of human primary somatic cells by overexpression of OCT4 alone (Zhu et

al., 2010b). Thus, alternative small molecules replacing *OCT4* might make the chemical-induced hiPSCs to be truth in future.

1.3 Transposable elements

1.3.1 Classification of transposable elements

Transposable elements (TEs), also known as transposons, are sequences of DNA that can move from one site in the genome to another. TEs are further classified as DNA transposons and retroelements (also called retrotransposons).

1.3.1.1 DNA transposon

DNA transposons have been identified in genomes of all of prokaryotic and eukaryotic organisms. They structurally contain a sequence encoding the transposase flanked by double inverted terminal repeats (ITRs). Certain DNA transposons encode for additional genes to modulate transposition (Ivics et al., 2004). The mechanism of DNA transposition does not involve an RNA intermediate. The DNA transposon is excised from the donor locus and subsequently re-integrates into a new genomic location by its transposase (so-called “cut and paste” transposition). This mechanism is not replicative. DNA transposon elements amplify their copy number by piggybacking host cellular mechanisms (e.g. replication, DNA repair). As the transposase can catalyze a transposition reaction in trans, and from an extrachromosomal molecule to the genome, DNA transposons are ideal non-viral gene delivery tools. *Sleeping Beauty* (Ivics et al., 1997) and *PiggyBac* (Ding et al., 2005), the two most famous transposons are widely used for various applications, including transgenesis, gene therapy or annotating gene function (Clark et al., 2004; Ivics and Izsvak, 2004; Ivics et al., 2011; Izsvak and Ivics, 2004).

1.3.1.2 Retroelements

Retroelements can amplify themselves in the genome. In detail, they are first transcribed from DNA to RNA, followed by reverse transcription of the RNA intermediate to DNA that is finally inserted at a new location in the genome. The reverse transcription step is catalyzed by a reverse transcriptase, which in autonomous elements is encoded by the retroelement itself. Non-autonomous elements are mobilized in trans. According to their structures, retroelements are subdivided to long terminal repeat (LTR) and non-LTR retroelements.

Non-LTR retroelements

Non-LTR retroelements consist of two sub-types, long interspersed elements (LINEs) and short interspersed elements (SINEs). They are widespread in eukaryotic genomes, and

are the most abundant classes of TEs in the human genome. LINE-1 (L1) is still transpositionally active, leading to a continuous variation of mammalian genome (Hancks and Kazazian, 2012; Moran et al., 1996). The intact L1 sequence is usually 6 kb long and contains an internal promoter, two open reading frames (ORFs), and a polyadenylation signal. After RNA polymerase II transcription, mRNA processing, and export to the cytoplasm, two proteins ORF1 and ORF2 are translated from intact L1 mRNA. ORF1 is responsible for re-entry of their encoding RNA into nucleus, while ORF2 is responsible for the subsequent re-integration of the copied DNA (Kazazian and Goodier, 2002).

Different from LTR retroelements and LINEs, SINEs are transcribed by RNA polymerase III into tRNA, 5S ribosomal RNA, and other small nuclear RNAs. In primates, the most common SINEs are called Alu(s). Alu does not encode reverse transcriptase, and depends on the L1-encoded machinery for retrotransposition (Babatz and Burns, 2013). In human, there is another group of primate-specific retroelement, called SVA named for the sources of their component parts, **S**INE-R, **V**NTR, and **A**lu sequences. Similarly to Alu, SVAs can be also mobilized, and reply on L1-encoded machinery for mobilization (Babatz and Burns, 2013).

LTR retroelements

The LTR retroelements structurally resemble exogenous retroviruses, encoding viral proteins gag, pro, pol, and env, which are flanked by LTRs. The LTR retroelements are derived from retroviruses, hence called endogenous retroviruses (ERVs). According to the similarity to exogenous retroviruses, ERVs are divided into three classes: class I (gamma retroviruses, also called ERV1), class II (beta retroviruses, also called ERV2 or ERVK), and class III (spuma retroviruses, also called ERVL) (Mager and Medstrand, 2003).

Human endogenous retroviruses (HERVs) make up about 8% of the genome (Belshaw et al., 2004). In each class, HERVs are subdivided into different families that are named with single-letter amino acid abbreviations of tRNA as primers using for reverse transcription from the primer binding site (PBS) (Cohen and Larsson, 1988). HERVs have accumulated mutations (insertion or deletion) in their LTRs and encoding internal sequences. Actually, HERVs are considered defective in retrotransposition. One exception might be HERVK (HML2), the youngest and human-specific subfamily of HERVs. Interestingly, HERVK (HML2) polymorphism is observed in different human populations (Moyes et al., 2007), and the increase of HERVK (HML2) copies is also observed in some cancer cell lines from different human populations, indicating that HERVK (HML2) elements may still be capable of retrotransposition in some conditions (Dube et al., 2014; Marchi et al., 2014; Wildschutte et al., 2014). However, while HERVK

(HML2)-derived viral proteins (e.g. *gag*, *pol* and *env*) and viral particles have been detected in some cells and tissues, there was no evidence that these viral particles would generate *de novo* insertions in the human genome (Agoni et al., 2013; Jha et al., 2011; Lemaitre et al., 2014).

1.3.2 Impacts of transposable elements on the host

A maize geneticist, Barbara McClintock has discovered TEs between 1940-1950 (McClintock, 1953). Notably, at the time of their discovery and for decades thereafter, the scientific community dismissed transposons as useless or "junk" DNA (Grant, 1981). In fact, scientists now assume that TE-derived sequences make up more than 60% of the human genome (de Koning et al., 2011). This significant part of the genome consists of sequences of generated by TE-derived activities during evolution. Notably, most of the TE-derived sequences are not capable of transposition, due to the accumulated inactivating mutations during evolution. McClintock, was among the first researchers to suggest that these mysterious mobile elements might play a regulatory role on gene expression (McClintock, 1967). Thus, she called them “controlling elements” (McClintock, 1956). Only recently, in the era of next-generation sequencing and epigenome, have biologists begun to entertain the possibility that "junk" DNA might not be junk after all.

1.3.2.1 Impacts of transposable elements on host gene expression

Today, it is widely accepted that TEs - especially retroelement sequences - heavily influence the host transcriptome. How might TEs influence host gene expression?

First, *de novo* insertion of retroelements into a gene may interfere with its function, and might generate a pathogenic phenotype, including cancer or neuronal diseases (Muotri et al., 2010; Shukla et al., 2013). In human, L1 retrotransposition might be triggered in pathogenic states and would generate *de novo* insertions (Bundo et al., 2014; Shukla et al., 2013).

While, retroelements are normally suppressed in somatic cells, they can express viral peptides/proteins under pathogenic conditions. These translational products can affect the host. For example, the expression of retroviral *env* protein can modulate the immune response, and could restrict host infection by exogenous retroviruses (Antony et al., 2011), but also promote tumorigenesis (Ishida et al., 2008; Reis et al., 2013; Wallace et al., 2014). Notably, due to inactivating mutations accumulated during evolution, most ERV transcripts cannot be translated into proteins, and mainly act as long non-coding RNAs (lncRNAs)(Kapusta et al., 2013; Kelley and Rinn, 2012). Most recently, Zeng et al. reported that ERV-derived RNA and cDNA could, respectively, trigger the RNA and DNA

sensors followed by quick activation of B cells (Zeng et al., 2014). B cell activation triggers the secretion of antibodies against the pathogenic antigens, implying the unexpected roles of ERV-derived non-coding transcripts in immune activation.

Furthermore, TEs, primarily retroelements (both active and inactive), can influence host genes by providing alternative regulatory sequences, including promoters, enhancers, splice sites, and polyadenylation signals. Acting as promoters and/or enhancers, retroelements not only promote ERV transcription, but also influence the transcription of adjacent host genes (Rebollo et al., 2011; Rowe et al., 2013b). Recently, deep sequencing revealed that 7% of transcripts in human are controlled by ERVs, and 40% of these transcripts show spatiotemporal expression and also lineage-specific (Conley et al., 2008; Faulkner et al., 2009). Retroelements can also provide new splice sites that may lead to exonization and alternative splicing (Piriyapongsa et al., 2007).

1.3.2.2 Domestication of transposable elements

Notably, the new role of a TE-derived novel gene can be very different from its original function of transposition. This evolutionary process is known as transposon domestication. There are many TE-derived, domesticated genes reported from the human genome. One of the best examples is *Syncytin*, derived from the retroviral *env* gene. *Syncytin* is highly expressed in the multinucleate syncytiotrophoblast layer of the placenta, playing functional roles during placentation, supported by evidence of promoting trophoblast cell fusion *in vitro* (Frendo et al., 2003; Mallet et al., 2004). Strikingly, the domestication events generating *Syncytin* occurred more than once during mammalian evolution, and occurred independently from each other, involving different retroelements (Blaise et al., 2003; Lavialle et al., 2013). In human, *Syncytin* derived from the *env* gene of a HERVW element (Mallet et al., 2004).

Some retroelement-derived genes exhibit a lineage-specific expression pattern and distinct functions in different species. For example, *L1TD1*, a gene domesticated from the L1 retrotransposon, is identified in both mouse and human ESCs. Although *L1TD1* is highly expressed in both mouse and human ESCs, it is only required for maintaining pluripotency in human, while it has no similar role in mice (Emani et al., 2015; Iwabuchi et al., 2011; Narva et al., 2012; Wong et al., 2011).

Besides domesticated genes that encode proteins, TEs can give rise to *de novo* lncRNAs. These lncRNAs usually exhibit cell-type specificity, defining tissue or developmental phase specificity (Kapusta et al., 2013). As an example, some retroelement-derived lncRNAs have crucial functions in cell identity and fate determination (Fort et al., 2014; Lu et al., 2014).

1.3.2.3 Retroelements can rewire the regulatory network of the host

Retroelements can carry multiple binding sites for transcription factors, and tend to integrate near active genes (Brady et al., 2009; Schroder et al., 2002; Zhang et al., 2008). The majority of retroelement insertions poses threats to the host, and is counter-selected. Still certain integration loci, whose regulation could be established, would be tolerated. These elements could trans-activate neighboring genes, and spread their chromatin states nearby (Groner et al., 2010). Due to the above properties and their repetitive nature, retroelements are suitable to establish novel gene regulatory networks in the host. These regulatory networks evolve together with the host defense mechanism controlling them.

In contrast to the heavily conserved TF sites in the genome, retroelement-driven regulatory networks are species-specific, reflecting the divergent distribution pattern of retroelements in distinct species (Sundaram et al., 2014). For example, ERVK and ERV1 contribute the most binding sites for OCT4 and NANOG in mESCs and hESCs, respectively (Kunarso et al., 2010). As epigenetic modification restricts the activity of retroelements in a cell type-specific manner, retroelement-driven regulatory networks are highly suitable to define cell identity (Chuong et al., 2013). In mice, the Trim28/KAP1-mediated, ERV-targeting transcriptional control has been co-opted to regulate cellular gene expression during early development (Groner et al., 2010).

1.3.3 The impact of transposable elements on embryogenesis and pluripotency

The tight regulation of TEs is crucial to prevent insertional mutagenesis during zygotic genome activation at 2-cell stage. However, during early development a genome-wide epigenetic reprogramming occurs to release the cell plasticity for subsequent commitment (Cantone and Fisher, 2013). During early development, the reprogramming process needs global epigenetic remodeling. The genome-wide DNA demethylation takes place in two waves (Kohli and Zhang, 2013). While the demethylation of the maternal genome of the zygote is considered as a passive process, the paternal genome goes through active demethylation. The embryo becomes globally *de novo* methylated after the blastocyst stage. The second wave of global demethylation occurs in the primordial germ cells (Seisenberger et al., 2013).

The global epigenetic changes result in massive reactivation of TEs. Both active TEs and certain elements already incapable of transposition can be transcriptionally activated. To defend against the potential retrotransposition events, the host has developed multiple defense mechanisms during evolution. TEs are silenced during early embryogenesis by a

combination of various repressing pathways. Intriguingly, different TEs have different reactivation pattern during embryogenesis (Rowe and Trono, 2011). The expression peaks of various TEs, exhibit developmental stage-specific patterns, implying that TE reactivation might be taken as hallmarks for certain stages of early development (Goke et al., 2015; Wang et al., 2014a). It is widely accepted that TE expression and repression are linked to the control of cellular genes through development (Peaston et al., 2004).

1.3.3.1 Reactivation of transposable elements during embryogenesis

In mice

Retroelements are relatively active in rodents, causing close to 10% of spontaneous mutations in inbred strains of mice (Maksakova et al., 2006). IAP, one of the active families of ERVs in mice, is expressed in the oocyte, but its expression declines dramatically after fertilization, and then peaks again at the blastocyst stage (Gifford et al., 2013). Another active family of ERVs, MusD/ETn is highly reactivated in postimplantation embryos (Loebel et al., 2004). Interestingly, transcriptional regulation of IAPs directly affects the pluripotency of mESCs (Ramirez et al., 2006). A recent comprehensive transcriptome analysis in mESCs suggests that a large population of *de novo*, ERV-derived transcripts (mainly non-coding) are affecting or required for maintaining pluripotency (Fort et al., 2014).

The reactivation of active TEs that are capable of transposition can be detrimental for embryogenesis. Indeed, L1 activity in oocyte was recently correlated with the fetal oocyte attrition (FOA) (Malki et al., 2014). FOA is an evolutionary conserved mechanism to eliminate early meiotic oocytes. Malki et al. reported that increased L1 activity induces cell-death in meiotic prophase I oocytes (Malki et al., 2014). Accordingly, mice lacking Mael that plays a role in transposon silencing have a shortened reproductive lifespan. Nevertheless, this study also suggests that limited L1 activity might be beneficial for selecting oocytes that are best suited for next generations by FOA.

Surprisingly, certain families of inactivated TEs are required for the normal development. As an example, ERVL-associated transcripts expressed at 2-cell stage are assumed to be crucial for totipotency in mice (Kigami et al., 2003; Svoboda et al., 2004). ERVL starts to be transcribed after fertilization, touch the peak at 2-cell stage, and then quickly get silenced at the blastocyst stage in mice. ERVL reactivation induces many 2-cell-specific transcripts, most of which are directly promoted by LTRs of ERVL, or are chimeric transcripts derived from ERVL (Macfarlan et al., 2012).

In mice, the second wave of global DNA demethylation occurs in primordial germ cells (PGCs) (E12.5-E13.5), followed by again the reactivation of retroelements (Hackett and

Surani, 2013). However, different from the blastocyst and mESCs, ETns are highly expressed in PGCs, while IAPs and several ERV1 (e.g. RLTR4) subfamilies are highly expressed in gonadal somatic cells (Liu et al., 2014). Once *Setdb1* is depleted, the transcription of a subset of IAPs and ETn are enhanced, whereas L1s are not significantly affected. Notably, the reactivation of retroelements exhibits sexual differences (Liu et al., 2014), raising the possibility that retroelements might have a direct influence on the PGC formation.

In human

Compared to mice less is known about the early phase of human development, regarding the reactivation patterns of TEs. In human sperm and oocytes, most retroelements are silenced owing to the DNA methylation, except L1 (Molaro et al., 2011). The expression levels of DNA transposons are relatively high in human early embryos (e.g. zygotes, 2-cells stage). Similarly to mice, ERVL transcription is gradually decreased during early development, while ERV1 and ERVK are reactivated from the 4-cell stage, and peak at the blastocyst stage (Smith et al., 2014). SVA, still capable of retrotransposition, is highly expressed, but restricted to a developmental window at morula stage (Guo et al., 2014). The transcription level of primate-specific L1 reflects their evolutionary age: the younger L1s are more highly expressed than the older ones at the blastocyst stage (Guo et al., 2014; Smith et al., 2014). According to the RNA-seq data of human early embryos (Yan et al., 2013), at morula and blastocyst stages the most of TE-derived transcriptional reads derive from the primate-specific HERVH (Wang et al., 2014a). Indeed, HERVH expression, including HERVH-driven lncRNAs and chimeric transcripts is most restricted to the pluripotent state.

1.3.3.2 Reactivation of transposable elements during induced pluripotency

Reprogramming of somatic cells is a reverse process compared to the embryo development. During the reprogramming process, the epigenetic status of somatic cells is remodeled in order to reset a pluripotent state. This epigenetic remodeling results in the reactivation of multiple TE families.

In mice, the transcription of IAPs, MusD and L1 are gradually increased during reprogramming, and peak at the late stage of reprogramming, while ERVL is just moderately transcribed during the whole process (Friedli et al., 2014). In the matured miPSCs, the expression of L1 and MusD remains at the highest level, while IAP transcription is mostly repressed, but still higher than in the parental somatic cells. The expression pattern of TEs in matured miPSCs and mESCs are similar (Friedli et al., 2014).

Interestingly, ERVL can also be reactivated *in vitro*. In serum condition ERVL is reactivated in a rare subpopulation of mESCs, while 2i condition erases the ERVL-positive subpopulation (Macfarlan et al., 2012). These ERVL-positive mESCs, sharing similar transcriptome with 2-cell stage, are totipotent, and can form fetus and placenta in a chimeric mice assay (Macfarlan et al., 2012).

Importantly, a failure to reactivate TE-enriched regions might interfere with the reprogramming process and affect the acquired pluripotency negatively. Recently, Matoba et al. reported that the quite low cloning efficiency by SCNT is associated with the resistance of TE-enriched regions to reprogramming (Matoba et al., 2014). They found that H3K9me3 were enriched at “reprogramming resistant regions (RRR)”, consisting primarily of TEs (e.g. ERVL, IAP, and L1 loci). These RRRs were associated with low expression level of ERVL (Matoba et al., 2014). Strikingly, the cloning efficiency could be improved by removing H3K9me3 at RRRs, reflecting the influence of TEs on the early development (Matoba et al., 2014).

In human, TE transcription is quickly reactivated, and gradually increased to the peak before hiPSCs are matured. In the matured hiPSCs, the expression levels of most of TEs are decreased back to the levels observed in hESCs (Friedli et al., 2014). Several (~400) copies of HERVH, a few genomic loci of HERVK(HML2) and L1 form exceptions. These TEs, similarly to hESCs are still highly expressed in hiPSCs, though their transcription levels are a bit lower than the pre-iPSCs. Those transcriptional active HERVK loci seem to be controlled by their neighboring regulatory elements, but not their own LTRs (Wang et al., 2014a).

1.3.3.3 Epigenetic regulation of transposable elements in embryos and pluripotent stem cells

In mice

The DNA of retroelements is preferentially methylated compared to the genome. TEs are controlled by CpG methylation by DNA methyltransferases including Dnmt3a, Dnmt3b and Dnmt1. *De novo* methylation of retroelements is re-established by Dnmt3a, Dnmt3b and Dnmt3l following the global genomic de-methylation in cleavage embryos and in PGCs (Lee et al., 2014b). Once established, the DNA methylation pattern of ERVs will be maintained by DNA methyltransferase, Dnmt1 during early embryogenesis and germline development (Gaudet et al., 2004).

The most active ERVs in mice are MusD/ETn and IAP (Lueders and Kuff, 1977). The repression of IAP ERVs in the mouse genome by DNA methylation mainly occurs in the postimplantation embryos and in differentiated cells (Howard et al., 2008; Hutnick et al.,

2010). The hypermethylation of IAPs is mainly regulated by Dnmt1 (Hutnick et al., 2010; Kato et al., 2007; Okano et al., 1999). However, depletion of *Dnmt1* in mESCs does not lead to dramatic elevation of IAP transcription, suggesting that alternative mechanisms are employed to maintain the silencing of IAPs in the preimplantation stage and in mESCs (Hutnick et al., 2010). Indeed, a few regions (mainly the imprinted genes and IAPs) in the mouse genome resistant to demethylation are marked with the repressive histone mark, H3K9me3, indicating the interplay between DNA methylation and trimethylation of H3K9 (H3K9me3). This suggests that in addition to DNA methylation, histone modification and consequent changes of the chromatin state also modulate transcription/silencing of ERVs during the early developmental stages.

To protect against retroelement activation, histone methylation complements DNA methylation. A DNA and repressive histone methyl marks together lead to the assembly of more compact chromatin. Generally, genome-wide maps of histone marks and DNA methylation have shown obvious linkage at TEs (Meissner et al., 2008; Mikkelsen et al., 2007; Okitsu and Hsieh, 2007). Specifically, the methylation of H3K9 and H3K27 are positively correlated, while H3K4 methylation is negatively correlated with the retroelement repression. In mESCs, retroelements, especially L1 and ERVs (class I and II), are marked with H3K9me2 and/or H3K9me3 (Mikkelsen et al., 2007). It has been known that H3K9 methylation is catalyzed by six KMTases such as Suv39h1 (also known as Kmt1a), Suv39h2 (Kmt1b), G9a (Kmt1c or Ehmt2), Glp (Kmt1d or Ehmt1) and Setdb1 (Kmt1e or Eset), all of which belong to the *Suv39* family of SET domain-containing proteins (Krishnan et al., 2011). These KMTases add one or more methyl groups to the ε-amino group of H3K9 to form dimethylated H3K9 (H3K9me2) and trimethylated H3K9 (H3K9me3). These KMTases have complementing roles. Though G9a catalyzes dimethylation of H3K9 at ERV regions, it is dispensable for the repression of ERVs in mESCs (Leung et al., 2011). Instead, G9a is required for *de novo* DNA methylation of ERV regions (Leung et al., 2011). Setdb1 catalyzes trimethylation of H3K9 in class I and II ERVs, and essential for their repression in the mouse blastocyst and ESCs. However, Setdb1 is not required for retroelement silencing in postimplantation embryos and committed cells (e.g. fibroblasts), providing a spatiotemporal control of the transcription of Class I and II ERVs (Karimi et al., 2011; Matsui et al., 2010). Interestingly, only intact retroelements (e.g. full-length L1 and ERVs) are marked with *Suv39h1/2*-dependent-H3K9me3, which safeguard the repression of intact L1 subfamily in mESCs (Bulut-Karslioglu et al., 2014).

H4K20me3 is also highly enriched at the ETn/MusD and IAP loci and overlaps with H3K9me3 (Kourmouli et al., 2004; Martens et al., 2005; Mikkelsen et al., 2007; Schotta et

al., 2004). While, H4K20me3 is significantly reduced in *Setdb1*-depleted mESCs (Matsui et al., 2010), no upregulation in ERV transcription was observed in mESCs deficient in the H4K20 KMTases *Suv420h1* and *Suv420h2* that caused the near complete loss of H4K20me3 at ERVs (Matsui et al., 2010). Thus, H4K20me3 seems not to be essential for the repression of ERVs. Intriguingly, H3K27me3 enrichment gets increased in some ERV loci while H3K9me3 levels is decreased in *Suv39-/-* mESCs (Lehnertz et al., 2003; Peters et al., 2003). These observations indicate that H3K27me3 may play a complementary role in controlling ERV transcription. Indeed, the polycomb group complex proteins—responsible for H3K27 trimethylation—are involved in silencing of ERVs (Golding et al., 2010; Leeb et al., 2010). Histone deacetylation has also been shown to be involved in silencing of ERVs (e.g. IAPs) in mESCs (Rowe et al., 2010) and L1 constructs in human embryonic cells (Garcia-Perez et al., 2010), but the mechanism needs further elucidation.

One pathway that links specificity to silencing machinery involves the Trim28 [tripartite motif protein 28, also known as KAP1 (KRAB-associated protein 1)] corepressor and its cofactors (Matsui et al., 2010; Rowe et al., 2010; Wolf and Goff, 2007, 2009). The specificity of Trim28-mediated silencing is provided by the DNA-binding Krüppel-associated box domain (KRAB)-containing zinc finger proteins (KRAB-ZFPs) (Jacobs et al., 2014; Maksakova et al., 2013; Schultz et al., 2002). KRAB-ZFPs bind the conserved sequence termed the primer-binding site (PBS), specific to the retroelements, while the KRAB domain is required for recruiting Trim28 (Thomas and Schneider, 2011). Remarkably, KRAB-ZFPs, since their appearance in tetrapods, have been under strong positive selection (Emerson and Thomas, 2009; Thomas and Schneider, 2011), and subject to rapid expansion, reflecting a continuous battle between the host and viruses. Mammalian genomes encode for hundreds of KRAB-ZFPs, and target specific DNA sequences through their zinc finger motifs (Najafabadi et al., 2015; Thomas and Schneider, 2011). The cofactor Yin Yang 1 (YY1) and ZFP809 can also bind the proviral LTRs of many retroviruses and enhance the recruitment of Trim28 (Schlesinger et al., 2013; Wolf and Goff, 2009; Wolf et al., 2015). In both mouse and human ESCs, the KRAB-ZFPs dock Trim28 at TEs. Trim28 serves as a scaffold and recruits Setdb1, histone deacetylases and HP1 (heterochromatin protein 1), which collectively induce transcriptional repression (Maksakova et al., 2013; Schultz et al., 2002). In contrast, ERVs are not re-activated by the disruption of *Trim28* in mouse embryonic fibroblasts (Rowe et al., 2010). It is assumed that once silenced in development, ERVs remain controlled in adult tissues by DNA methylation (Martens et al., 2005; Meissner et al., 2008; Mikkelsen et al., 2007). Thus, during early development (preimplantation stage), the key mechanism to silence ERVs is histone modification, while DNA methylation is

often present, but dispensable. The further enrollment of DNMTs by Trim28 results in permanent silencing marks, which are subsequently maintained throughout the development (Quenneville et al., 2012; Rowe et al., 2013a).

In contrast, the ancient, no longer active, ERVL elements (Class III) are regulated slightly differently from Class I and II ERVs. ERVL is characteristically and massively expressed at 2-cell stage of mouse embryos (Kigami et al., 2003; Svoboda et al., 2004). ERVL is not regulated by Setdb1, but by the H3K4 demethylase Kdm1a (also called Lsd1). The loss of *Kdm1a* in mESCs results in an increase of H3K4me3 and H3K27ac at ERVL loci, accompanied by the enhanced transcription of ERVL (Macfarlan et al., 2011; Macfarlan et al., 2012; Maksakova et al., 2013). These data indicate that distinct histone modifications by different histone modifiers regulate different classes of ERVs.

Small RNAs are also able to regulate diverse families of highly polymorphic TEs. The germline specific Piwi-interacting RNA (piRNA) pathway was first characterized in *Drosophila* (Aravin et al., 2001), however Piwi-like proteins (Miwi, Mili and Miwi2) are also found in mice (Girard et al., 2006). piRNAs are 24–30 nt long sequences that are bound by Piwi or Aubergine proteins and target mRNA from TEs. Importantly, the small RNA-based silencing mechanism is cross talking with H3K9me3 histone modification (Brower-Toland et al., 2007; Pezic et al., 2014). While the Piwi-mediated silencing mechanism is Dicer-independent, Dicer-dependent small interfering RNAs (siRNAs) that are produced from long double-stranded RNAs also cooperate to suppress TEs during development (Ciaudo et al., 2013).

Furthermore, APOBEC (apolipoprotein B mRNA-editing enzyme and catalytic enzymes) proteins serve as important post-transcriptional ERV blockers in early embryos and in germ cells (Wissing et al., 2011). The APOBEC proteins act during the first steps of reverse transcription, and can introduce multiple cytosine-to-uracil changes by deamination of the retroviral negative-strand DNA (Harris et al., 2003; Mangeat et al., 2003; Richardson et al., 2014; Yang et al., 2007). Such deaminated retroelement is subject to degradation by uracil removal and abasic nuclease cleavage.

L1 may be also repressed by DNA-repair (Gasior et al., 2008). Furthermore, TEs possessing bidirectional promoters may even block themselves by antisense transcription.

In sum, various defense mechanisms, responsible for repressing TEs during early development form a complex regulatory network. The different repressing mechanisms complement each other and seem to control TEs in a specific temporal and spatial

pattern. Intriguingly, recent studies suggest that some cellular genes crucial during early phase of development are co-regulated with TEs.

In human

To date most of the research regarding the mechanisms TE activation and silencing are performed in mice, and less is known about the human scenario. Compared to mice, significantly less active elements are present in the human genome. Some studies in human cancer cells suggest that in normal somatic cells retroelements (e.g. L1 and certain subfamilies of ERVs) are regulated by DNA methylation (Muotri et al., 2010). These observations would indicate that the basic mechanism by which retroelements are repressed during human early development and in pluripotent stem cells is evolutionally conserved. The main difference between mouse and human would reflect the differential set of TEs, accumulated species-specifically during evolution.

Recent single cell deep sequencing data show that while the transcription from certain retroelements do reactivate in human preimplantation embryos, the repressed ones are positively correlated with DNA methylation and repressive histones marks (especially H3K9me3) (Guo et al., 2014; Smith et al., 2014; Smith et al., 2012; Wang et al., 2014a). Similarly to mice, TRIM28 and H3K9me3 were confirmed to repress retroelements (e.g. SVA, L1, subfamilies of class I and class II ERVs) in human pluripotent cells (Jacobs et al., 2014; Turelli et al., 2014). In turn, upon *TRIM28* depletion, the increase of transcription of these retroelements is accompanied by replacement of H3K9me3 with active histone marks (e.g. H3K4me1/3) (Turelli et al., 2014).

Similarly to mice, different TEs seem to be regulated by alternative pathways. For example, in addition to TRIM28-based repression, L1 is also repressed by the small RNA-based RNAi system, and by APOBEC proteins (Marchetto et al., 2013; Muckenfuss et al., 2006; Wissing et al., 2011). Interestingly, while the different pathways complement each other, they act on evolutionarily distinct sets of elements. Recent work has revealed that in hESCs, the TRIM28 system represses a discrete subset of L1 lineages predicted to have entered the ancestral genome between 26.8 million and 7.6 million years (MYA) ago (Castro-Diaz et al., 2014). In contrast, the small RNA-based, PIWI-piRNA (PIWI-interacting RNA) pathway seems to suppress the youngest L1 lineages (Castro-Diaz et al., 2014). Furthermore, TRIM28-mediated repression targets SVAs (older than 3.5 MYA), while the younger, human-specific SVA lineages are less frequently regulated (Turelli et al., 2014). Class II HERVs that were endogenized after humans and chimpanzees diverged (<7 MYA), are controlled by TRIM28. However, Class III HERVs, the oldest identifiable HERVs, which can be traced back to some 100 MYA, are not controlled by TRIM28 any longer, probably due to the accumulation of mutations around their TSSs,

indicative of extensive purifying selection (Turelli et al., 2014). Their DNA might already be irreversibly silenced by methylation.

2 Objectives

In my thesis, I focus on the relationship between retroelements and the pluripotency in human.

2.1 Generation of human induced pluripotent stem cells using the *Sleeping Beauty* transposon system

As the first one reconstructed DNA transposon, the *Sleeping Beauty* transposon is the most thoroughly studied vertebrate transposon to date, and it has shown efficient transposition in mammalian cells (Ivics et al., 1997). The *Sleeping Beauty* transposon system is composed of a transposase source and a transposon vector flanked by inverted terminal repeats (ITRs) (Ivics et al., 2004). In the presence of the transposase, the transposon is mobilized by the "cut and paste" transposition. During the transposition reaction, the transposon is excised and then gets reintegrated into a different locus in the genome (Izsvak and Ivics, 2004). The *Sleeping Beauty* transposon system can be used as a non-viral gene delivery tool, and thereby opens up new possibilities for genetic manipulation in cells, animal models as well as for human gene therapy.

Compared with the viral vectors currently in use, the *Sleeping Beauty* transposon has many favorable advantages as a gene delivery system, including its reduced immunogenicity (Yant et al., 2000), relaxed limitation on the size of expression cassettes (Wang et al., 2014b; Zayed et al., 2004), and improved safety/toxicity profiles (Huang et al., 2010a; Ivics et al., 2007). In comparison to the *PiggyBac* transposon whose integration features resemble retroviral vectors (de Jong et al., 2014), the *Sleeping Beauty* transposon integration occurs fairly randomly in the genome (Huang et al., 2010a). In addition, by means of SB100X, a novel hyperactive transposase developed in our lab (Mates et al., 2009), the *Sleeping Beauty* transposon shows robust transposition activity. Furthermore, the *Sleeping Beauty* transposon seems to trigger significantly milder epigenetic changes at the genomic insertion sites (Zhu et al., 2010a). More importantly, the *Sleeping Beauty* transposon is capable of supporting stable, long-term transgene expression both *in vitro* (Grabundzija et al., 2010) and *in vivo* (Mates et al., 2009) with lower transgene silencing than viral vectors.

As described in the "Introduction" part, the reprogramming efficiency of somatic cells, especially in human, using non-integrative methods is quite low, while the integrated approaches like viral vectors and the *PiggyBac* transposon may raise the safety concerns by disrupting endogenous genes (e.g. reactivating oncogenes or repressing tumor

repressors) in iPSCs. Given the advantages of the *Sleeping Beauty* transposon, I tried to establish an alternative approach for reprogramming both mouse and human fibroblasts into iPSCs, based on the *Sleeping Beauty* transposon system. Importantly, when compared to the *PiggyBac* transposon, *Sleeping Beauty*-transfected cells exhibit fewer numbers of integrants per cell, and it is easier to derive single-copy integrants (Huang et al., 2010a). The Cre-loxP system is integrated into the *Sleeping Beauty* transposon system, in order to exchange the Yamanaka factors with the genes of interest. Using this combined method, it is expected that the monogenetic patient-derive iPSCs can be corrected with the therapeutic gene meanwhile the Yamanaka factors can be removed. Additionally, this *Sleeping Beauty*-based reprogramming method can be used to generate transgenic animals from "single-copy-insertion" iPSCs.

2.2 The roles of human endogenous retroviruses in acquisition and maintenance of human pluripotency

Recently, more and more studies show the functional roles for TEs, especially retroelements, during early development. For example, ERVL elements have been shown to specify totipotency at mouse embryo 2-cell stage and in mESCs cultured in serum condition, by fine-tuning the regulatory network of totipotent cells (Macfarlan et al., 2012). Furthermore, silencing and reactivation of retroelements affect the normal development of mouse embryos and the determination of mESC fate. However, the human transposome is quite different from mouse. Thus, how the human-specific TEs affect early development is still elusive. In this study, using the RNA-seq, ChIP-seq and genome-editing techniques, and hESCs/iPSCs as models, I tried to systematically analyze the genome-wide expression profile of TEs in human pluripotent stem cells, and explore whether and/or how TEs, particularly human endogenous retroviruses (HERVs), affect acquisition and maintenance of human pluripotency.

3 Publications

3.1 Manuscript 1

Ivana Grabundzija*, **Jichang Wang***, Attila Sebe*, Zsuzsanna Erdei, Robert Kajdi, Anantharam Devaraj, Doris Steinemann, Károly Szuhai, Ulrike Stein, Tobias Cantz, Axel Schambach, Christopher Baum, Zsuzsanna Izsvák#, Balázs Sarkadi#, and Zoltán Ivics#. *Sleeping Beauty* transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells. ***Nucleic Acids Res***. 2013; 41(3): 1829-1847.

link:
http://dx.doi.org/10.1093/nar/gks1305

In this project, Jichang Wang made contributions as follows:

1) Design and clone the vector pT2/RMCE-OSKM(L)-EOS-mCherry and pT2/RMCE-OSKM-miRNA302/367;
2) Generate miPSCs using the *Sleeping Beauty*/RMCE-based reprogramming system;
3) Characterize the integration profile of the sleeping beauty transposon in miPSCs;
4) Perform the RMCE experiment on one single-copy miPSC clone;
5) Characterize the exchanged miPSC clone by series of assays (qRT-PCR, DNA methylation);
6) Perform the *in vitro* differentiation assay on miPSCs;
7) Generate hiPSCs using the *Sleeping Beauty*/RMCE-based reprogramming system;
8) Characterize hiPSCs by series of assays (qRT-PCR, DNA methylation and immunostaining);
9) Determine silencing of the reprogramming factors in hiPSCs;
10) Perform the *in vitro* differentiation assay on hiPSCs;
11) Write parts of the manuscript.

3.2 Manuscript 2

Jichang Wang*, Gangcai Xie*, Manvedra Singh, Avazeh T. Ghanbarian, Tamás Raskó, Attila Szvetnik, Huiqiang Cai, Daniel Besser, Alessandro Prigione, Nina Fuchs, Gerald Schumann, Wei Chen, Matthew C. Lorincz, Zoltán Ivics, Laurence D. Hurst#, Zsuzsanna Izsvák#. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. ***Nature***. 2014; 516: 405-409.

link:

http://dx.doi.org/10.1038/nature13804

In this project, Jichang Wang made contributions as follows:

1) Conceive ideas for the project;
2) Design and perform experiments with the exceptions of EMSA and the confocal imaging;
3) Analyze and interpret data
4) Participate in bioinformatics analyses
5) Write parts of the manuscript.

* co-first author; # co-corresponding author

4 Discussion

During my PhD studies, I first generated iPSCs from both mouse and human fibroblasts using defined transcription factors. This protocol is based on the *Sleeping Beauty* transposon system, and generates iPSCs with the relative high efficiencies, comparable to the retroviral methods. Using the *Sleeping Beauty* transposon system, the single-copy iPSC colonies can be obtained with higher frequency than other approaches (e.g. *PiggyBac* transposon). Importantly, I have demonstrated that it is feasible to remove the reprogramming transgenes from the matured iPSCs or replace them with a gene of interest (e.g. therapeutic genes) (section 4.1).

To explore whether and/or how retroelements affect acquisition and maintenance of human pluripotency, I have used the RNA-seq technique to determine global transcription profiles of retroelements in hiPSCs, and in their parental and differentiated cells. By comparative transcriptome analysis, I did identify a primate-specific endogenous retrovirus, HERVH as being essential for human pluripotency. I have used a series of gain of function and loss of function assays to confirm my hypothesis. Importantly, I have observed that a HERVH-marked sub-population in hPSCs might represent an alternative pluripotent state, resembling the naïve mESCs. These naïve-like stem cells share key features with epiblast cells within the human blastocyst. My research raises the possibility that the ground state of human pluripotency might be captured and maintained indefinitely *in vitro*. My research indicates that HERVH transcription plays crucial roles in defining naïve human pluripotency. The presence of the HERVH-derived regulatory circuitry partially explains why human embryonic stem cells (hESCs) differ considerably from those of mice (section 4.2).

4.1 Reprogramming of somatic cells into iPSCs using the novel *Sleeping Beauty*/RMCE-based reprogramming system

4.1.1 Comparison of the *Sleeping Beauty*/RMCE-based reprogramming system with other approaches

Yamanaka and his colleagues generated iPSCs from somatic cells for the first time (2006), using a defined set of transcription factors (Takahashi and Yamanaka, 2006). Since 2006, developing alternative iPSC techniques became a very popular research area in the world. In the last decade, researchers have developed various strategies to derive iPSCs, using different gene delivery methods. Although, the originally used retroviral delivery is relatively efficient and widely used, this method might raise certain

safety issues regarding clinical translations, owing to the insertional mutagenesis of retroviral vectors. Alternatively to the viral vectors, a series of non-viral or non-integrative methods have been established (see Introduction 1.2.2.1). The price-to-pay for the non-integrative strategies include the relatively low reprogramming efficiency, due to the gradual decrease of transgene expression during the reprogramming process. Although there are multiple protocols are available to generate iPSCs, there are still problems to be solved to catch up with the enormous demand for iPSCs for drug screening, disease modeling and potential clinical applications in regenerative medicine.

Our approach, based on the combination of the *Sleeping Beauty* transposon system and Cre-mediated recombination exchange (Cre/RMCE) strategy, shows series of advantages compared to viral-based and non-integrative methods. First, the hyperactive transposase SB100X-mediated transposition of the *Sleeping Beauty* transposon supports long-term stable expression of transgenes in both mice and human cells (Mates et al., 2009; Xue et al., 2009). The sustained expression of reprogramming factors is essential to obtain authentic iPSCs with a good frequency (Hockemeyer et al., 2008). Importantly, our reprogramming system can efficiently obtain iPSCs, comparable to retroviral vectors. Second, it is considerably simpler and more cost-effective to prepare the plasmid-based reprogramming vectors in comparison to viral vectors. Unlike the mRNA-based reprogramming method, which needs multiple rounds of transfections during the reprogramming process (Yakubov et al., 2010), using the *Sleeping Beauty* transposon system requires only a single transfection. Third, genomic integrity is crucial for iPSC application in drug screening and cell therapy. The *Sleeping Beauty* transposon is originated from the fish, and there are no homologous sequences to the *Sleeping Beauty* transposon existing in the human genome. Thus, the *Sleeping Beauty* transposon is precisely integrated into the host genome without any undesired genomic effects (e.g. genomic instability). The integration profile of the *Sleeping Beauty* transposon is fairly random (Huang et al., 2010a; Huang et al., 2010b). Fourth, SB100X promotes single-copy insertion of the reprogramming cassette. The *Sleeping Beauty* transposon supports a good expression of the reprogramming transgenes even from a single copy, and would not compromise the reprogramming efficiency. Therefore, we can efficiently screen single-copy iPSC clones, and select those that are integrated into a “safe-harbor” locus (Papapetrou et al., 2011). In combination with the Cre/RMCE system, we can successfully exchange the reprogramming cassette with genes of interest in single-copy iPSCs. This strategy makes it possible to correct monogenetic, patient-derived iPSCs by exchanging the reprogramming cassette with the therapeutic gene. The integration

footprint of the *Sleeping Beauty* transposon can be used as a "molecular identity" for different iPSC clones from different patients.

When compared, the presence of *PiggyBac*-like sequences in the mammalian genome (Sarkar et al., 2003) represents an unpredictable risk, and might argue against using the *PiggyBac* transposon system in a clinical setup. Furthermore, approximately 5 % of the excision sites, generated by the *PiggyBac* transposase are not precise, and contain microdeletions (Liang et al., 2009). The *PiggyBac* transposon is also reported to share a similar integration profile with gamma-retroviral vectors, and has a bias toward transcription start sites (TSSs) of oncogenes and tumor repressors, leaving the risks of oncogene reactivation or disruption of tumor repressors (de Jong et al., 2014; Galvan et al., 2009). Finally, the *PiggyBac* transposon hardly allows single-copy insertion of the reprogramming cassette in iPSCs, without reducing the reprogramming efficiency (Yusa et al., 2009). It is much more difficult to remove the reprogramming cassette, when it is present in multiple copies. In sum, compared to the *PiggyBac* transposon, the *Sleeping Beauty*-based reprogramming system is much safer to meet the criteria of iPSC generation for therapeutic application.

However, our reprogramming system has also limitations. While, the *Sleeping Beauty* transposon system is efficient to reprogram fibroblasts, we have difficulties to efficiently translate our approach to blood cells (e.g. B and T cells). Notably, it was recently shown that integration of exogenous sequences into the host genome was restricted by host cell-cycle-related factors, which can be overcome by using small molecules (Yu et al., 2015). Therefore, by the large-scale chemical screening, it is possible to identify chemical compounds that can improve *Sleeping Beauty* transposition in blood cells.

4.1.2 Selection for the authentic hiPSCs using a pluripotency reporter

Reprogramming somatic cells using a defined set of TFs is a rare event. Usually, highly experienced researchers in hPSC culturing isolate the putative hiPSC clones based on cell morphologies. These clones are than subjected to further time-consuming characterization of pluripotency. Reporters that could ease the laborious identification and characterization work are highly appreciated by stem cell researchers. Indeed, fibroblast, isolated from transgenic mice, expressing pluripotency reporters proved to be valuable to monitor the reprogramming process. For example, GFP that is knocked into the *Oct4* locus is used to monitor the reactivation of endogenous pluripotency genes (Yoshimizu et al., 1999), and isolate matured mouse iPSC (miPSC) clones in culture. Similarly, given the large heterogeneity of hPSCs, it could be very helpful to use reporter(s) to monitor the reprogramming process of human somatic cells and select matured hiPSC clones. To

efficiently select authentic hiPSC clones, I integrated a pluripotency reporter into the *Sleeping Beauty*-based reprogramming system. In this system, the reporter is driven by EOS, a retrotransposon-derived promoter that is specifically active in the pluripotent state (Hotta et al., 2009). This pluripotency reporter can be used to monitor and optimize the reprogramming process of human somatic cells. It can also be used to efficiently isolate authentic patient-specific iPSCs from a small number of biopsies. As undifferentiated hPSCs might form teratomas *in vivo*, the detection and removal of undifferentiated hPSCs is key for therapeutic application of hPSCs. Notably, our pluripotency reporter can also be used to detect and sort out undifferentiated cells from the differentiated cell population prior to transplantation.

4.1.3 Mechanisms of somatic cell reprogramming

Somatic cell reprogramming is inherently an epigenetic remodeling event. During the process, stepwise changes of transcriptome and cell identity are accompanied by slow epigenetic reprogramming, including histone modification and DNA methylation (Lee et al., 2014b; Papp and Plath, 2011; Theunissen and Jaenisch, 2014). According to the genome-wide gene expression analysis, the reprogramming process can be divided into three phases: the early, intermediate and late stages (O'Malley et al., 2013; Polo et al., 2012; Samavarchi-Tehrani et al., 2010). These stages are distinguished by stage-specific hallmarks and molecular events.

At the early stage, the expression of genes involved in cell proliferation, metabolism, and cytoskeletal organization is upregulated, while developmental genes are downregulated (Hansson et al., 2012; Polo et al., 2012). Consistently, using the *Sleeping Beauty*-based reprogramming strategy, I observed that the mesenchymal-to-epithelial transition (MET) occurred at the early stage. This stage was characterized by upregulation of epithelial genes, while mesenchymal genes were downregulated. Notably, MET is the first rate-limiting event during the reprogramming process (Li et al., 2010; Samavarchi-Tehrani et al., 2010), but could be modulated by certain miRNAs (Liao et al., 2011). Specifically, a hESC-enriched miRNA cluster, miRNA302/367 (located in chr4) was shown to repress the TGFβ signaling pathway that induced epithelial-to-mesenchymal transition (EMT) (Subramanyam et al., 2011). Thus, miRNA302/367 overexpression acts by promoting the initial phase of reprogramming, by overcoming the TGFβ-induced EMT. Indeed, I reported that miRNA302/367 overexpression accelerated the initial phase of the reprogramming process, evidenced by small colonies formed 2-3 days earlier (Grabundzija et al., 2013). Intriguingly, miRNA302/367 overexpression might be detrimental to the formation of miPSCs. I did observe that miRNA302/367 overexpression

delayed the reprogramming process of mouse embryonic fibroblasts, and induced more flat (not domed) iPSC colonies, resembling hESCs (Grabundzija et al., 2013). Why does miRNA302/367 have opposite functions in reprogramming in mouse and human somatic cells? It might be related to the developmental feature of mESCs vs hESCs —the former represents a naïve pluripotent state, while the latter represents a primed pluripotent state. A recent report showed that ES-cell-specific cell-cycle-regulating (ESCC) miRNAs were negatively related to maintenance of naïve state in mESCs (Kumar et al., 2014). It is very interesting to explore the distinct roles of miRNA302/367 cluster in mESCs and hESCs.

At the intermediate stage, the ESC-like colonies start to be formed, but accompanied by non-ESC-like colonies. The stage-specific embryonic antigen 1 (SSEA1), a surface marker for mESCs, starts to express in a subpopulation of reprogrammed somatic cells. In these SSEA1-positive cells, many pluripotency-associated genes (e.g. *Sall4*, *Nanog*, *Utf1*, *Dppa3*) are gradually upregulated. SSEA1 expression is a marker for successful reprogramming. However, certain cells would fail to downregulate the mesenchymal genes, and would become refractory to reprogramming. These cells would express Thy1, a fibroblast-specific surface marker (Polo et al., 2012). These findings further demonstrate that MET is key for the reprogramming. Consistent with these changes in transcriptome of reprogrammed cells, our pluripotency reporter was first giving signals at end of the intermediate stage, reflecting its accuracy of reporting the activation of endogenous pluripotency genes (e.g. *Oct4*, *Sox2*). Another hallmark of the intermediate stage is the metabolic transition from OXPHOS to glycolysis (Folmes et al., 2011). This phenomenon is characterized by upregulation of glycolysis-associated genes and downregulation of OXPHOS-associated genes.

At the late stage, most ESC-like colonies will be matured, and self-renew independently from the transgene expression, evidenced by the activation of the endogenous pluripotency regulatory network. In optimal scenario, in matured iPSCs the exogenous reprogramming factors are silenced (Takahashi and Yamanaka, 2006). Interestingly, the *Sleeping Beauty*-based reprogramming cassette behaved differently in mice and human iPSCs. While in mice the transgene cassette was not silenced even in matured miPSCs. In mouse fibroblast reprogramming, although the endogenous pluripotency genes were reactivated (e.g. *Oct4*-GFP), the transgenes were still detectable, evidenced by the expression of antibiotic selection markers. This is consistent with that the *Sleeping Beauty* transposon is resistant to silencing in mESCs in which retroviral vectors are usually silenced (Ivics et al., 2014a; Ivics et al., 2014b; Ivics et al., 2014c). As the ectopic expression of exogenous reprogramming factor might interfere differentiation of miPSCs, it is necessary to remove the reprogramming cassette from the matured miPSCs.

Importantly, once the exogenous reprogramming factors are removed, these miPSCs are still capable of self-renewal and acquire full differentiation potentials (Grabundzija et al., 2013). These data indicate that the pluripotency of miPSCs, generated by the *Sleeping Beauty*-based system are independent from the transgene expression. By contrast, the reprogramming transgenes, pluripotency reporter and the antibiotic selection markers were readily silenced in the stable hiPSC colonies at the end of reprogramming process. The global gene expression analysis confirmed that the endogenous pluripotency regulatory network had been fully activated in these stable hiPSC colonies. Importantly, these stable hiPSC colonies can be differentiated into three germ layers, demonstrating that they are fully pluripotent (Grabundzija et al., 2013). Why do these reprogramming transgenes delivered by the *Sleeping Beauty* transposon exhibit distinct transcription behaviour in miPSCs and hiPSCs? I hypothesize that this phenomenon is associated with the different nature of miPSCs and hiPSCs. Mouse iPSCs represent the naïve state that is resistant to the overexpression of core pluripotency factors. Indeed, overexpression of *Oct4*, *Klf4*, and *Nanog* promotes mESCs into a synthetic super-pluripotent state (Geula et al., 2015). These super-pluripotent mESCs are refractory to differentiation even upon LIF withdrawal. In contrast, hiPSCs represent an advanced developmental stage where some differentiation program has been already triggered. In such state, pluripotency genes (e.g. *OCT4*, *SOX2*, *KLF4*) also act as lineage specifiers (Loh and Lim, 2011). Thus, ectopic expression of these pluripotency genes might destabilize the endogenous pluripotency regulatory network of hiPSCs, and initiate differentiation towards certain lineages. So, at the late stage, the pre-matured hiPSCs undergo a selection process. Only those colonies could be matured, where the reprogramming transgenes has been silenced. However, the ones that are still dependent on transgene expression would be collapsed, indicating that the transgene silencing is crucial for hiPSC maturation.

As somatic cell reprogramming is essentially a series of epigenetic reprogramming events, it is not surprised that dynamic rearrangement of the epigenetic landscape is observed during this process. At the early stage, a mark associated with euchromatin, H3K4me2 is redistributed genome-wide (Koche et al., 2011). CpG-poor pluripotency genes are activated, due to global DNA demethylation, followed by acquisition of active histone mark, H3K4me3 and loss of repressive mark, H3K27me3 at the late stage (Lee et al., 2014a). MET-associated genes with CpG-rich promoters are reactivated early in the reprogramming process via removal of H3K27me3 in their promoter regions (Lee et al., 2014a). In contrast, the fibroblast-specific and developmental/differentiation genes are repressed by DNA methylation or acquisition of H3K27me3 at the early and intermediate stages (Lee et al., 2014a; Polo et al., 2012). These findings support the view that the

global and coordinated regulation of DNA methylation and H3K27me3 histone modification are required for the successful reprogramming. While DNA methylation and H3K27me3 assist "biphasic" changes during the reprogramming process, H3K9me3 and H3K79me2/3 modification are barriers to the reprogramming process (Chen et al., 2013a; Mattout et al., 2011; Onder et al., 2012). Their gradual, but global erasure is crucial for successful reprogramming. The reprogramming kinetics of human fibroblasts can be promoted by depletion of H3K9 and H3K79 methyltransferases *SUV39H1/2* and *DOT1L*, respectively. However, depletion of *SETDB1*, another H3K9 methyltransferase, compromises the reprogramming of human fibroblasts, and decreases the reprogramming efficiency, indicating a distinct function of different H3K9 methyltransferases (Onder et al., 2012). In fact, SETDB1-mediated H3K9me3 distribution might be beneficial for reprogramming, being consistent with the finding that Setdb1 is required for the acquisition and maintenance of mouse pluripotency (Dodge et al., 2004; Lohmann et al., 2010). Notably, Setdb1 restricts reactivation of retroelements (mainly ERV1 and ERV2) (see Introduction 1.3.3.3). Therefore, one rational speculation could be that SETDB1-mediated H3K9me3 histone marks safeguard reprogrammed cells from the activation of retroelements. Thus, SETDB1 might protect against genome instability of hiPSCs that would be otherwise induced by global epigenetic changes.

4.2 Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells

4.2.1 HERVH is a specific marker for human pluripotency

While many genes are involved in the maintenance of pluripotency in PSCs, recent evidence from mouse and human has suggested that, expression owing to binding of transcription factors to transposable elements (TEs) plays an important role in the process (Kunarso et al., 2010). The transcriptional activation of TEs during embryogenesis reflects a history of successful TE invasion (which to be successful must have occurred in cells in the lineage from zygote to germ cells). While only a particular subfamily of L1 element is capable of autonomous retrotransposition in the human genome that is capable of *trans*-mobilizing other retroelements (e.g. Alu and SVA), many retroelements and their relics are only transcriptionally active. As different genomes have been invaded by different TEs, the involvement of TE-derived transcripts has the potential to establish lineage-specific transcriptional circuitry (Sundaram et al., 2014). Strikingly, different TEs have wired a different set of genes into the core regulatory network of embryonic stem cells in humans and mice (Kunarso et al., 2010). While in mice ERVL is implicated in regulating the transitional state between toti- and pluripotency (Macfarlan et

al., 2012), less is known about the role of human ERVs (HERVs) in transcriptional regulation. I thus surveyed RNA-seq data to look for TE families that are specifically abundant in the hPSC transcriptome. I have shown that several TEs are expressed at significantly higher levels in hPSCs compared to fibroblasts. The most significantly upregulated elements, HERVHs belong to the ERV1 type of LTR retroelement. HERVH elements are 15-30 million years old and present exclusively in Apes and Old World Monkeys (OWMs) (Mager and Freeman, 1995). Each copy has ~80% identity to the Repbase consensus sequence. A higher level of transcription was associated with elements containing consensus LTR7 rather than the diverged variants LTR7C, LTR7B or LTR7Y (Wang et al., 2014a). Different from HERVH, other families of retroelements, such as HERVK/HML2, HERVW/HERV17, HERVL, are slightly transcribed in certain hiPSC lines, indicating that their transcription might be sporadic. Due to the repetitive feature of TEs, analysing RNA-seq data is very challenging. To exclude possible bias when RNA-seq reads are mapped to the genome, I have suggested two strategies: including multiple mapping reads, and unique mapping. Importantly, these two types of analyses (performed by Dr. Gangcai Xie, a bioinformatic colleague) showed similar expression profiles of TEs, confirming the robustness and reliability of the RNA-seq analysis. I did also use qRT-PCR, to confirm basic findings, predicted by the RNA-seq analysis.

To address how specific HERVH transcription is to pluripotent cell types, RNA-seq data from public databases were collected to compare multiple hESC and hiPSC lines and differentiated cell lines (performed by Dr. Gangcai Xie, a bioinformatic colleague). The comparative analysis showed that HERVH was specifically expressed in hPSCs, and the vast majority of the transcribed loci were identical between hiPSCs and hESCs. Consistently with recent reports, our results suggest that HERVH is a specific marker for human pluripotency (Kelley and Rinn, 2012; Santoni et al., 2012).

4.2.2 HERVH is involved in the regulatory network for human pluripotency

The specificity of HERVH transcription in hPSCs raises the possibility that it has been recruited to the regulatory network for human pluripotency. I approached this possibility in two steps.

As I've discussed it earlier (see Introduction 1.3.3.3), the host has developed a series of mechanisms to restrict TEs, including epigenetic regulation. How does HERVH escape from the epigenetic repression in hPSCs? I did explore the chromatin landscapes of all of HERVH loci in hPSCs and in differentiated cells. Expectedly, in the pluripotent state, the active histone mark, H3K4me3 was enriched in transcribed loci of HERVH, but removed from the inactive loci. Furthermore, H3K4me3 was missing at most of the HERVH loci in

differentiated cells. In contrast, the repressive histone mark, H3K9me3 was lost at active loci, but was relatively enriched at inactive loci of HERVH in hPSCs (collaborating with Dr. Gangcai XIe who did bioinformatics analysis) (Wang et al., 2014a). As TRIM28 has been implicated to regulate TEs via H3K9me3 marks, TRIM28 is an obvious candidate that could also regulate HERVH in hPSCs. However, *TRIM28* depletion in hESCs did not result in significant and global increase of HERVH transcription (Turelli et al., 2014), suggesting that the epigenetic regulation of HERVH is more complicated than expected. One possible explanation is TRIM28 is not the primary regulator of HERVH. It is also possible that the depletion of *TRIM28* by shRNA was not complete, and the residual protein can still provide full activity on HERVH regulation. Thus, it is necessary to knock out *TRIM28* in order to elucidate its role in regulation of HERVH transcription in hPSCs. Beside these dynamic changes of chromatin landscapes, DNA methylation was also shown to regulate HERVH expression in hPSCs. DNA methylation was negatively correlated with HERVH expression in hPSCs (Wang et al., 2014a; Xie et al., 2013).

As described above, HERVH expression is heterogeneous in hPSCs, and the highest expression level is driven by LTR7 rather than the variance version LTR7Y/C/B. This difference might be associated with the binding sites they provide for the pluripotency TFs. Indeed, ChIP-seq analysis confirmed this hypothesis. At the active loci of HERVH, OCT4, NANOG and KLF4 (but not SOX2) are remarkably enriched in LTR7. Importantly, a novel TF, LBP9 (also called Tfcp2l1) recently implicated to play a role in pluripotency in mice, was identified to regulate HERVH transcription (Wang et al., 2014a). The binding of LBP9 on HERVH was identified by combined computational prediction, and was experimentally confirmed. Tfcp2l1/LBP9 has been reported to interact with Oct4, and act as a core factor for maintaining the ground state of mESCs (Martello et al., 2013; van den Berg et al., 2010; Ye et al., 2013). Importantly, HERVH transcription in hPSCs is directly affected by depleting these TFs. These results suggest that HERVH, as a crucial element, has been recruited into the transcription regulatory circuitry of human pluripotency. However, In contrast to mice in which Tfcp2l1/LBP9 genome occupancy profile does not overlap with that of canonical pluripotency factors Oct4, Sox2, and Nanog (Chen et al., 2008; Martello et al., 2013), LBP9 and OCT4/NANOG/KLF4 are clustered on HERVH. These findings indicate that HERVH provides TF binding sites for the key pluripotency factors, and due to its repetitive nature might form a novel regulatory circuitry.

Gain of function and loss of function assays confirmed that HERVH had a biological function on the acquisition and maintenance of human pluripotency. My results are consistent with other reports, showing the gradually increased of HERVH transcription during the reprogramming process (Friedli et al., 2014). In agreement with previous

studies, I reported that many hPSC-specific lncRNAs were derived from HERVH, two of which, LINC-ROR, and LINC00458, have been shown to promote reprogramming and maintain the hPSC identity (Kelley and Rinn, 2012; Loewer et al., 2010; Ng et al., 2012). Intriguingly, these HERVH-derived lncRNAs have a conserved region that might be functional. Indeed, the conserved region of HERVH-derived lncRNAs has positive impact on reprogramming. In this study, a novel HERVH-derived transcript, *ESRG* was identified to be required for human pluripotency (Wang et al., 2014a). *ESRG* is present in primates, and annotated as an lncRNA. However, unlike in other primates, a long putative open reading frame is embedded in the human *ESRG*, implying its coding potential. Thus, *ESRG* seems to be a HERVH-derived, human specific protein-encoding gene. Besides HERVH-derived lncRNAs, a list of hPSC-specific *de novo* transcripts were discovered in my study, all of which were associated with HERVH transcription. For example, *SCGB3A2* was transcribed into different isoforms controlled by LTR7 as its alternative promoter. In some loci, HERVH was also partially exonized and fused with the neighboring genes, to form chimeric transcripts. Most of these chimeric transcripts were hPSC-specific, and related to the biological functions of stem cells, indicating that they had potential roles in maintenance of hPSC identity (Wang et al., 2014a). It is unclear if these HERVH-associated transcripts are capable of translation into proteins. Recently, several reports showed that putative lncRNAs could be translated into proteins that have biological functions (Anderson et al., 2015; Bazzini et al., 2014; Magny et al., 2013). It will be interesting to explore the coding potential of these HERVH-associated transcripts in hPSCs and during differentiation.

4.2.3 HERVH-derived transcription defines naïve-like state of hPSCs

The view that pluripotency is not a restricted state but rather a spectrum that ranges from naïve to more developmentally primed states is universally accepted. The term, naïve state means that stem cells are free of developmental and epigenetic restriction, similar to the ground state epiblast cells in preimplantation embryos (Nichols and Smith, 2009). While, the naïve state has been well characterized in mice, it is quite difficult to reproduce it in human. Although, hESCs exhibit some molecular and biological features similar to naïve mESCs, the wide species-specific genetic differences might explain why the mouse naïve state cannot be reproduced in other species (e.g. human).

As discussed above, TEs, especially retroelements, have been recruited into the pluripotency network as regulatory elements. In my study, HERVH has been confirmed to be required for maintenance of human pluripotency. While hPSCs are generally more similar to primed mouse EpiSCs, the high expression of HERVH marks naive-like stem

cells. Importantly, single-cell RNA-seq reveals that HERVH transcription is remarkably enhanced in epiblast cells within human blastocyst embryos, implying that HERVH might define the naïve state of human pluripotency (Wang et al., 2014a). To explore this possibility, the HERVH-based reporter, pT2-LTR7-GFP#2 was integrated into the genome of either mouse or human PSCs by the *Sleeping Beauty* transposon-mediated gene transfer, providing stable transgene expression. While all of mESC colonies homogeneously express GFP, only ~4% of cells in each hESC colony show a strong GFP signal (GFP^{high}), indicating cellular heterogeneity. The rest of the cells show either weak GFP signal (GFP^{low}) or GFP silencing (GFP^{-}). Interestingly, when comparing HERVH transcription in different single cells from the same hESC clones, HERVH expression also exhibits heterogeneity (Wang et al., 2014a). Correlation analysis reveals that HERVH transcription level is positively correlated with several pluripotency-associated genes, including naïvety-associated TFs. This observation might indicate that the GFP^{high} subpopulation represents a naïve-like state. To examine this possibility, a series of analysis were performed involving GFP^{high} and GFP^{low} cells. Remarkably, GFP^{high} cells can be cultured *in vitro*, forming tight, uniformly expressing 3-dimentional (3D) colonies characteristic of naïve mESCs, while GFP^{low} cells are maintained in only the conventional hESC condition rather than the naïve condition.

In mice, reactivation of X chromosomes is taken as the hallmark for naïve state of stem cells, which is restricted into the female epiblast within the blastocyst and in naïve mESCs (Bao et al., 2009; Mak et al., 2004; Okamoto et al., 2004). Although, X reactivation is also observed in human early embryos, it occurs more dynamically (Okamoto et al., 2011). Interestingly, reactivated X chromosome would be silenced again during the transition from GFP^{high} to GFP^{low} cells, indicating that GFP^{high} cells have different epigenetic landscapes from GFP^{low} cells, consistent with recent reports on naïve human pluripotency (Gafni et al., 2013; Takashima et al., 2014). Intriguingly, the Jaenisch lab has generated an alternative lineage of human naïve-like hPSCs, where the X chromosome was not reactivated, but showing similar transcriptional profiling to naïve mESCs (Theunissen et al., 2014). How is this possible? I assume there might be two possibilities: 1) X reactivation is not the hallmark for naïve human pluripotency; 2) there are alternative synthetic pluripotent states, possibly representing the different developmental stages of human pluripotency *in vivo*.

Cross-species comparison of genome-wide transcription profiling between GFP-marked cells and other published lineages, shows that GFP^{high} clusters with previously reported naïve hPSCs and mESCs, while GFP^{low} clusters with previously reported primed hPSCs and mouse EpiSCs (Gafni et al., 2013; Hanna et al., 2009a). Importantly, the transcription

profile of GFP^{high} cells is the closest to human ICM (Vassena et al., 2011), compared with any of the previously reported naïve hPSCs (Wang et al., 2014a). This would indicate that GFP^{high} cells exhibit the transcriptional profiling of naïve pluripotency. However, one should note that the term naïve was used to characterize ground state pluripotency in mice, and due to the HERVH-derived regulatory circuitry, it should be different from mice. In fact, the presence of the primate-specific, regulatory network raises the issue, whether naive mESCs can be taken as 'golden standard' for naïve human pluripotency.

4.2.4 The 'golden standard' for evaluating naïve human pluripotency

Murine naïve ESCs have a series of unusual properties: they are released from the epigenetic restriction, X chromosomes are active, they form 3D rounded clusters, resembling the E4.5 epiblast of preimplantation blastocyst (Boroviak et al., 2014), and they don't expresses genes typical of differentiated cells. Recently, the global transcriptional and epigenetic analyses shows that the epiblast cells within the mouse blastocyst share many similar properties with human (Guo et al., 2014; Smith et al., 2014; Yan et al., 2013), indicating that there are, to some extent, conserved signaling pathway and regulatory network governing the pluripotency.

However, many studies reveal species-specific features during early development and in ESCs. As mentioned above, to maintain pluripotency, naïve mESCs require LIF, while the differentiation stimuli are suppressed by inhibiting the Erk/MAPK and GSK3β signaling pathways, with small molecule inhibitors (2i condition) (Ying et al., 2008). In contrast, human naïve-like cells and primed hESCs are dependent on bFGF and/or Activin/TGFβ signaling (Chan et al., 2013; Gafni et al., 2013; Theunissen et al., 2014; Ware et al., 2014). *In vivo* studies also show that the mechanisms of lineage specification in human embryos differ remarkably from mice: the human blastocyst is not sensitive to the inhibition of FGF/MAPK signaling pathway in the same manner as the mouse blastocyst (Kuijk et al., 2012; Roode et al., 2012; Yamanaka et al., 2010). My results demonstrate that primate-specific retroelements are involved in defining human pluripotency. My findings collectively suggest that the regulatory circuitry for pluripotency is not as evolutionally conserved as one might expect. Therefore, using naïve mESCs as the gold standard for derivation and evaluation of naïve hPSCs might be shoehorning the latter into a mouse-shaped box. It raises the question: how to define and evaluate the naïve human pluripotency?

Given that naïve pluripotency is defined by the features similar to the epiblast cells within the blastocyst, and reflects the ability to self-renew while retaining the potential for unbiased differentiation and germline contribution, naivety could be defined by

functionality. If so, what should be the criteria? Behavior within a chimera is thought to be one of the most stringent functional assays. Consistent with this view, in contrast to EpiSCs, naive mESCs can efficiently integrate into the ICM of blastocysts and generate normal chimeras, indicating their full developmental potential.

However, there are issues to consider when using human-mouse chimeras to assess functionality of 'naïve' hPSCs. Most importantly, given the species-specific differences substantiated by our work, it isn't clear how to interpret the behavior of human ESCs in the context of mouse blastocyst. Indeed, the current human-mice chimeras demonstrate no more than human naïve-like cells can engraft into mouse blastocyst at higher frequency than conventional hESCs (Gafni et al., 2013; James et al., 2006). However, as the engrafted cells do not develop, engraftment alone does not demonstrate functionality *in vivo* (e.g. the ability to form functional organs). Furthermore, a recent study could not reproduce the human-mouse chimera assay using their own and Hanna's human naïve-like cells (Theunissen et al., 2014), – indicating that this method presents reproducibility issues when used to evaluate the function of human naïve-like cells *in vivo*.

One might suggest that tests on chimeras involving closer relatives to human might make for a cleaner interpretation, with fewer potential incompatibilities. However, a recent study on the rhesus monkey chimeras suggests that using primate pluripotent cells (including ICM) may not be feasible, due to the species-specific nature of primate embryos (Tachibana et al., 2012). Importantly, ethical issues here are foremost.

I suggest that transcription and epigenetic profiles closest resembling ICM/epiblast cells in the human blastocyst might be a more useful metric. To this end, the cells enriched using the HERVH reporter are good models of naïve cells as they cluster nearest to the ICM when compared with the 'novel naïve' cells obtained by other groups.

However, I need to emphasize that HERVH-driven transcriptional profiles in the current naïve-like hPSC lines (including the GFP^{high} cells) are slightly different from the human ICM, suggesting that these naive cell lines need further optimization. Our HERVH-based reporter could be a powerful tool for enabling optimization of naïve-like hPSC culture conditions. As shown in this study, HERVH hyper-activation is required for the maintenance of human pluripotency *in vitro*, but it may interfere the differentiation of hPSCs. In conjunction with single-cell next-generation sequencing, the efforts to characterize the transcriptome and epigenome of human early embryos will be helpful to deeply understand the nature and specificity of human pluripotency. It may be anticipated to derive a novel stable state of hPSCs that can closely mimic the human ICM.

4.2.5 How artificial the naïve-like ESCs cultures are

The purpose of culturing hESCs was always to expand cells under well-controlled conditions, without loosing their pluripotency. However, the concept of keeping cells long-term *in vitro* is rather artificial in itself. Importantly, cells possessing superior culturing properties might keep their pluripotency, but would not be identical to any true developmental stage. In fact, recent studies performed in mice have shown that there are multiple, alternative stages of pluripotency (Tonge et al., 2014).

In human, the optimization strategies to establish naïve-like cultures generated various naïve-like hPSC line (Chan et al., 2013; Gafni et al., 2013; Takashima et al., 2014; Theunissen et al., 2014; Ware et al., 2014). Importantly, the approaches to mimic naïve mESCs resulted in naïve-like cultures, where the mouse pluripotent features are dominant, while the primate ones were suppressed. It is not clear at the moment, whether further optimization attempts to generate naïve hPSC cultures, being more similar to their murine counterparts would solve the hPSC culturing problems.

The specific expression pattern of HERVH, characteristic of hPSCs recapitulates the features of the ICM of blastocyst. High-levels of HERVH expression not only mark cells in a naive state, but also apparently play a role in maintenance of this state, while inhibiting differentiation *in vitro*. However, I need to emphasize that in cultured hESCs both the number and the intensity of HERVH-expression is higher when compared to ICM. In fact, too high a level of HERVH expression in the ICM may interfere the normal embryo development, as hyper-activation of HERVH in hiPSCs causes defect of differentiation toward neuronal lineage (Koyanagi-Aoi et al., 2013). By tuning down HERVH expression the cells might differentiate normally. The lack of diapause behavior in human embryos suggests that hyper-activation of HERVH/LTR7 during the postimplantation developmental stages would not be adaptive**.**

4.2.6 Human pluripotency and host defense

What seems to be a real enigma is why does evolution tinker with something that doesn't obviously need tinkering with? How did we evolve a circuitry particular to us, humans? We know that some transcription factors related to LBP9 are important in suppressing viruses in primates (Parada et al., 1995; Romerio et al., 1997; Yoon et al., 1994) – perhaps this is at the heart of the conundrum?

The role of LBP9, comparable to HERVH, is to support self-renewal, while transcriptional down-regulation of either LBP9 or HERVH potentiates differentiation. Being central to the activity of HERVH in early embryos LBP9 functions as a switch that regulates HERVH. In

mESCs, the knockdown phenotype of *Tfcp2l1/LBP9* is less severe, and does not affect the self-renewal (Ye et al., 2013), but rather the differentiation potential, suggesting that LBP9 has a slightly different function in human vs mouse.

In fact, by clustering binding sites of the key pluripotency factors, including LBP9, HERVH was expressed at the key times to define pluripotency. However, the host defense against HERVs had to develop a way to repress HERVH expression. While, LBP9 is associated with nonAUG codon usage (Zhou et al., 2000) and pluripotency (Martello et al., 2013; Ye et al., 2013) in mammals, it seems to be engaged with ERV regulation only in primates. Perhaps, LBP9 in conjunction with other members of the LBP family (Parada et al., 1995) has been adopted to defend the host against retroviruses, and now also functions as a repressor of HERVH. If so, then HERVH was recruited to the pluripotency network by serendipitous modification of a pluripotency factor detailed to defend the cell against it. Consistent with this, LBP9, a member of a CP2 family of transcription factors, can form heteromer complexes with other family members, and functions either as a transcriptional activator or a repressor, depending upon the interacting partner (To et al., 2010). Furthermore, unlike other members of the family, LBP9 expression shows specific spatiotemporal regulation, which combined with its ability to bind LTRs, would have eased recruitment of HERVH to the pluripotency network.

5 Conclusions

(1) The *Sleeping Beauty*-based reprogramming system allows to simply and efficiently reprogram both mouse and human somatic cells into pluripotent stem cells. This approach, in combination with the Cre-mediated recombination system can be used to generate and correct patient-specific iPSCs, suitable for therapeutic application;

(2) By means of RNA-seq and ChIP-seq, a primate-specific retroelement family HERVH is identified as a specific marker of human pluripotent stem cells;

(3) HERVH has been recruited into the regulatory circuitry of human pluripotency, and plays functional roles in acquisition and maintenance of human pluripotency;

(4) A rare sub-population within hPSCs shares many features with naive mESCs and the human ICM, and can be isolated easily using the HERVH-based reporter system;

(5) The LBP9-HERVH-driven transcription circuitry defines naive-like hPSCs;

(6) The HERVH-driven human-specific regulatory network could at least partially explain why mouse and human ESCs are basically different;

(7) Therefore, the human ICM, rather than mESC, should be taken as the ‘golden standard’ for evaluating naive human pluripotency.

6 References

Agoni, L., Guha, C., and Lenz, J. (2013). Detection of Human Endogenous Retrovirus K (HERV-K) Transcripts in Human Prostate Cancer Cell Lines. Frontiers in oncology *3*, 180.

Anderson, D.M., Anderson, K.M., Chang, C.L., Makarewich, C.A., Nelson, B.R., McAnally, J.R., Kasaragod, P., Shelton, J.M., Liou, J., Bassel-Duby, R.*, et al.* (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell *160*, 595-606.

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

Antony, J.M., Deslauriers, A.M., Bhat, R.K., Ellestad, K.K., and Power, C. (2011). Human endogenous retroviruses and multiple sclerosis: innocent bystanders or disease determinants? Biochimica et biophysica acta *1812*, 162-176.

Aoi, T., Yae, K., Nakagawa, M., Ichisaka, T., Okita, K., Takahashi, K., Chiba, T., and Yamanaka, S. (2008). Generation of pluripotent stem cells from adult mouse liver and stomach cells. Science *321*, 699-702.

Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Current biology : CB *11*, 1017-1027.

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

Babatz, T.D., and Burns, K.H. (2013). Functional impact of the human mobilome. Current opinion in genetics & development *23*, 264-270.

Ban, H., Nishishita, N., Fusaki, N., Tabata, T., Saeki, K., Shikamura, M., Takada, N., Inoue, M., Hasegawa, M., Kawamata, S.*, et al.* (2011). Efficient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus vectors. Proceedings of the National Academy of Sciences of the United States of America *108*, 14234-14239.

Bao, S., Tang, F., Li, X., Hayashi, K., Gillich, A., Lao, K., and Surani, M.A. (2009). Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. Nature *461*, 1292-1295.

Bazzini, A.A., Johnstone, T.G., Christiano, R., Mackowiak, S.D., Obermayer, B., Fleming, E.S., Vejnar, C.E., Lee, M.T., Rajewsky, N., Walther, T.C.*, et al.* (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. The EMBO journal *33*, 981-993.

Belshaw, R., Pereira, V., Katzourakis, A., Talbot, G., Paces, J., Burt, A., and Tristem, M. (2004). Long-term reinfection of the human genome by endogenous retroviruses. Proceedings of the National Academy of Sciences of the United States of America *101*, 4894-4899.

Benvenisty, N., Leder, A., Kuo, A., and Leder, P. (1992). An embryonically expressed gene is a target for c-Myc regulation via the c-Myc-binding sequence. Genes & development *6*, 2513-2523.

Blaise, S., de Parseval, N., Benit, L., and Heidmann, T. (2003). Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. Proceedings of the National Academy of Sciences of the United States of America *100*, 13013-13018.

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nature cell biology *16*, 516-528.

Boroviak, T., and Nichols, J. (2014). The birth of embryonic pluripotency. Philosophical transactions of the Royal Society of London Series B, Biological sciences *369*.

Bowles, J., Teasdale, R.P., James, K., and Koopman, P. (2003). Dppa3 is a marker of pluripotency and has a human homologue that is expressed in germ cell tumours. Cytogenetic and genome research *101*, 261-265.

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

Brady, T., Lee, Y.N., Ronen, K., Malani, N., Berry, C.C., Bieniasz, P.D., and Bushman, F.D. (2009). Integration target site selection by a resurrected human endogenous retrovirus. Genes & development *23*, 633-642.

Brinster, R.L., and Troike, D.E. (1979). Requirements for blastocyst development in vitro. Journal of animal science *49 Suppl 2*, 26-34.

Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A.*, et al.* (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191-195.

Brower-Toland, B., Findley, S.D., Jiang, L., Liu, L., Yin, H., Dus, M., Zhou, P., Elgin, S.C., and Lin, H. (2007). Drosophila PIWI associates with chromatin and interacts directly with HP1a. Genes & development *21*, 2300-2311.

Buecker, C., Chen, H.H., Polo, J.M., Daheron, L., Bu, L., Barakat, T.S., Okwieka, P., Porter, A., Gribnau, J., Hochedlinger, K.*, et al.* (2010). A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. Cell stem cell *6*, 535-546.

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

Buganim, Y., Markoulaki, S., van Wietmarschen, N., Hoke, H., Wu, T., Ganz, K., Akhtar-Zaidi, B., He, Y., Abraham, B.J., Porubsky, D.*, et al.* (2014). The developmental potential of iPSCs is greatly influenced by reprogramming factor selection. Cell stem cell *15*, 295-309.

Bulut-Karslioglu, A., De La Rosa-Velazquez, I.A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J., Galan, C., Winter, G.E., Engist, B., Gerle, B.*, et al.* (2014). Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. Molecular cell *55*, 277-290.

Bundo, M., Toyoshima, M., Okada, Y., Akamatsu, W., Ueda, J., Nemoto-Miyauchi, T., Sunaga, F., Toritsuka, M., Ikawa, D., Kakita, A.*, et al.* (2014). Increased l1 retrotransposition in the neuronal genome in schizophrenia. Neuron *81*, 306-313.

Byrne, J.A., Pedersen, D.A., Clepper, L.L., Nelson, M., Sanger, W.G., Gokhale, S., Wolf, D.P., and Mitalipov, S.M. (2007). Producing primate embryonic stem cells by somatic cell nuclear transfer. Nature *450*, 497-502.

Cantone, I., and Fisher, A.G. (2013). Epigenetic programming and reprogramming during development. Nature structural & molecular biology *20*, 282-289.

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development *132*, 885-896.

Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., Duc, J., Jang, S.M., Turelli, P., and Trono, D. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. Genes & development *28*, 1397-1409.

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

Chan, M.M., Smith, Z.D., Egli, D., Regev, A., and Meissner, A. (2012). Mouse ooplasm confers context-specific reprogramming capacity. Nature genetics *44*, 978-980.

Chan, Y.S., Goke, J., Ng, J.H., Lu, X., Gonzales, K.A., Tan, C.P., Tng, W.Q., Hong, Z.Z., Lim, Y.S., and Ng, H.H. (2013). Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. Cell stem cell *13*, 663-675.

Chang, C.W., Lai, Y.S., Pawlik, K.M., Liu, K., Sun, C.W., Li, C., Schoeb, T.R., and Townes, T.M. (2009). Polycistronic lentiviral vector for "hit and run" reprogramming of adult skin fibroblasts to induced pluripotent stem cells. Stem cells *27*, 1042-1049.

Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., Liang, H., Chen, Y., Chen, J., Wu, Y.*, et al.* (2013a). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nature genetics *45*, 34-42.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J.*, et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106-1117.

Chen, Y., Blair, K., and Smith, A. (2013b). Robust self-renewal of rat embryonic stem cells requires fine-tuning of glycogen synthase kinase-3 inhibition. Stem cell reports *1*, 209-217.

Cherry, S.R., Biniszkiewicz, D., van Parijs, L., Baltimore, D., and Jaenisch, R. (2000). Retroviral expression in embryonic stem cells and hematopoietic stem cells. Molecular and cellular biology *20*, 7419-7426.

Chew, J.L., Loh, Y.H., Zhang, W., Chen, X., Tam, W.L., Yeap, L.S., Li, P., Ang, Y.S., Lim, B., Robson, P.*, et al.* (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Molecular and cellular biology *25*, 6031-6046.

Cho, H.J., Lee, C.S., Kwon, Y.W., Paek, J.S., Lee, S.H., Hur, J., Lee, E.J., Roh, T.Y., Chu, I.S., Leem, S.H.*, et al.* (2010). Induction of pluripotent stem cells from adult somatic cells by protein-based reprogramming without genetic manipulation. Blood *116*, 386-395.

Chuong, E.B., Rumi, M.A., Soares, M.J., and Baker, J.C. (2013). Endogenous retroviruses function as species-specific enhancer elements in the placenta. Nature genetics *45*, 325-329.

Ciaudo, C., Jay, F., Okamoto, I., Chen, C.J., Sarazin, A., Servant, N., Barillot, E., Heard, E., and Voinnet, O. (2013). RNAi-dependent and independent control of LINE1 accumulation and mobility in mouse embryonic stem cells. PLoS genetics *9*, e1003791.

Clark, K.J., Geurts, A.M., Bell, J.B., and Hackett, P.B. (2004). Transposon vectors for gene-trap insertional mutagenesis in vertebrates. Genesis *39*, 225-233.

Cohen, M., and Larsson, E. (1988). Human endogenous retroviruses. BioEssays : news and reviews in molecular, cellular and developmental biology *9*, 191-196.

Conley, A.B., Piriyapongsa, J., and Jordan, I.K. (2008). Retroviral promoters in the human genome. Bioinformatics *24*, 1563-1567.

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

Daheron, L., and D'Souza, S. (2008). Blood - SeV derived fibroblast generated iPSCs. In StemBook (Cambridge (MA)).

de Jong, J., Akhtar, W., Badhai, J., Rust, A.G., Rad, R., Hilkens, J., Berns, A., van Lohuizen, M., Wessels, L.F., and de Ridder, J. (2014). Chromatin landscapes of retroviral and transposon integration profiles. PLoS genetics *10*, e1004250.

de Koning, A.P., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive elements may comprise over two-thirds of the human genome. PLoS genetics *7*, e1002384.

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., and Xu, T. (2005). Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell *122*, 473-483.

Do, D.V., Ueda, J., Messerschmidt, D.M., Lorthongpanich, C., Zhou, Y., Feng, B., Guo, G., Lin, P.J., Hossain, M.Z., Zhang, W.*, et al.* (2013). A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo. Genes & development *27*, 1378-1390.

Dodge, J.E., Kang, Y.K., Beppu, H., Lei, H., and Li, E. (2004). Histone H3-K9 methyltransferase ESET is essential for early development. Molecular and cellular biology *24*, 2478-2486.

Dube, D., Contreras-Galindo, R., He, S., King, S.R., Gonzalez-Hernandez, M.J., Gitlin, S.D., Kaplan, M.H., and Markovitz, D.M. (2014). Genomic flexibility of human endogenous retrovirus type K. Journal of virology *88*, 9673-9682.

Dunn, S.J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naive pluripotency. Science *344*, 1156-1160.

Eggan, K., Baldwin, K., Tackett, M., Osborne, J., Gogos, J., Chess, A., Axel, R., and Jaenisch, R. (2004). Mice cloned from olfactory sensory neurons. Nature *428*, 44-49.

Elling, U., Klasen, C., Eisenberger, T., Anlag, K., and Treier, M. (2006). Murine inner cell mass-derived lineages depend on Sall4 function. Proceedings of the National Academy of Sciences of the United States of America *103*, 16319-16324.

Emani, M.R., Narva, E., Stubb, A., Chakroborty, D., Viitala, M., Rokka, A., Rahkonen, N., Moulder, R., Denessiouk, K., Trokovic, R.*, et al.* (2015). The L1TD1 Protein Interactome Reveals the Importance of Post-transcriptional Regulation in Human Pluripotency. Stem cell reports.

Emerson, R.O., and Thomas, J.H. (2009). Adaptive evolution in zinc finger transcription factors. PLoS genetics *5*, e1000325.

Esteban, M.A., Wang, T., Qin, B., Yang, J., Qin, D., Cai, J., Li, W., Weng, Z., Chen, J., Ni, S.*, et al.* (2010). Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. Cell stem cell *6*, 71-79.

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

Fang, R., Liu, K., Zhao, Y., Li, H., Zhu, D., Du, Y., Xiang, C., Li, X., Liu, H., Miao, Z.*, et al.* (2014). Generation of naive induced pluripotent stem cells from rhesus monkey fibroblasts. Cell stem cell *15*, 488-496.

Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T.*, et al.* (2009). The regulated retrotransposon transcriptome of mammalian cells. Nature genetics *41*, 563-571.

Faunes, F., Hayward, P., Descalzo, S.M., Chatterjee, S.S., Balayo, T., Trott, J., Christoforou, A., Ferrer-Vaquer, A., Hadjantonakis, A.K., Dasgupta, R.*, et al.* (2013). A membrane-associated beta-catenin/Oct4 complex correlates with ground-state pluripotency in mouse embryonic stem cells. Development *140*, 1171-1183.

Feng, B., Jiang, J., Kraus, P., Ng, J.H., Heng, J.C., Chan, Y.S., Yaw, L.P., Zhang, W., Loh, Y.H., Han, J.*, et al.* (2009). Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. Nature cell biology *11*, 197-203.

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

Folmes, C.D., Nelson, T.J., Martinez-Fernandez, A., Arrell, D.K., Lindor, J.Z., Dzeja, P.P., Ikeda, Y., Perez-Terzic, C., and Terzic, A. (2011). Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. Cell metabolism *14*, 264-271.

Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C.A., Saxena, A., Bonetti, A., Voineagu, I., Bertin, N., Kratz, A.*, et al.* (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. Nature genetics *46*, 558-566.

Frendo, J.L., Olivier, D., Cheynet, V., Blond, J.L., Bouton, O., Vidaud, M., Rabreau, M., Evain-Brion, D., and Mallet, F. (2003). Direct involvement of HERV-W Env glycoprotein in human trophoblast cell fusion and differentiation. Molecular and cellular biology *23*, 3566-3574.

Friedli, M., Turelli, P., Kapopoulou, A., Rauwel, B., Castro-Diaz, N., Rowe, H.M., Ecco, G., Unzu, C., Planet, E., Lombardo, A.*, et al.* (2014). Loss of transcriptional control over

endogenous retroelements during reprogramming to pluripotency. Genome research *24*, 1251-1259.

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., and Hasegawa, M. (2009). 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. Proceedings of the Japan Academy Series B, Physical and biological sciences *85*, 348-362.

Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A.*, et al.* (2013). Derivation of novel human ground state naive pluripotent stem cells. Nature *504*, 282-286.

Galvan, D.L., Nakazawa, Y., Kaja, A., Kettlun, C., Cooper, L.J., Rooney, C.M., and Wilson, M.H. (2009). Genome-wide mapping of PiggyBac transposon integrations in primary human T cells. Journal of immunotherapy *32*, 837-844.

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

Garcia-Perez, J.L., Morell, M., Scheys, J.O., Kulpa, D.A., Morell, S., Carter, C.C., Hammer, G.D., Collins, K.L., O'Shea, K.S., Menendez, P.*, et al.* (2010). Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. Nature *466*, 769-773.

Gasior, S.L., Roy-Engel, A.M., and Deininger, P.L. (2008). ERCC1/XPF limits L1 retrotransposition. DNA repair *7*, 983-989.

Gaudet, F., Rideout, W.M., 3rd, Meissner, A., Dausman, J., Leonhardt, H., and Jaenisch, R. (2004). Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing. Molecular and cellular biology *24*, 1640-1648.

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A.A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y.S.*, et al.* (2015). Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. Science *347*, 1002-1006.

Gianotti-Sommer, A., Rozelle, S.S., Sullivan, S., Mills, J.A., Park, S.M., Smith, B.W., Iyer, A.M., French, D.L., Kotton, D.N., Gadue, P.*, et al.* (2008). Generation of human induced pluripotent stem cells from peripheral blood using the STEMCCA lentiviral vector. In StemBook (Cambridge (MA)).

Gifford, W.D., Pfaff, S.L., and Macfarlan, T.S. (2013). Transposable elements as genetic regulatory substrates in early development. Trends in cell biology *23*, 218-226.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199-202.

Goke, J., Lu, X., Chan, Y.S., Ng, H.H., Ly, L.H., Sachs, F., and Szczerbinska, I. (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell stem cell *16*, 135-141.

Golding, M.C., Zhang, L., and Mann, M.R. (2010). Multiple epigenetic modifiers induce aggressive viral extinction in extraembryonic endoderm stem cells. Cell stem cell *6*, 457-467.

Grabundzija, I., Irgang, M., Mates, L., Belay, E., Matrai, J., Gogol-Doring, A., Kawakami, K., Chen, W., Ruiz, P., Chuah, M.K.*, et al.* (2010). Comparative analysis of transposable element vector systems in human cells. Molecular therapy : the journal of the American Society of Gene Therapy *18*, 1200-1209.

Grabundzija, I., Wang, J., Sebe, A., Erdei, Z., Kajdi, R., Devaraj, A., Steinemann, D., Szuhai, K., Stein, U., Cantz, T.*, et al.* (2013). Sleeping Beauty transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells. Nucleic acids research *41*, 1829-1847.

Grant, B. (1981). The safe-neighborhood hypothesis of junk DNA. Journal of theoretical biology *90*, 149-150.

Gridina, M.M., and Serov, O.L. (2010). Bidirectional reprogramming of mouse embryonic stem cell/fibroblast hybrid cells is initiated at the heterokaryon stage. Cell and tissue research *342*, 377-389.

Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Denervaud, N., Bucher, P., and Trono, D. (2010). KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. PLoS genetics *6*, e1000869.

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J.*, et al.* (2014). The DNA methylation landscape of human early embryos. Nature *511*, 606-610.

Gurdon, J.B. (1962a). Adult frogs derived from the nuclei of single somatic cells. Developmental biology *4*, 256-273.

Gurdon, J.B. (1962b). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. Journal of embryology and experimental morphology *10*, 622-640.

Gurdon, J.B., and Wilmut, I. (2011). Nuclear transfer to eggs and oocytes. Cold Spring Harbor perspectives in biology *3*.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L.*, et al.* (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295-300.

Haase, A., Olmer, R., Schwanke, K., Wunderlich, S., Merkert, S., Hess, C., Zweigerdt, R., Gruh, I., Meyer, J., Wagner, S.*, et al.* (2009). Generation of induced pluripotent stem cells from human cord blood. Cell stem cell *5*, 434-441.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C.*, et al.* (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell stem cell *13*, 360-369.

Hackett, J.A., Dietmann, S., Murakami, K., Down, T.A., Leitch, H.G., and Surani, M.A. (2013). Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency. Stem cell reports *1*, 518-531.

Hackett, J.A., and Surani, M.A. (2013). DNA methylation dynamics during the mammalian life cycle. Philosophical transactions of the Royal Society of London Series B, Biological sciences *368*, 20110328.

Hackett, J.A., and Surani, M.A. (2014). Regulatory principles of pluripotency: from the ground state up. Cell stem cell *15*, 416-430.

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

Hancks, D.C., and Kazazian, H.H., Jr. (2012). Active human retrotransposons: variation and disease. Current opinion in genetics & development *22*, 191-203.

Hanna, J., Cheng, A.W., Saha, K., Kim, J., Lengner, C.J., Soldner, F., Cassady, J.P., Muffat, J., Carey, B.W., and Jaenisch, R. (2010). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. Proceedings of the National Academy of Sciences of the United States of America *107*, 9222-9227.

Hanna, J., Markoulaki, S., Mitalipova, M., Cheng, A.W., Cassady, J.P., Staerk, J., Carey, B.W., Lengner, C.J., Foreman, R., Love, J.*, et al.* (2009a). Metastable pluripotent states in NOD-mouse-derived ESCs. Cell stem cell *4*, 513-524.

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

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

Hansson, J., Rafiee, M.R., Reiland, S., Polo, J.M., Gehring, J., Okawa, S., Huber, W., Hochedlinger, K., and Krijgsveld, J. (2012). Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. Cell reports *2*, 1579-1592.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., and Malim, M.H. (2003). DNA deamination mediates innate immunity to retroviral infection. Cell *113*, 803-809.

Hasegawa, K., Cowan, A.B., Nakatsuji, N., and Suemori, H. (2007). Efficient multicistronic expression of a transgene in human embryonic stem cells. Stem cells *25*, 1707-1712.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W.*, et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.

Heng, J.C., Feng, B., Han, J., Jiang, J., Kraus, P., Ng, J.H., Orlov, Y.L., Huss, M., Yang, L., Lufkin, T.*, et al.* (2010). The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. Cell stem cell *6*, 167-174.

Hochedlinger, K., and Jaenisch, R. (2002). Monoclonal mice generated by nuclear transfer from mature B and T donor cells. Nature *415*, 1035-1038.

Hockemeyer, D., Soldner, F., Cook, E.G., Gao, Q., Mitalipova, M., and Jaenisch, R. (2008). A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. Cell stem cell *3*, 346-353.

Hotta, A., Cheung, A.Y., Farra, N., Vijayaragavan, K., Seguin, C.A., Draper, J.S., Pasceri, P., Maksakova, I.A., Mager, D.L., Rossant, J.*, et al.* (2009). Isolation of human iPS cells using EOS lentiviral vectors to select for pluripotency. Nature methods *6*, 370-376.

Hotta, A., and Ellis, J. (2008). Retroviral vector silencing during iPS cell induction: an epigenetic beacon that signals distinct pluripotent states. Journal of cellular biochemistry *105*, 940-948.

Hou, P., Li, Y., Zhang, X., Liu, C., Guan, J., Li, H., Zhao, T., Ye, J., Yang, W., Liu, K.*, et al.* (2013). Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. Science *341*, 651-654.

Houghton, F.D., Thompson, J.G., Kennedy, C.J., and Leese, H.J. (1996). Oxygen consumption and energy metabolism of the early mouse embryo. Molecular reproduction and development *44*, 476-485.

Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., and Eden, A. (2008). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene *27*, 404-408.

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

Huang, P., He, Z., Ji, S., Sun, H., Xiang, D., Liu, C., Hu, Y., Wang, X., and Hui, L. (2011). Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. Nature *475*, 386-389.

Huang, P., Zhang, L., Gao, Y., He, Z., Yao, D., Wu, Z., Cen, J., Chen, X., Liu, C., Hu, Y., *et al.* (2014b). Direct reprogramming of human fibroblasts to functional and expandable hepatocytes. Cell stem cell *14*, 370-384.

Huang, X., Guo, H., Tammana, S., Jung, Y.C., Mellgren, E., Bassi, P., Cao, Q., Tu, Z.J., Kim, Y.C., Ekker, S.C., *et al.* (2010a). Gene transfer efficiency and genome-wide integration profiling of Sleeping Beauty, Tol2, and piggyBac transposons in human primary T cells. Molecular therapy : the journal of the American Society of Gene Therapy *18*, 1803-1813.

Huang, X., Haley, K., Wong, M., Guo, H., Lu, C., Wilber, A., and Zhou, X. (2010b). Unexpectedly high copy number of random integration but low frequency of persistent expression of the Sleeping Beauty transposase after trans delivery in primary human T cells. Human gene therapy *21*, 1577-1590.

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E., and Melton, D.A. (2008a). Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nature biotechnology *26*, 795-797.

Huangfu, D., Osafune, K., Maehr, R., Guo, W., Eijkelenboom, A., Chen, S., Muhlestein, W., and Melton, D.A. (2008b). Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. Nature biotechnology *26*, 1269-1275.

Hutnick, L.K., Huang, X., Loo, T.C., Ma, Z., and Fan, G. (2010). Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism. The Journal of biological chemistry *285*, 21082-21091.

Ichida, J.K., Blanchard, J., Lam, K., Son, E.Y., Chung, J.E., Egli, D., Loh, K.M., Carter, A.C., Di Giorgio, F.P., Koszka, K., *et al.* (2009). A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. Cell stem cell *5*, 491-503.

Ieda, M., Fu, J.D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell *142*, 375-386.

Ishida, T., Obata, Y., Ohara, N., Matsushita, H., Sato, S., Uenaka, A., Saika, T., Miyamura, T., Chayama, K., Nakamura, Y., *et al.* (2008). Identification of the HERV-K gag antigen in prostate cancer by SEREX using autologous patient serum and its immunogenicity. Cancer immunity *8*, 15.

Ivics, Z., Garrels, W., Mates, L., Yau, T.Y., Bashir, S., Zidek, V., Landa, V., Geurts, A., Pravenec, M., Rulicke, T.*, et al.* (2014a). Germline transgenesis in pigs by cytoplasmic microinjection of Sleeping Beauty transposons. Nature protocols *9*, 810-827.

Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvak, Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell *91*, 501-510.

Ivics, Z., Hiripi, L., Hoffmann, O.I., Mates, L., Yau, T.Y., Bashir, S., Zidek, V., Landa, V., Geurts, A., Pravenec, M.*, et al.* (2014b). Germline transgenesis in rabbits by pronuclear microinjection of Sleeping Beauty transposons. Nature protocols *9*, 794-809.

Ivics, Z., and Izsvak, Z. (2004). Transposable elements for transgenesis and insertional mutagenesis in vertebrates: a contemporary review of experimental strategies. Methods in molecular biology *260*, 255-276.

Ivics, Z., Izsvak, Z., Chapman, K.M., and Hamra, F.K. (2011). Sleeping Beauty transposon mutagenesis of the rat genome in spermatogonial stem cells. Methods *53*, 356-365.

Ivics, Z., Katzer, A., Stuwe, E.E., Fiedler, D., Knespel, S., and Izsvak, Z. (2007). Targeted Sleeping Beauty transposition in human cells. Molecular therapy : the journal of the American Society of Gene Therapy *15*, 1137-1144.

Ivics, Z., Kaufman, C.D., Zayed, H., Miskey, C., Walisko, O., and Izsvak, Z. (2004). The Sleeping Beauty transposable element: evolution, regulation and genetic applications. Current issues in molecular biology *6*, 43-55.

Ivics, Z., Mates, L., Yau, T.Y., Landa, V., Zidek, V., Bashir, S., Hoffmann, O.I., Hiripi, L., Garrels, W., Kues, W.A.*, et al.* (2014c). Germline transgenesis in rodents by pronuclear microinjection of Sleeping Beauty transposons. Nature protocols *9*, 773-793.

Iwabuchi, K.A., Yamakawa, T., Sato, Y., Ichisaka, T., Takahashi, K., Okita, K., and Yamanaka, S. (2011). ECAT11/L1td1 is enriched in ESCs and rapidly activated during iPSC generation, but it is dispensable for the maintenance and induction of pluripotency. PloS one *6*, e20461.

Izsvak, Z., and Ivics, Z. (2004). Sleeping beauty transposition: biology and applications for molecular therapy. Molecular therapy : the journal of the American Society of Gene Therapy *9*, 147-156.

Jacobs, F.M., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A.D., Katzman, S., Paten, B., Salama, S.R., and Haussler, D. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature *516*, 242-245.

James, D., Noggle, S.A., Swigut, T., and Brivanlou, A.H. (2006). Contribution of human embryonic stem cells to mouse blastocysts. Developmental biology *295*, 90-102.

Jha, A.R., Nixon, D.F., Rosenberg, M.G., Martin, J.N., Deeks, S.G., Hudson, R.R., Garrison, K.E., and Pillai, S.K. (2011). Human endogenous retrovirus K106 (HERV-K106) was infectious after the emergence of anatomically modern humans. PloS one *6*, e20234.

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

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

Kapusta, A., Kronenberg, Z., Lynch, V.J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., and Feschotte, C. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS genetics *9*, e1003470.

Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M.*, et al.* (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell stem cell *8*, 676-687.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Human molecular genetics *16*, 2272-2280.

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

Kazazian, H.H., Jr., and Goodier, J.L. (2002). LINE drive. retrotransposition and genome instability. Cell *110*, 277-280.

Kelley, D., and Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome biology *13*, R107.

Kelly, K.F., Ng, D.Y., Jayakumaran, G., Wood, G.A., Koide, H., and Doble, B.W. (2011). beta-catenin enhances Oct-4 activity and reinforces pluripotency through a TCF-independent mechanism. Cell stem cell *8*, 214-227.

Kigami, D., Minami, N., Takayama, H., and Imai, H. (2003). MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. Biology of reproduction *68*, 651-654.

Kim, J.B., Zaehres, H., Wu, G., Gentile, L., Ko, K., Sebastiano, V., Arauzo-Bravo, M.J., Ruau, D., Han, D.W., Zenke, M.*, et al.* (2008). Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. Nature *454*, 646-650.

Koche, R.P., Smith, Z.D., Adli, M., Gu, H., Ku, M., Gnirke, A., Bernstein, B.E., and Meissner, A. (2011). Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell stem cell *8*, 96-105.

Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. Nature *502*, 472-479.

Kourmouli, N., Jeppesen, P., Mahadevhaiah, S., Burgoyne, P., Wu, R., Gilbert, D.M., Bongiorni, S., Prantera, G., Fanti, L., Pimpinelli, S.*, et al.* (2004). Heterochromatin and tri-methylated lysine 20 of histone H4 in animals. Journal of cell science *117*, 2491-2501.

Koyanagi-Aoi, M., Ohnuki, M., Takahashi, K., Okita, K., Noma, H., Sawamura, Y., Teramoto, I., Narita, M., Sato, Y., Ichisaka, T.*, et al.* (2013). Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. Proceedings of the National Academy of Sciences of the United States of America *110*, 20569-20574.

Krishnan, S., Horowitz, S., and Trievel, R.C. (2011). Structure and function of histone H3 lysine 9 methyltransferases and demethylases. Chembiochem : a European journal of chemical biology *12*, 254-263.

Kuijk, E.W., van Tol, L.T., Van de Velde, H., Wubbolts, R., Welling, M., Geijsen, N., and Roelen, B.A. (2012). The roles of FGF and MAP kinase signaling in the segregation of the epiblast and hypoblast cell lineages in bovine and human embryos. Development *139*, 871-882.

Kumar, R.M., Cahan, P., Shalek, A.K., Satija, R., DaleyKeyser, A.J., Li, H., Zhang, J., Pardee, K., Gennert, D., Trombetta, J.J.*, et al.* (2014). Deconstructing transcriptional heterogeneity in pluripotent stem cells. Nature *516*, 56-61.

Kunarso, G., Chia, N.Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.S., Ng, H.H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nature genetics *42*, 631-634.

Lassar, A.B., Thayer, M.J., Overell, R.W., and Weintraub, H. (1989). Transformation by activated ras or fos prevents myogenesis by inhibiting expression of MyoD1. Cell *58*, 659-667.

Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C., and Heidmann, T. (2013). Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. Philosophical transactions of the Royal Society of London Series B, Biological sciences *368*, 20120507.

Lee, D.S., Shin, J.Y., Tonge, P.D., Puri, M.C., Lee, S., Park, H., Lee, W.C., Hussein, S.M., Bleazard, T., Yun, J.Y.*, et al.* (2014a). An epigenomic roadmap to induced pluripotency reveals DNA methylation as a reprogramming modulator. Nature communications *5*, 5619.

Lee, H.J., Hore, T.A., and Reik, W. (2014b). Reprogramming the methylome: erasing memory and creating diversity. Cell stem cell *14*, 710-719.

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

Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., and Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. Genes & development *24*, 265-276.

Leese, H.J. (2012). Metabolism of the preimplantation embryo: 40 years on. Reproduction *143*, 417-427.

Leese, H.J., and Barton, A.M. (1984). Pyruvate and glucose uptake by mouse ova and preimplantation embryos. Journal of reproduction and fertility *72*, 9-13.

Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Current biology : CB *13*, 1192-1200.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A.*, et al.* (2013). Naive pluripotency is associated with global DNA hypomethylation. Nature structural & molecular biology *20*, 311-316.

Lemaitre, C., Harper, F., Pierron, G., Heidmann, T., and Dewannieux, M. (2014). The HERV-K human endogenous retrovirus envelope protein antagonizes Tetherin antiviral activity. Journal of virology *88*, 13626-13637.

Leung, D.C., Dong, K.B., Maksakova, I.A., Goyal, P., Appanah, R., Lee, S., Tachibana, M., Shinkai, Y., Lehnertz, B., Mager, D.L.*, et al.* (2011). Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing. Proceedings of the National Academy of Sciences of the United States of America *108*, 5718-5723.

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

Li, W., Wei, W., Zhu, S., Zhu, J., Shi, Y., Lin, T., Hao, E., Hayek, A., Deng, H., and Ding, S. (2009). Generation of rat and human induced pluripotent stem cells by combining genetic reprogramming and chemical inhibitors. Cell stem cell *4*, 16-19.

Li, Y., McClintick, J., Zhong, L., Edenberg, H.J., Yoder, M.C., and Chan, R.J. (2005). Murine embryonic stem cell differentiation is promoted by SOCS-3 and inhibited by the zinc finger transcription factor Klf4. Blood *105*, 635-637.

Liang, Q., Kong, J., Stalker, J., and Bradley, A. (2009). Chromosomal mobilization and reintegration of Sleeping Beauty and PiggyBac transposons. Genesis *47*, 404-408.

Liao, B., Bao, X., Liu, L., Feng, S., Zovoilis, A., Liu, W., Xue, Y., Cai, J., Guo, X., Qin, B., *et al.* (2011). MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. The Journal of biological chemistry *286*, 17359-17364.

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

Liu, H., Zhu, F., Yong, J., Zhang, P., Hou, P., Li, H., Jiang, W., Cai, J., Liu, M., Cui, K.*, et al.* (2008). Generation of induced pluripotent stem cells from adult rhesus monkey fibroblasts. Cell stem cell *3*, 587-590.

Liu, S., Brind'Amour, J., Karimi, M.M., Shirane, K., Bogutz, A., Lefebvre, L., Sasaki, H., Shinkai, Y., and Lorincz, M.C. (2014). Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. Genes & development *28*, 2041-2055.

Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H., and Wong, J. (2013). UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nature communications *4*, 1563.

Loebel, D.A., Tsoi, B., Wong, N., O'Rourke, M.P., and Tam, P.P. (2004). Restricted expression of ETn-related sequences during post-implantation mouse development. Gene expression patterns : GEP *4*, 467-471.

Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.H., Thomas, K., Park, I.H., Garber, M., Curran, M., Onder, T., Agarwal, S.*, et al.* (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nature genetics *42*, 1113-1117.

Loh, K.M., and Lim, B. (2011). A precarious balance: pluripotency factors as lineage specifiers. Cell stem cell *8*, 363-369.

Loh, Y.H., Ng, J.H., and Ng, H.H. (2008). Molecular framework underlying pluripotency. Cell cycle *7*, 885-891.

Lohmann, F., Loureiro, J., Su, H., Fang, Q., Lei, H., Lewis, T., Yang, Y., Labow, M., Li, E., Chen, T.*, et al.* (2010). KMT1E mediated H3K9 methylation is required for the maintenance of embryonic stem cells by repressing trophectoderm differentiation. Stem cells *28*, 201-212.

Lu, X., Sachs, F., Ramsay, L., Jacques, P.E., Goke, J., Bourque, G., and Ng, H.H. (2014). The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nature structural & molecular biology *21*, 423-425.

Lueders, K.K., and Kuff, E.L. (1977). Sequences associated with intracisternal A particles are reiterated in the mouse genome. Cell *12*, 963-972.

Lyssiotis, C.A., Foreman, R.K., Staerk, J., Garcia, M., Mathur, D., Markoulaki, S., Hanna, J., Lairson, L.L., Charette, B.D., Bouchez, L.C.*, et al.* (2009). Reprogramming of murine fibroblasts to induced pluripotent stem cells with chemical complementation of Klf4. Proceedings of the National Academy of Sciences of the United States of America *106*, 8912-8917.

Macfarlan, T.S., Gifford, W.D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., Andrews, S.E., Franco, L., Rosenfeld, M.G., Ren, B.*, et al.* (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes & development *25*, 594-607.

Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57-63.

Macia, A., Blanco-Jimenez, E., and Garcia-Perez, J.L. (2014). Retrotransposons in pluripotent cells: Impact and new roles in cellular plasticity. Biochimica et biophysica acta.

Maekawa, M., Yamaguchi, K., Nakamura, T., Shibukawa, R., Kodanaka, I., Ichisaka, T., Kawamura, Y., Mochizuki, H., Goshima, N., and Yamanaka, S. (2011). Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. Nature *474*, 225-229.

Mager, D.L., and Freeman, J.D. (1995). HERV-H endogenous retroviruses: presence in the New World branch but amplification in the Old World primate lineage. Virology *213*, 395-404.

Mager, D.L., and Medstrand, P. (2003). Retroviral repeat sequences (Hampshire, England.: Macmillan Publishers Ltd.).

Magny, E.G., Pueyo, J.I., Pearl, F.M., Cespedes, M.A., Niven, J.E., Bishop, S.A., and Couso, J.P. (2013). Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. Science *341*, 1116-1120.

Maherali, N., and Hochedlinger, K. (2009). Tgfbeta signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. Current biology : CB *19*, 1718-1723.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R.*, et al.* (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell stem cell *1*, 55-70.

Mak, W., Nesterova, T.B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A.P., and Brockdorff, N. (2004). Reactivation of the paternal X chromosome in early mouse embryos. Science *303*, 666-669.

Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., and Mager, D.L. (2006). Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS genetics *2*, e2.

Maksakova, I.A., Thompson, P.J., Goyal, P., Jones, S.J., Singh, P.B., Karimi, M.M., and Lorincz, M.C. (2013). Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. Epigenetics & chromatin *6*, 15.

Malki, S., van der Heijden, G.W., O'Donnell, K.A., Martin, S.L., and Bortvin, A. (2014). A role for retrotransposon LINE-1 in fetal oocyte attrition in mice. Developmental cell *29*, 521-533.

Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G., Bonnaud, B., Lucotte, G., Duret, L., and Mandrand, B. (2004). The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. Proceedings of the National Academy of Sciences of the United States of America *101*, 1731-1736.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature *424*, 99-103.

Marchetto, M.C., Narvaiza, I., Denli, A.M., Benner, C., Lazzarini, T.A., Nathanson, J.L., Paquola, A.C., Desai, K.N., Herai, R.H., Weitzman, M.D.*, et al.* (2013). Differential L1 regulation in pluripotent stem cells of humans and apes. Nature *503*, 525-529.

Marchi, E., Kanapin, A., Magiorkinis, G., and Belshaw, R. (2014). Unfixed endogenous retroviral insertions in the human population. Journal of virology *88*, 9529-9537.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A.*, et al.* (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590-604.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J.*, et al.* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.

Martello, G., Bertone, P., and Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. The EMBO journal *32*, 2561-2574.

Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Gottgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. Cell stem cell *11*, 491-504.

Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. The EMBO journal *24*, 800-812.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proceedings of the National Academy of Sciences of the United States of America *78*, 7634-7638.

Martin, K.L., and Leese, H.J. (1995). Role of glucose in mouse preimplantation embryo development. Molecular reproduction and development *40*, 436-443.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A.*, et al.* (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nature cell biology *9*, 625-635.

Mates, L., Chuah, M.K., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J.*, et al.* (2009). Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nature genetics *41*, 753-761.

Matoba, S., Liu, Y., Lu, F., Iwabuchi, K.A., Shen, L., Inoue, A., and Zhang, Y. (2014). Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. Cell *159*, 884-895.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature *464*, 927-931.

Mattout, A., Biran, A., and Meshorer, E. (2011). Global epigenetic changes during somatic cell reprogramming to iPS cells. Journal of molecular cell biology *3*, 341-350.

McClintock, B. (1953). Induction of Instability at Selected Loci in Maize. Genetics *38*, 579-599.

McClintock, B. (1956). Controlling elements and the gene. Cold Spring Harbor symposia on quantitative biology *21*, 197-216.

McClintock, B. (1967). Regulation of pattern of gene expression by controlling elements in maize. Carnegie Inst Wash Yrbk *65*, 568-578.

Meek, S., Wei, J., Sutherland, L., Nilges, B., Buehr, M., Tomlinson, S.R., Thomson, A.J., and Burdon, T. (2013). Tuning of beta-catenin activity is required to stabilize self-renewal of rat embryonic stem cells. Stem cells *31*, 2104-2115.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.

Meissner, A., Wernig, M., and Jaenisch, R. (2007). Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. Nature biotechnology *25*, 1177-1181.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

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

Miyoshi, N., Ishii, H., Nagano, H., Haraguchi, N., Dewi, D.L., Kano, Y., Nishikawa, S., Tanemura, M., Mimori, K., Tanaka, F.*, et al.* (2011). Reprogramming of mouse and human cells to pluripotency using mature microRNAs. Cell stem cell *8*, 633-638.

Molaro, A., Hodges, E., Fang, F., Song, Q., McCombie, W.R., Hannon, G.J., and Smith, A.D. (2011). Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. Cell *146*, 1029-1041.

Montserrat, N., Nivet, E., Sancho-Martinez, I., Hishida, T., Kumar, S., Miquel, L., Cortina, C., Hishida, Y., Xia, Y., Esteban, C.R.*, et al.* (2013). Reprogramming of human fibroblasts to pluripotency with lineage specifiers. Cell stem cell *13*, 341-350.

Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., and Kazazian, H.H., Jr. (1996). High frequency retrotransposition in cultured mammalian cells. Cell *87*, 917-927.

Morgani, S.M., Canham, M.A., Nichols, J., Sharov, A.A., Migueles, R.P., Ko, M.S., and Brickman, J.M. (2013). Totipotent embryonic stem cells arise in ground-state culture conditions. Cell reports *3*, 1945-1957.

Moyes, D., Griffiths, D.J., and Venables, P.J. (2007). Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease. Trends in genetics : TIG *23*, 326-333.

Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., Flory, E., Schumann, G.G., and Munk, C. (2006). APOBEC3 proteins inhibit human LINE-1 retrotransposition. The Journal of biological chemistry *281*, 22161-22172.

Muotri, A.R., Marchetto, M.C., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K., and Gage, F.H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. Nature *468*, 443-446.

Nagy, A., Gocza, E., Diaz, E.M., Prideaux, V.R., Ivanyi, E., Markkula, M., and Rossant, J. (1990). Embryonic stem cells alone are able to support fetal development in the mouse. Development *110*, 815-821.

Najafabadi, H.S., Mnaimneh, S., Schmitges, F.W., Garton, M., Lam, K.N., Yang, A., Albu, M., Weirauch, M.T., Radovani, E., Kim, P.M.*, et al.* (2015). C2H2 zinc finger proteins greatly expand the human regulatory lexicon. Nature biotechnology.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nature biotechnology *26*, 101-106.

Nakatake, Y., Fukui, N., Iwamatsu, Y., Masui, S., Takahashi, K., Yagi, R., Yagi, K., Miyazaki, J., Matoba, R., Ko, M.S.*, et al.* (2006). Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. Molecular and cellular biology *26*, 7772-7782.

Narva, E., Rahkonen, N., Emani, M.R., Lund, R., Pursiheimo, J.P., Nasti, J., Autio, R., Rasool, O., Denessiouk, K., Lahdesmaki, H.*, et al.* (2012). RNA-binding protein L1TD1 interacts with LIN28 via RNA and is required for human embryonic stem cell self-renewal and cancer cell proliferation. Stem cells *30*, 452-460.

Ng, S.Y., Johnson, R., and Stanton, L.W. (2012). Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. The EMBO journal *31*, 522-533.

Nichols, J., Silva, J., Roode, M., and Smith, A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development *136*, 3215-3222.

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell stem cell *4*, 487-492.

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

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

O'Malley, J., Skylaki, S., Iwabuchi, K.A., Chantzoura, E., Ruetz, T., Johnsson, A., Tomlinson, S.R., Linnarsson, S., and Kaji, K. (2013). High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature *499*, 88-91.

Ohnuki, M., Tanabe, K., Sutou, K., Teramoto, I., Sawamura, Y., Narita, M., Nakamura, M., Tokunaga, Y., Nakamura, M., Watanabe, A.*, et al.* (2014). Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. Proceedings of the National Academy of Sciences of the United States of America *111*, 12426-12431.

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. Science *303*, 644-649.

Okamoto, I., Patrat, C., Thepot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., Wolf, J.P., Renard, J.P., Duranthon, V.*, et al.* (2011). Eutherian mammals use diverse

strategies to initiate X-chromosome inactivation during development. Nature *472*, 370-374.

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

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313-317.

Okitsu, C.Y., and Hsieh, C.L. (2007). DNA methylation dictates histone H3K4 methylation. Molecular and cellular biology *27*, 2746-2757.

Onder, T.T., Kara, N., Cherry, A., Sinha, A.U., Zhu, N., Bernt, K.M., Cahan, P., Marcarci, B.O., Unternaehrer, J., Gupta, P.B.*, et al.* (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature *483*, 598-602.

Papapetrou, E.P., Lee, G., Malani, N., Setty, M., Riviere, I., Tirunagari, L.M., Kadota, K., Roth, S.L., Giardina, P., Viale, A.*, et al.* (2011). Genomic safe harbors permit high beta-globin transgene expression in thalassemia induced pluripotent stem cells. Nature biotechnology *29*, 73-78.

Papp, B., and Plath, K. (2011). Reprogramming to pluripotency: stepwise resetting of the epigenetic landscape. Cell research *21*, 486-501.

Parada, C.A., Yoon, J.B., and Roeder, R.G. (1995). A novel LBP-1-mediated restriction of HIV-1 transcription at the level of elongation in vitro. The Journal of biological chemistry *270*, 2274-2283.

Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D., and Knowles, B.B. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Developmental cell *7*, 597-606.

Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y.*, et al.* (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Molecular cell *12*, 1577-1589.

Pezic, D., Manakov, S.A., Sachidanandam, R., and Aravin, A.A. (2014). piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. Genes & development *28*, 1410-1428.

Piriyapongsa, J., Polavarapu, N., Borodovsky, M., and McDonald, J. (2007). Exonization of the LTR transposable elements in human genome. BMC genomics *8*, 291.

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

Qin, H., Diaz, A., Blouin, L., Lebbink, R.J., Patena, W., Tanbun, P., LeProust, E.M., McManus, M.T., Song, J.S., and Ramalho-Santos, M. (2014). Systematic identification of barriers to human iPSC generation. Cell *158*, 449-461.

Quenneville, S., Turelli, P., Bojkowska, K., Raclot, C., Offner, S., Kapopoulou, A., and Trono, D. (2012). The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. Cell reports *2*, 766-773.

Ramirez, M.A., Pericuesta, E., Fernandez-Gonzalez, R., Moreira, P., Pintado, B., and Gutierrez-Adan, A. (2006). Transcriptional and post-transcriptional regulation of retrotransposons IAP and MuERV-L affect pluripotency of mice ES cells. Reproductive biology and endocrinology : RB&E *4*, 55.

Rebollo, R., Karimi, M.M., Bilenky, M., Gagnier, L., Miceli-Royer, K., Zhang, Y., Goyal, P., Keane, T.M., Jones, S., Hirst, M.*, et al.* (2011). Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. PLoS genetics *7*, e1002301.

Redmer, T., Diecke, S., Grigoryan, T., Quiroga-Negreira, A., Birchmeier, W., and Besser, D. (2011). E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. EMBO reports *12*, 720-726.

Reis, B.S., Jungbluth, A.A., Frosina, D., Holz, M., Ritter, E., Nakayama, E., Ishida, T., Obata, Y., Carver, B., Scher, H.*, et al.* (2013). Prostate cancer progression correlates with increased humoral immune response to a human endogenous retrovirus GAG protein. Clinical cancer research : an official journal of the American Association for Cancer Research *19*, 6112-6125.

Richardson, S.R., Narvaiza, I., Planegger, R.A., Weitzman, M.D., and Moran, J.V. (2014). APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. eLife *3*, e02008.

Romerio, F., Gabriel, M.N., and Margolis, D.M. (1997). Repression of human immunodeficiency virus type 1 through the novel cooperation of human factors YY1 and LSF. Journal of virology *71*, 9375-9382.

Roode, M., Blair, K., Snell, P., Elder, K., Marchant, S., Smith, A., and Nichols, J. (2012). Human hypoblast formation is not dependent on FGF signalling. Developmental biology *361*, 358-363.

Rowe, H.M., Friedli, M., Offner, S., Verp, S., Mesnard, D., Marquis, J., Aktas, T., and Trono, D. (2013a). De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. Development *140*, 519-529.

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., *et al.* (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature *463*, 237-240.

Rowe, H.M., Kapopoulou, A., Corsinotti, A., Fasching, L., Macfarlan, T.S., Tarabay, Y., Viville, S., Jakobsson, J., Pfaff, S.L., and Trono, D. (2013b). TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. Genome research *23*, 452-461.

Rowe, H.M., and Trono, D. (2011). Dynamic control of endogenous retroviruses during development. Virology *411*, 273-287.

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

Santoni, F.A., Guerra, J., and Luban, J. (2012). HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. Retrovirology *9*, 111.

Sarkar, A., Sim, C., Hong, Y.S., Hogan, J.R., Fraser, M.J., Robertson, H.M., and Collins, F.H. (2003). Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences. Molecular genetics and genomics : MGG *270*, 173-180.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A.H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nature medicine *10*, 55-63.

Schlesinger, S., Lee, A.H., Wang, G.Z., Green, L., and Goff, S.P. (2013). Proviral silencing in embryonic cells is regulated by Yin Yang 1. Cell reports *4*, 50-58.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes & development *18*, 1251-1262.

Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521-529.

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes & development *16*, 919-932.

Seisenberger, S., Peat, J.R., Hore, T.A., Santos, F., Dean, W., and Reik, W. (2013). Reprogramming DNA methylation in the mammalian life cycle: building and breaking

epigenetic barriers. Philosophical transactions of the Royal Society of London Series B, Biological sciences *368*, 20110330.

Seki, T., Yuasa, S., Oda, M., Egashira, T., Yae, K., Kusumoto, D., Nakata, H., Tohyama, S., Hashimoto, H., Kodaira, M.*, et al.* (2010). Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells. Cell stem cell *7*, 11-14.

Shipony, Z., Mukamel, Z., Cohen, N.M., Landan, G., Chomsky, E., Zeliger, S.R., Fried, Y.C., Ainbinder, E., Friedman, N., and Tanay, A. (2014). Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. Nature *513*, 115-119.

Shu, J., Wu, C., Wu, Y., Li, Z., Shao, S., Zhao, W., Tang, X., Yang, H., Shen, L., Zuo, X.*, et al.* (2013). Induction of pluripotency in mouse somatic cells with lineage specifiers. Cell *153*, 963-975.

Shukla, R., Upton, K.R., Munoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., Ghisletti, S.*, et al.* (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell *153*, 101-111.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722-737.

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

Simonsson, S., and Gurdon, J. (2004). DNA demethylation is necessary for the epigenetic reprogramming of somatic cell nuclei. Nature cell biology *6*, 984-990.

Smith, Z.D., Chan, M.M., Humm, K.C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., and Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. Nature *511*, 611-615.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339-344.

Sommer, C.A., Sommer, A.G., Longmire, T.A., Christodoulou, C., Thomas, D.D., Gostissa, M., Alt, F.W., Murphy, G.J., Kotton, D.N., and Mostoslavsky, G. (2010). Excision of reprogramming transgenes improves the differentiation potential of iPS cells generated with a single excisable vector. Stem cells *28*, 64-74.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D.*, et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490-495.

Stadtfeld, M., Apostolou, E., Ferrari, F., Choi, J., Walsh, R.M., Chen, T., Ooi, S.S., Kim, S.Y., Bestor, T.H., Shioda, T.*, et al.* (2012). Ascorbic acid prevents loss of Dlk1-Dio3

imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. Nature genetics *44*, 398-405, S391-392.

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

Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G., and Hochedlinger, K. (2008). Induced pluripotent stem cells generated without viral integration. Science *322*, 945-949.

Stein, S., Ott, M.G., Schultze-Strasser, S., Jauch, A., Burwinkel, B., Kinner, A., Schmidt, M., Kramer, A., Schwable, J., Glimm, H.*, et al.* (2010). Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. Nature medicine *16*, 198-204.

Subramanyam, D., Lamouille, S., Judson, R.L., Liu, J.Y., Bucay, N., Derynck, R., and Blelloch, R. (2011). Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. Nature biotechnology *29*, 443-448.

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M.P., and Wang, T. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome research *24*, 1963-1976.

Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G.J., and Schultz, R.M. (2004). RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. Developmental biology *269*, 276-285.

Szabo, E., Rampalli, S., Risueno, R.M., Schnerch, A., Mitchell, R., Fiebig-Comyn, A., Levadoux-Martin, M., and Bhatia, M. (2010). Direct conversion of human fibroblasts to multilineage blood progenitors. Nature *468*, 521-526.

Tachibana, M., Amato, P., Sparman, M., Gutierrez, N.M., Tippner-Hedges, R., Ma, H., Kang, E., Fulati, A., Lee, H.S., Sritanaudomchai, H.*, et al.* (2013). Human embryonic stem cells derived by somatic cell nuclear transfer. Cell *153*, 1228-1238.

Tachibana, M., Sparman, M., Ramsey, C., Ma, H., Lee, H.S., Penedo, M.C., and Mitalipov, S. (2012). Generation of chimeric rhesus monkeys. Cell *148*, 285-295.

Tada, M., Tada, T., Lefebvre, L., Barton, S.C., and Surani, M.A. (1997). Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. The EMBO journal *16*, 6510-6520.

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., and Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Current biology : CB *11*, 1553-1558.

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

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

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W.*, et al.* (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell *158*, 1254-1269.

Tan, S.S., Williams, E.A., and Tam, P.P. (1993). X-chromosome inactivation occurs at different times in different tissues of the post-implantation mouse embryo. Nature genetics *3*, 170-174.

Tee, W.W., Shen, S.S., Oksuz, O., Narendra, V., and Reinberg, D. (2014). Erk1/2 activity promotes chromatin features and RNAPII phosphorylation at developmental promoters in mouse ESCs. Cell *156*, 678-690.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196-199.

Theunissen, T.W., and Jaenisch, R. (2014). Molecular control of induced pluripotency. Cell stem cell *14*, 720-734.

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L.*, et al.* (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell stem cell *15*, 471-487.

Thomas, J.H., and Schneider, S. (2011). Coevolution of retroelements and tandem zinc finger genes. Genome research *21*, 1800-1812.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145-1147.

Thomson, M., Liu, S.J., Zou, L.N., Smith, Z., Meissner, A., and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell *145*, 875-889.

Thuan, N.V., Kishigami, S., and Wakayama, T. (2010). How to improve the success rate of mouse cloning technology. The Journal of reproduction and development *56*, 20-30.

To, S., Rodda, S.J., Rathjen, P.D., and Keough, R.A. (2010). Modulation of CP2 family transcriptional activity by CRTR-1 and sumoylation. PloS one *5*, e11702.

Tonge, P.D., Corso, A.J., Monetti, C., Hussein, S.M., Puri, M.C., Michael, I.P., Li, M., Lee, D.S., Mar, J.C., Cloonan, N.*, et al.* (2014). Divergent reprogramming routes lead to alternative stem-cell states. Nature *516*, 192-197.

Turelli, P., Castro-Diaz, N., Marzetta, F., Kapopoulou, A., Raclot, C., Duc, J., Tieng, V., Quenneville, S., and Trono, D. (2014). Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome research *24*, 1260-1270.

Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L.E., Trotter, M.W., Cho, C.H., Martinez, A., Rugg-Gunn, P.*, et al.* (2009). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development *136*, 1339-1349.

van den Berg, D.L., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010). An Oct4-centered protein interaction network in embryonic stem cells. Cell stem cell *6*, 369-381.

Vassena, R., Boue, S., Gonzalez-Roca, E., Aran, B., Auer, H., Veiga, A., and Izpisua Belmonte, J.C. (2011). Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. Development *138*, 3699-3709.

Vastenhouw, N.L., and Schier, A.F. (2012). Bivalent histone modifications in early embryogenesis. Current opinion in cell biology *24*, 374-386.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Sudhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. Nature *463*, 1035-1041.

Viswanathan, S.R., Daley, G.Q., and Gregory, R.I. (2008). Selective blockade of microRNA processing by Lin28. Science *320*, 97-100.

Wakayama, S., Jakt, M.L., Suzuki, M., Araki, R., Hikichi, T., Kishigami, S., Ohta, H., Van Thuan, N., Mizutani, E., Sakaide, Y.*, et al.* (2006). Equivalency of nuclear transfer-derived embryonic stem cells to those derived from fertilized mouse blastocysts. Stem cells *24*, 2023-2033.

Wakayama, T., Perry, A.C., Zuccotti, M., Johnson, K.R., and Yanagimachi, R. (1998). Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. Nature *394*, 369-374.

Wallace, T.A., Downey, R.F., Seufert, C.J., Schetter, A., Dorsey, T.H., Johnson, C.A., Goldman, R., Loffredo, C.A., Yan, P., Sullivan, F.J.*, et al.* (2014). Elevated HERV-K mRNA expression in PBMC is associated with a prostate cancer diagnosis particularly in older men and smokers. Carcinogenesis *35*, 2074-2083.

Wang, J., Xie, G., Singh, M., Ghanbarian, A.T., Rasko, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N.V.*, et al.* (2014a). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature *516*, 405-409.

Wang, T., Chen, K., Zeng, X., Yang, J., Wu, Y., Shi, X., Qin, B., Zeng, L., Esteban, M.A., Pan, G.*, et al.* (2011). The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. Cell stem cell *9*, 575-587.

Wang, Y., Wang, J., Devaraj, A., Singh, M., Jimenez Orgaz, A., Chen, J.X., Selbach, M., Ivics, Z., and Izsvak, Z. (2014b). Suicidal autointegration of sleeping beauty and piggyBac transposons in eukaryotic cells. PLoS genetics *10*, e1004103.

Ware, C.B., Nelson, A.M., Mecham, B., Hesson, J., Zhou, W., Jonlin, E.C., Jimenez-Caliani, A.J., Deng, X., Cavanaugh, C., Cook, S.*, et al.* (2014). Derivation of naive human embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America *111*, 4484-4489.

Welstead, G.G., Brambrink, T., and Jaenisch, R. (2008). Generating iPS cells from MEFS through forced expression of Sox-2, Oct-4, c-Myc, and Klf4. Journal of visualized experiments : JoVE.

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

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318-324.

Wildschutte, J.H., Ram, D., Subramanian, R., Stevens, V.L., and Coffin, J.M. (2014). The distribution of insertionally polymorphic endogenous retroviruses in breast cancer patients and cancer-free controls. Retrovirology *11*, 62.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810-813.

Wissing, S., Montano, M., Garcia-Perez, J.L., Moran, J.V., and Greene, W.C. (2011). Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. The Journal of biological chemistry *286*, 36427-36437.

Wolf, D., and Goff, S.P. (2007). TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. Cell *131*, 46-57.

Wolf, D., and Goff, S.P. (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature *458*, 1201-1204.

Wolf, G., Yang, P., Fuchtbauer, A.C., Fuchtbauer, E.M., Silva, A.M., Park, C., Wu, W., Nielsen, A.L., Pedersen, F.S., and Macfarlan, T.S. (2015). The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. Genes & development *29*, 538-554.

Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hamalainen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M.*, et al.* (2009). piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature *458*, 766-770.

Wong, R.C., Ibrahim, A., Fong, H., Thompson, N., Lock, L.F., and Donovan, P.J. (2011). L1TD1 is a marker for undifferentiated human embryonic stem cells. PloS one *6*, e19355.

Worringer, K.A., Rand, T.A., Hayashi, Y., Sami, S., Takahashi, K., Tanabe, K., Narita, M., Srivastava, D., and Yamanaka, S. (2014). The let-7/LIN-41 pathway regulates reprogramming to human induced pluripotent stem cells by controlling expression of prodifferentiation genes. Cell stem cell *14*, 40-52.

Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). 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.

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

Wu, Z., Chen, J., Ren, J., Bao, L., Liao, J., Cui, C., Rao, L., Li, H., Gu, Y., Dai, H.*, et al.* (2009). Generation of pig induced pluripotent stem cells with a drug-inducible system. Journal of molecular cell biology *1*, 46-54.

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

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

Xue, X., Huang, X., Nodland, S.E., Mates, L., Ma, L., Izsvak, Z., Ivics, Z., LeBien, T.W., McIvor, R.S., Wagner, J.E.*, et al.* (2009). Stable gene transfer and expression in cord blood-derived CD34+ hematopoietic stem and progenitor cells by a hyperactive Sleeping Beauty transposon system. Blood *114*, 1319-1330.

Yakubov, E., Rechavi, G., Rozenblatt, S., and Givol, D. (2010). Reprogramming of human fibroblasts to pluripotent stem cells using mRNA of four transcription factors. Biochemical and biophysical research communications *394*, 189-193.

Yamaji, M., Ueda, J., Hayashi, K., Ohta, H., Yabuta, Y., Kurimoto, K., Nakato, R., Yamada, Y., Shirahige, K., and Saitou, M. (2013). PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. Cell stem cell *12*, 368-382.

Yamanaka, Y., Lanner, F., and Rossant, J. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. Development *137*, 715-724.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J.*, et al.* (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nature structural & molecular biology *20*, 1131-1139.

Yang, B., Chen, K., Zhang, C., Huang, S., and Zhang, H. (2007). Virion-associated uracil DNA glycosylase-2 and apurinic/apyrimidinic endonuclease are involved in the

degradation of APOBEC3G-edited nascent HIV-1 DNA. The Journal of biological chemistry *282*, 11667-11675.

Yant, S.R., Meuse, L., Chiu, W., Ivics, Z., Izsvak, Z., and Kay, M.A. (2000). Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. Nature genetics *25*, 35-41.

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

Yeo, J.C., Jiang, J., Tan, Z.Y., Yim, G.R., Ng, J.H., Goke, J., Kraus, P., Liang, H., Gonzales, K.A., Chong, H.C.*, et al.* (2014). Klf2 is an essential factor that sustains ground state pluripotency. Cell stem cell *14*, 864-872.

Yi, F., Pereira, L., Hoffman, J.A., Shy, B.R., Yuen, C.M., Liu, D.R., and Merrill, B.J. (2011). Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. Nature cell biology *13*, 762-770.

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

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-523.

Yoon, J.B., Li, G., and Roeder, R.G. (1994). Characterization of a family of related cellular transcription factors which can modulate human immunodeficiency virus type 1 transcription in vitro. Molecular and cellular biology *14*, 1776-1785.

Yoshida, K., Chambers, I., Nichols, J., Smith, A., Saito, M., Yasukawa, K., Shoyab, M., Taga, T., and Kishimoto, T. (1994). Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways. Mechanisms of development *45*, 163-171.

Yoshimizu, T., Sugiyama, N., De Felice, M., Yeom, Y.I., Ohbo, K., Masuko, K., Obinata, M., Abe, K., Scholer, H.R., and Matsui, Y. (1999). Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. Development, growth & differentiation *41*, 675-684.

Yu, B., He, Z.Y., You, P., Han, Q.W., Xiang, D., Chen, F., Wang, M.J., Liu, C.C., Lin, X.W., Borjigin, U.*, et al.* (2013). Reprogramming fibroblasts into bipotential hepatic stem cells by defined factors. Cell stem cell *13*, 328-340.

Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., La Russa, M., Xie, M., Ding, S.*, et al.* (2015). Small Molecules Enhance CRISPR Genome Editing in Pluripotent Stem Cells. Cell stem cell *16*, 142-147.

Yu, J., Chau, K.F., Vodyanik, M.A., Jiang, J., and Jiang, Y. (2011). Efficient feeder-free episomal reprogramming with small molecules. PloS one *6*, e17557.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, II, and Thomson, J.A. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. Science *324*, 797-801.

Yu, J., Vodyanik, M.A., He, P., Slukvin, II, and Thomson, J.A. (2006). Human embryonic stem cells reprogram myeloid precursors following cell-cell fusion. Stem cells *24*, 168-176.

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

Yusa, K., Rad, R., Takeda, J., and Bradley, A. (2009). Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. Nature methods *6*, 363-369.

Zayed, H., Izsvak, Z., Walisko, O., and Ivics, Z. (2004). Development of hyperactive sleeping beauty transposon vectors by mutational analysis. Molecular therapy : the journal of the American Society of Gene Therapy *9*, 292-304.

Zeng, M., Hu, Z., Shi, X., Li, X., Zhan, X., Li, X.D., Wang, J., Choi, J.H., Wang, K.W., Purrington, T.*, et al.* (2014). MAVS, cGAS, and endogenous retroviruses in T-independent B cell responses. Science *346*, 1486-1492.

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

Zhang, Y., Maksakova, I.A., Gagnier, L., van de Lagemaat, L.N., and Mager, D.L. (2008). Genome-wide assessments reveal extremely high levels of polymorphism of two active families of mouse endogenous retroviral elements. PLoS genetics *4*, e1000007.

Zhao, Y., Yin, X., Qin, H., Zhu, F., Liu, H., Yang, W., Zhang, Q., Xiang, C., Hou, P., Song, Z.*, et al.* (2008). Two supporting factors greatly improve the efficiency of human iPSC generation. Cell stem cell *3*, 475-479.

Zhou, H., Wu, S., Joo, J.Y., Zhu, S., Han, D.W., Lin, T., Trauger, S., Bien, G., Yao, S., Zhu, Y.*, et al.* (2009). Generation of induced pluripotent stem cells using recombinant proteins. Cell stem cell *4*, 381-384.

Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J., and Melton, D.A. (2008). In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. Nature *455*, 627-632.

Zhou, W., Choi, M., Margineantu, D., Margaretha, L., Hesson, J., Cavanaugh, C., Blau, C.A., Horwitz, M.S., Hockenbery, D., Ware, C.*, et al.* (2012). HIF1alpha induced switch

from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. The EMBO journal *31*, 2103-2116.

Zhou, W., Clouston, D.R., Wang, X., Cerruti, L., Cunningham, J.M., and Jane, S.M. (2000). Induction of human fetal globin gene expression by a novel erythroid factor, NF-E4. Molecular and cellular biology *20*, 7662-7672.

Zhu, J., Park, C.W., Sjeklocha, L., Kren, B.T., and Steer, C.J. (2010a). High-level genomic integration, epigenetic changes, and expression of sleeping beauty transgene. Biochemistry *49*, 1507-1521.

Zhu, S., Li, W., Zhou, H., Wei, W., Ambasudhan, R., Lin, T., Kim, J., Zhang, K., and Ding, S. (2010b). Reprogramming of human primary somatic cells by OCT4 and chemical compounds. Cell stem cell *7*, 651-655.

Zhu, S., Rezvani, M., Harbell, J., Mattis, A.N., Wolfe, A.R., Benet, L.Z., Willenbring, H., and Ding, S. (2014). Mouse liver repopulation with hepatocytes generated from human fibroblasts. Nature *508*, 93-97.

Appendix I Abbreviations

APOBEC	apolipoprotein B mRNA-editing enzyme and catalytic enzymes
array-CGH	microarray-based comparative genomic hybridisation
Fgf	fibroblast growth factor
bFGF	basic fibroblast growth fator
BMP4	bone morphogenetic protein 4
bp	base pair
ChIP-seq	chromatin immunoprecipitation with massively parallel DNA sequencing
EMSA	electrophoretic mobility shift assay
EMT	epithelial-to-mesenchymal-transition
EpiSCs	epiblast stem cells
ERVK	endogenous retrovirus, class K
ERVL	endogenous retrovirus, class L
ERVs	endogenous retroviruses
ESCs	embryonic stem cells
FACS	fluorescence-activated cell sorting
FISH	fluorescence in situ hybridization
FOA	fetal oocyte attrition
GFP	green fluorescent protein
H3K27ac	acetylated histone h3 lysine 27/acetylation of histone h3 lysine 27
H3K27me3	trimethylated histone h3 lysine 27/trimethylation of histone h3 lysine 27
H3K4me1	monomethylated histone h3 lysine 4/monomethylation of histone h3 lysine 4
H3K4me2	dimethylated histone h3 lysine 4/dimethylation of histone h3 lysine 4
H3K4me3	trimethylated histone h3 lysine 4/trimethylation of histone h3 lysine 4
H3K79me2	dimethylated histone h3 lysine 79/dimethylation of histone h3 lysine 79
H3K79me3	trimethylated histone h3 lysine 79/trimethylation of histone h3 lysine 79
H3K9me1	monomethylated histone h3 lysine 9/monomethylation of histone h3 lysine 9
H3K9me2	dimethylated histone h3 lysine 9/dimethylation of histone h3 lysine 9
H3K9me3	trimethylated histone h3 lysine 9//trimethylation of histone h3 lysine 9
H4K20me3	trimethylated histone h4 lysine 20/trimethylation of histone h4 lysine 20
HDAC	histone deacetylase
HERVH	human endogenous retrovirus, family H
HERVK	human endogenous retrovirus, family K
HERVW	human endogenous retrovirus, family W
hESCs	human embryonic stem cells

hiPSCs	human induced pluripotent stem cells
hPSCs	human pluripotent stem cells
IAP	intracisternal A-type particle
ICM	inner cell mass
iPSCs	induced pluripotent stem cells
ITR	inverted terminal repeat
KRAB-ZFP	Kruppel-associated box domain-containg zinc finger protein
L1	LINE-1
LIF	leukemia inhibitory factor
LINC-ROR	Long Intergenic Non-protein Coding RNA, Regulator Of Reprogramming
lincRNA	long intergenic non-coding RNA
LINE	long interspersed element
lncRNA	long non-coding RNA
LTR	long terminal repeat
MAPK	mitogen-activated protein kinase
MEF	mouse embryonic fibroblast
mESCs	mouse embryonic stem cells
MET	mesenchymal-to-epithelial transition
miPSC	mouse induced pluripotent stem cells
miRNA	microRNA
mRNA	messenger RNA
nt	nucleotides
ORF	open reading frame
OSK	Oct4, Sox2 and Klf4
OSKM	Oct4, Sox2, Klf4 and c-Myc
OXPHOS	oxidative phosphorylation
PBS	primer-binding site
PGCs	primordial germ cells
piRNA	Piwi-interacting RNA
PRC2	Polycomb Repressive Complex 2
PSCs	pluripotent stem cells
RMCE	recombinase-mediated cassette exchange
RNA-seq	RNA sequencing
RNAi	RNA interference
SCNF	somatic cell nuclear transfer
shRNA	short hairpin RNA

SINE	short interspersed element
siRNA	small interfering RNA
SSEA1	stage-specific embryonic antigen 1
TEs	transposable elements
TFs	transcription factors
TSS	transcription start site
VPA	vlproic acid

Appendix II CURRICULUM VITAE

Name: Wang, Jichang

Nationality: Chinese

EDUCATION

12/2009 - present Max-Delbrück-Center for Molecular Medicine (MDC)
& Freie Universität Berlin, Germany
PhD student, majoring in Biology

09/2006 – 07/2009 School of Basic Medical Sciences, Nanjing Medical University, China
Master of Degree majoring in Pharmacology, July 2009.

09/2001 – 07/2006 Fourth School of Clinical Medicine, Nanjing Medical University, China
Bachelor's Degree majoring in Medicine (Pediatrics), July 2006.

PROFESSIONAL EXPERIENCE

12/2009 – present Scientist. Max-Delbrück-Center for Molecular Medicine (MDC), Berlin, Germany

07/2009 – 11/2009 CRA. Jiangsu Hengrui Medicine Co., Ltd, China

01/2006 – 06/2006 Internship in the Second affiliated Hospital, Nanjing Medical University, Nanjing, China

06/2005 – 12/2005 Internship in Nanjing Children's Hospital, Nanjing Medical University, Nanjing, China

PUBLICATIONS

(* co-first author; # co-corresponding author)

1. **Jichang Wang***, Gangcai Xie*, Manvedra Singh, Avazeh T. Ghanbarian, Tamás Raskó, Attila Szvetnik, Huiqiang Cai, Daniel Besser, Alessandro Prigione, Nina Fuchs, Gerald Schumann, Wei Chen, Matthew C. Lorincz, Zoltán Ivics, Laurence D. Hurst#, Zsuzsanna Izsvák#. Primate-specific endogenous retrovirus driven transcription defines naïve-like stem cells. ***Nature***. 2014; 516: 405-409. doi: 10.1038/nature13804.

2. Yongming Wang*, **Jichang Wang***, Anatharam Devaraj, Manvendra Singh, Ana Jimenez Orgaz, Jia-Xuan Chen, Matthias Selbach, Zoltán Ivics#, Zsuzsanna Izsvák#. Suicidal autointegration of sleeping beauty and piggyback transposons in eukaryotic cells. ***PLOS Genet***. 2014;10(3):e1004103. doi: 10.1371/journal.pgen.1004103.

3. Wenli Zhang; Manish Solanki; Melanie Ebel; Chistina Rauschhuber; Nadine Müther; **Jichang Wang**; Chuanbo Sun; Zsuzsanna Izsvak; Anja Ehrhardt. Hybrid adeno-associated viral vectors utilizing transposase-mediated somatic integration for stable transgene expression in human cells. ***PLOS ONE.*** 2013; 8(10):e76771. doi: 10.1371/journal.pone.0076771.

4. Wenli Zhang, Martin Muck-Hausl, **Jichang Wang**, Chuanbo Sun, Maren Gebbing, Csaba Miskey, Zoltan Ivics, Zsuzsanna Izsvak and Anja Ehrhardt**.** Integration profile and safety of an adenovirus hybrid-vector utilizing hyperactive Sleeping Beauty transposase for somatic integration. ***PLOS ONE.*** 2013; 8(10):e75344. doi: 10.1371/journal.pone.0075344.

5. Ivana Grabundzija*, **Jichang Wang***, Attila Sebe*, Zsuzsanna Erdei, Robert Kajdi, Anantharam Devaraj, Doris Steinemann, Károly Szuhai, Ulrike Stein, Tobias Cantz, Axel Schambach, Christopher Baum, Zsuzsanna Izsvák#, Balázs Sarkadi#, and Zoltán Ivics#. Sleeping Beauty transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells. ***Nucleic Acids Res***. 2013; 41(3): 1829-1847. doi: 10.1093/nar/gks1305.

6. Ruizhen Shi, **Jichang Wang**, Songhua Huang, Xiaojun Wang, Qingping Li. Angiotensin II induces vascular endothelial growth factor synthesis in mesenchymal stem cells. ***Exp Cell Res***. 2009; 315: 10-15. doi: 10.1016/j.yexcr.2008.09.024.

PRESENTATIONS

1. **Jichang Wang**, Gangcai Xie, Manvedra Singh, Avazeh T. Ghanbarian, Tamás Raskó, Attila Szvetnik, Huiqiang Cai, Daniel Besser, Alessandro Prigione, Nina Fuchs, Gerald Schumann, Wei Chen, Matthew C. Lorincz, Zoltán Ivics, Laurence D. Hurst, Zsuzsanna Izsvák. Primate-specific endogenous retrovirus driven transcription defines naïve-like stem cells (Selected oral presentation). **2nd International Annual Conference of the German Stem Cell Network (GSCN). November 3rd-5th, 2014, Heidelberg, Germany.**

2. **Jichang Wang**. Sleeping Beauty transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells (Selected oral presentation). **Conference of Transposition & Genome Engineering 2013**. September 18th-21st, 2013, Budapest, Hungary.

3. **Jichang Wang**, Manuel Grez, Zoltán Ivics, Zsuzsanna Izsvák. Gene therapy of chronic granulomatous disease with the *sleeping beauty* transposon system (oral presentation). **Annual meeting of iGenome projects**. March 15th, 2012, Frankfurt (Main), Germany.

4. **Jichang Wang**, Ivana Grabundzija, Attila Sebe, Balazs Sarkadi, Zoltán Ivics,

Zsuzsanna Izsvák. Safety of iPS cassette exchange by the *Sleeping Beauty* transposon system and Cre-mediated RMCE (Poster). **EuroSyStem: Consortium Meeting 2011**. June 6th - 07th, 2011, Prague, Czech.

5. **Jichang Wang**, Zoltán Ivics, Zsuzsanna Izsvák. Safety of iPS cassette exchange by the *Sleeping Beauty* transposon system and Cre-mediated RMCE (Oral presentation). **ReGene Meeting**. April 14th, 2011, Hanover, Germany.

6. **Jichang Wang**, Manuel Grez, Zoltán Ivics, Zsuzsanna Izsvák. Gene therapy of chronic granulomatous disease with the *sleeping beauty* transposon system (Poster). **Clinical Gene transfer: state of the art & European congress on human iPS stem cell reprogramming**. April 7th - 09th, 2011, Paris, France.