THE UNIVERSITY OF CHICAGO

GENE POTENCY REGULATION: ORCHESTRATING STEMNESS AND LINEAGE DIFFERENTIATION

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON GENETICS, GENOMICS, AND SYSTEMS BIOLOGY

BY

BOHOU WU

CHICAGO, ILLINOIS

AUGUST 2024

Table of Contents

LIST OF FIGURES

ACKNOWLEDGEMENTS

First, I extend my deepest appreciation to my advisor, Dr. Bruce Lahn, for all his guidance and support throughout my graduate career. I met Bruce during my summer research at the University of Chicago. I was impressed by his profound knowledge and engaging personalities, and deeply drawn to the beauty of his gene occlusion theory. Working alongside him, I found myself tackling increasingly complex challenges, fueling my aspiration to become a true scientist. I am equally grateful to Dr. Marcelo Nobrega, my second mentor, whose indispensable support became pivotal when the Lahn lab closed. Marcelo gave me a lot of suggestions for my research and academic career before I joined in Nobrega lab. After Lahn lab was closed, he encouraged me to insist my dream and supported me to continue my research in his lab. Without his guidance and unwavering assistance, my achievements would not have been possible. I would like to thank my thesis committee, Dr. Heng-Chi Lee and Dr. Alex Ruthenburg. They guided me through the most challenging phases of my journey. I am extremely grateful for all their advices and kindly help.

My sincere appreciation extends to the members of both the Lahn and Nobrega labs, including Jae Hyun Lee, Li Zhang, Croydon J. Fernandes, Debora Sobreira, Ivy Aneas, Robert Mitchell, Kate Farris, Noboru Jo Sakabe, Isabella Salamone, Zach Weber and Alexis Thornburg. It is my honor to have a chance to work with them. Special thanks to Jae and Li, whose steadfast support and contributions were invaluable throughout my graduate studies and this dissertation. I want to also thank Xiaoyang Dou, for her generous assistance with bioinformatics and suggestions for my career.

The past four years have been extremely difficult for me, particularly following my father's diagnosis with late-stage cancer. In this difficult period, I found solace in the unwavering support of friends, colleagues, and medical professionals. I would like to thank Yicheng Hou, Yue Liu,

Zhongyu Zou, Xianbin Yu, Yu Xiao, Cai Qi, Xiangbin Ruan, Dr. Xiaochang Zhang, Dr. Grace Yu and all my friends in the University of Chicago. It is my great fortune to meet them and have them when I need help. I want to thank Shu Yang, Qingping Li, Chunli Liu, Rongrong Hao, Xiaofang Hu, Hongbo Liu, Hongjuan Li, Kai Li, Jing Wang, Ruliang Wang, Fei Xin and all other friends, doctors and nurses who saved my father's life many times. I want to thank Xiaodan Xiao, Zuming Zhang, Li Deng, Kairui Zhao, Yanzhen Cai, Zhiyu Mao, Nannan Zheng, Xintang Gao who are struggling with me to find a cure for my father. Finally, I am profoundly grateful to my family, especially my wife, Boyang Gao, whose unwavering support and understanding have been my pillar of strength. Together, we have weathered every storm and celebrated every triumph. As I continue on my journey, I am confident that a brighter future awaits, fueled by the support and love of those who have stood by me.

ABSTRACT

Multicellular organisms are capable of generating a wide range of different cell types from the same genome through differentiation, during which distinct gene expression patterns are gradually specified along a branchwork of lineages. The highly orchestrated and robust developmental process requires tight regulation of cell potency, which limits developmental competence towards given lineages to specific stem cell populations. It remains elusive whether cell potency is actively acquired in lineage intermediates, by the activation of pioneer transcription factors, or progressively and irreversibly restricted during development. Importantly, the molecular basis of cell potency is unclear. We speculate that gene potency, which refers to the ability of a gene to be activated in response to proper transcriptional activators or developmental cues, is the foundation of cell potency. In order to measure gene potency, we took advantage of cell fusion assay, wherein genome of specific cell type was exposed to regulators from a disparate lineage. The potency of a silent gene could be evaluated by whether it could be activated post cell fusion. Transcriptome profiling of hybrid cells led to discovery of two group of silent genes: occluded genes and activatable genes. Despite both being silent before cell fusion, activatable and occluded genes showed distinct transcriptional potential in response to transcriptional activators from fusion partners. While activatable genes are readily activated after cell fusion, occluded genes remain silent even in the fusion cells where the same genes of the fusion partner are actively expressed. Fusion of the same mouse SPRET fibroblast cell line with a serious of mouse cell lines of various differentiation potential indicates that restriction of cell potency is accompanied by decreasing of activatable genes and increasing of occluded genes. Importantly, while naïve pluripotent embryonic stem cells (ESCs) are able to reprogram occluded genes into a potent activatable or active status, de-occlusion ability is disabled in primed pluripotent stem cells

(EpiSCs), which are derived from epiblast corresponding to developmental stage prior to lineage differentiation. The shutoff of de-occlusion capacity before onset of differentiation renders gene occlusion irreversible in every downstream lineage, suggesting that developmental competence is progressively restricted along developmental process. Consistently, neural differentiation of fibroblast x EpiSC fusion cells suggests that occluded fibroblast genes, but not potent EpiSC genes, cannot be activated during the lineage differentiation process, arguing against the theory of active acquirement of developmental potential in lineage intermediates. Collectively, these finding indicates the fundamental role of irreversible gene occlusion in progressive restriction of cell potency in developmental process.

Mechanismly, we found that chromatinization with recombinant histones without post-translational modifications commits a minigene to occlusion after electroporated into somatic cells, indicating that gene occlusion is the default choice of unprotected chromatin. In contrast, activatable status requires protection by binding of placeholders to the target genes. Deletion of Sox2 or Olig2 impaired target gene potency, switching activatable or expressed genes into occluded status. Based on the observations, we propose a placeholder model in regulation of gene potency during lineage differentiation process. Before differentiation, stem cells possess placeholders that bind to and prevent fully chromatinization of regulatory regions of genes to be activated in downstream lineages, assuring their activatable status. In response to differentiation signal, stem cell placeholders disappear and different groups of activatable genes are activated in different downstream lineages. Genes of alternative lineages loss protection from placeholders and are spontaneously occluded by chromatin so that they will never be activated even if their transcription activators reappear in the future.

Chapter 1: Introduction

1.1 Cell identity and cell potency

Development of complex multicellular organism starts with a single cell. Following cell proliferation and lineage differentiation, the single fertilized egg transformed progressively into a wide range of specialized cell types. Understanding this process is one of the ultimate goals of biologists.

During the developmental process, individual cells become more and more restricted in their differentiation potency. Zygote and blastomeres during the early cleavage of the embryo are totipotent, able to generate both embryonic and extra-embryonic cells. Slightly restricted in developmental potential is pluripotent stem cells, which is capable of producing all somatic lineages but not extra-embryonic tissues. The first pluripotent stem cells isolated from normal embryo are mouse embryonic stem cells (ESCs) [1]. Mouse embryonic stem cells correspond to the inner cell mass (ICM) of preimplantation embryo. When injected into blastocysts, they contribute cells to tissues originated from all three germ layers, as well as to germlines, of chimeric animals. Pluripotency is retained in the epiblast of implanted embryo. Explanting epiblast of postimplantation mouse embryo give rise to an independent pluripotent cell line named EpiSCs [2][3]. Similar to ESCs, EpiSCs display full developmental potential evaluated by multilineage differentiation in embryoid bodies and teratomas. However, EpiSCs also possess features distinct from ESCs, such as flat colony morphology, different requirement of signals for pluripotency maintenance, X chromosome inactivation and extremely low efficiency in generating chimeras. Therefore, EpiSCs are usually called primed pluripotent stem cells, as opposed to ESCs that are commonly referred as naïve pluripotent stem cells. Importantly, it is reported that EpiSCs do have the capacity to generate chimeras when grafted to postimplantation rather than blastocysts or

morula [4], or supported by FGF4 in culture medium before embryo injection [5]. Those observations further validated the pluripotency of EpiSCs. Derivation of human embryonic stem cells from blastocysts was reported in 1998 [6]. Interestingly, by a variety of criteria, the so called human ESCs are more similar to mouse EpiSCs. Naïve pluripotent, or ground state, human ESCs were obtained in later studies [7]. Further development of pluripotent cells leads to diversification of germ layers and subsequent production of a series of multipotent stem cells, unipotent progenitor cells and, finally, terminally differentiated cells. Notably, in contrast to the view of gradual restriction of cell potency, it has been proposed that developmental competence is actively acquired in the multipotent lineage intermediates, by the function of pioneer transcription factors such as FOXA1 and FOXA2, rather than remnant from upstream stem cells [37]. Highly specialized terminally differentiated cells, that possess no potency to other cell fates, are the major bearers of structural and functional work of adult multicellular organisms. Nonetheless, cells with different level of potency are usually retained in the adult tissue. Those cells, usually referred as somatic stem cells or adult stem cells, play important roles in homeostasis, tissue repair and regeneration. In mammals, the olfactory epithelium contains a population of somatic stem cells that is able to produce neurons called olfactory receptor cells. Therefore, olfactory epithelium is capable of self-renew and recover from tissue damage. In contrast, the auditory epithelium and retinal epithelium lack stem cells. Consequently, cell loss due to damage or degenerative processes during aging is permanent to them [8]. Notably, some species, such as planarians, possess pluripotent somatic stem cells in the adults [9], enabling them to regenerate essentially any missing body part.

Reprogramming and transdifferentiation challenge the traditional view of irreversible cell differentiation. Four approaches have been reported to reprogram somatic cells to a pluripotent

state: nuclear transplantation, cell fusion, overexpression of pluripotency factors and chemical stimulation [10][11]. In 2006, Takahashi and Yamanaka reported the groundbreaking discovery that forced expression of four transcription factors, Oct4, Sox2, Klf4 and c-Myc, converted somatic fibroblasts to induced pluripotent stem cells (iPSCs) [12]. Since then, the reprogramming field quickly expanded, leading to advances not only in efficiency and applicability, but also understanding of the mechanism. On the other hand, transdifferentiation studies apply approaches similar to reprogramming, such as ectopic expression of transcription factors and modification of culture conditions. Despite success in several systems, however, transdifferentiation practices generally suffer from aberrant gene expression patterns of induced cells and incomplete switch into new lineages. Conversion of fibroblasts into muscle cells by overexpressing Myod represents one of the earliest achievements of transdifferentiation. Notably, the endogenous Myod gene was not activated in transdifferentiated muscle cells, indicating incomplete cell fate transfer [13]. Induced melanocytes from fibroblasts [14], induced macrophages from B cells [15], induced neurons from fibroblasts [16] as well as many other studies all share similar problems that transdifferentiated cells reserve memories from original identities and are not completely switched to new fates. A key difference between reprogramming and transdifferentiation is that the former employs the natural ability of early embryo to reset cell potency while the latter attempts to create an unnatural process. This may explain the greater difficulties and inconsistency of transdifferentiation practices compared to reprogramming studies.

Overall, global view of lineage differentiation of developmental process has been depicted in several model organisms. Experimentally, cells of various potency have been characterized both in vivo and in vitro. Practices of manipulating cell potency and identities were also proven successful to various extents. Despite those achievements, however, the underlying mechanisms

are still poorly understood. It remains unclear whether there is a grant unified theory to explain cell differentiation process.

1.2 Epigenetic regulation of gene expression and cell differentiation

Advances in molecular biology, high throughput sequencing techniques and, especially, achievements in animal cloning work, lead to the common agreements that cell identities within multicellular organism is determined by differential gene expression rather than different genetic material. Epigenetics, focusing on inheritable gene expression regulation without altering DNA sequences, provides an attractive view to understand cell potency, identity and differentiation process. Numerous mechanisms were reported to be involved in epigenetic gene regulation, including DNA modifications, histone modifications, chromatin accessibility, chromatin structure, R-loop formation, noncoding RNAs and phase separation. From a broader view, differential gene expression due to transcription factors or RNA modifications also fall into the category of epigenetics. Epigenetic regulation, especially chromatin modifications, provides appealing explanations for cell identity maintenance and inheritance due to its feature of stability. However, it is still unclear whether the epigenetic modifications are leaders regulating gene expression or followers regulated to reinforce already established gene expression pattern. Importantly, the well characterized epigenetics regulators generally lack specificity by themselves. It is difficult to explain how they select target genes or functional cis elements from the beginning. Consistently, unlike the successes in explaining epigenetic profile maintenance, models explaining establishment of the profile are unsatisfactory.

Discovered in 1948, DNA methylation, C5-methylcytosine (5mC) to be specific, remains to be the best-studied epigenetic modifications. It is essential for important biological events such

as imprinting, X chromosome inactivation and silencing of transposons [18][19]. In mammals, DNA methylation is dynamically regulated during the developmental process. Primordial germ cells first undergo genome-wide de-methylation [20], which is followed by sex-specific re-establishment of DNA methylation [21]. After fertilization, the paternal and maternal genomes experience global DNA de-methylation through active and passive mechanisms [22]. Subsequent re-establishment of DNA methylation quickly occurs after implantation [23]. Notably, although DNA methylation is mostly regarded as a repressive modification, it is reported to associate with active genes in some circumstances [24][25]. Histone modifications represent a larger and more complicated category of epigenetic marks. There are more than 130 sites on five canonical and around 30 histone variants that undergo at least 12 post-translational modifications [26]. Among them, methylation of H3K4 and acetylation of H3K27 are well known for their association with active transcription, while methylation of H3K27 and H3K9 are wide spreading marks for silent genes and heterochromatin. Similar to DNA methylation, histone modifications are important for cell identity maintenance and are dynamically regulated during developmental process. An intriguing phenomenon worth mentioning is the existence of bivalent domains, which refers to the co-occupancy of H3K4me3 and H3K27me3 at the same genomic regions. Bivalent domains were first described in mESCs [27], which was found to be associated with enhancers and promoters. The co-occurrence of permissive and repressive marks was reported in lineage-committed stem cells and terminally differentiated cells in subsequent studies [28][29]. Functionally, bivalent domains were proposed to reflect gene potency. Genes associated with bivalent modifications are generally silent but poised for activation in stem cells. Upon differentiation, bivalent genes become either fully activated or repressed in different in different lineages. Bivalent modifications provide an explanation for developmental potential by linking cell potency to gene potency. However, the

causal relationship between bivalency and gene potency was unclear. Additionally, the potency of genes devoid of bivalent modifications is not able to be explained by bivalency. Most importantly, bivalent domains remain to be descriptive. It is unknown how bivalent domains are established in stem cells and regulated during differentiation specifically at target genes. In addition to DNA and histone modifications, chromatin structure, accessibility, noncoding RNAs, R-loop formation and phase separation of chromatin regulators added new layers of complexity to epigenetic regulation. Notably, most epigenetic studies focus on the relationship between epigenetic features and gene activity. Regulation of gene potency has been limited to description of bivalent modifications and a few chromatin features [30].

The inheritance of epigenetic modifications, which is important for maintenance of cell identity, has been extensively investigated. DNA methylation represents one of the most faithful epigenetic marks inherited during cell division. DNMT1, methyltransferase mediating addition of 5mC to DNA, recognizes methylation of CpGs on the parental strand and methylates that on the daughter strand, thus copying the DNA methylation patterns to the daughter cells [31]. In addition to the replication-coupled mechanism of DNA methylation transmission, a replication-uncoupled mechanism mediated by interaction between UHRF1 and methylated H3K9 was used to further guarantee faithful inheritance of DNA methylation [32]. Similar to DNA methylation, H3K9me3 and H3K27me3 utilize machineries with intrinsic property of self-propagation as well. SUV39H1 and Polycomb repressive complex 2 (PRC2), major writers responsible for H3K9me3 and H3K27me3 respectively, are able to recognize modifications on the parental histones and modify newly incorporated histones during mitosis [33]. Intriguingly, the modifications with intrinsic mechanisms of inheritance are mostly repressive. Active modifications generally lack the ability to copy parental epigenetic information to newly synthesized chromatin during DNA replication.

Consistently, parental nucleosomes of the repressive chromatin are distributed locally to the same genomic locus of the daughter chromatin after DNA replication, preserving epigenetic information. In contrast, nucleosomes of the active chromatin are dispersed after DNA replication, making the parental modifications un-trackable [34]. Nevertheless, it is reported that accessibility of active chromatin is re-established within 30 minutes to 2 hours, which is faster than restoring of repressive histone modifications [35]. Importantly, the transcriptionally active RNA Pol II complex retains proximal to nascent DNA strands after replications [38]. In yeast, transcription factors transiently evicted by DNA replication machinery quickly re-bind to target regions within minutes [36]. Those observations indicate that active epigenetic information can be faithfully transmitted during DNA replication. But unlike repressive information, this process is likely to be achieved by preserving chromatin accessibility rather than copying chemical modifications on parental nucleosomes.

1.3 Gene occlusion

Cell fusion assay has been utilized to study the stability an of cell identities, which is essentially the plasticity of lineage specific gene expression patterns [39][40]. Detailed measurements of transcripts expressed from each genome within the hybrid cell revealed existence of distinct gene silencing status: activatable and occluded [41]. Albeit not expressed, activatable genes are transcriptionally competent. They can be activated in the hybrid cells where proper transcriptional activators are provided. On the contrary, occluded genes represent a more stable silencing mode. They remain silent in fusion cells although the same genes are actively transcribed from the fusion partner. Systematic mapping of occluded genes by cell fusion assay revealed that master regulators of cell fates, exemplified by Myf5 and Myod1, were under regulation of occlusion in improper cell types. Importantly, when the endogenously occluded genes were

introduced into corresponding cells by BAC transgenes, which were devoid of any eukaryotic epigenetic modifications, the transgenes were robustly expressed in most cases [42]. This observation suggested that the cellular environment of somatic cells was supportive to expression of considerable amount of lineage inappropriate genes. Without gene occlusion, improper expression of master regulators of alternative lineages would possibility alter the original cellular characteristics, which was confirmed by the Myf5 BAC transgene induced myotube formation in fibroblasts [42]. Collectively, gene occlusion was suggested to be critical for safeguarding cell identities.

The development of metazoan begins with rebooting the epigenetic profile of highly specialized zygotic genome to build pluripotency [46]. Reminiscent of the global reprogramming event, ESCs possess remarkable de-occlusion ability [44], which is absent in terminally differentiated somatic cells or multipotent stem cells. Intriguingly, in contrast to the fast activation of activatable genes following cell fusion, reprogramming of occluded genes is a progressive event which requires DNA replication. The kinetical difference in terms of activation between occluded genes and activatable genes confirmed the stability of gene occlusion. The regulation of occlusion during development and its role in cell identity maintenance led to the occlusis model in explanation of lineage differentiation [45]. Compared to the transcriptome view of cell identity, which focus on gene expression, occlusis model emphasizes gene potency. According to the occlusis model, the capacity of stem cells to differentiate into multiple lineages is based on the activatable status of genes to be activated in downstream cell fates. As differentiation process going on, different set of genes are occluded, progressively restricting cell fates.

Although the occlusis model provided a descent explanation for lineage differentiation, enabling quantitative evaluation of cell potency according to occludome, the molecular basis of gene occlusion remains elusive. The observation that BAC transgenes carrying endogenously occluded genes are actively expressed in corresponding cells leads to the assumption that additional epigenetic modifications are required for gene occlusion. Importantly, when the BACs were introduced into embryos and assayed for their expression after the transgenes experienced the differentiation process, they were occluded similar to endogenous genes in improper lineages [42]. Accordingly, gene occlusion was proposed to be established by a machinery adding epigenetic modifications during the differentiation process. Nonetheless, the hypothesized occlusion machinery has to be disabled after differentiation to accommodate to the observation that BAC transgenes were not occluded in post-differentiation cells. In the cases of stepwise differentiation, the activity of occlusion machinery has to be switched on and off repeatedly, which requires complicated regulatory mechanisms that is difficult to achieve robustness. Additionally, intensive profiling of epigenetic modifications failed to distinguish occluded and activatable genes [43][47]. Consistently, treatment of somatic hybrid cells with inhibitors for DNA methyltransferase or histone deacetylase only affected limited number of occluded genes. Therefore, understanding of occlusion mechanisms and regulation requires more elegant studies.

Chapter 2: The role of gene potency reduction in lineage restriction

2.1 Introduction

Lineage differentiation, the basis of complexity of multicellular organisms, is characterized by two complementary features: specification of cellular identities and restriction of cell potency. While the former endows various cell types with specific biological functions, the latter ensures the orderliness and regulatory robustness of developmental process. It has been well established cell identities are determined by specific gene expression patterns, which require cooperation of master regulators and shaping of epigenetic features at key cis-regulatory elements such as super-enhancers [48][49]. However, the understanding of cell potency regulation is limited. Both trans-acting factors and cis-acting epigenetic mechanisms have been investigated to explore the molecular basis and regulation of cell potency. Transcriptome comparison of various stem cell lines and terminally differentiated cell lines were conducted in search of factors governing stemness, or developmental potency [50]. While components of typical biological processes, such as JAK/STAT signaling and cell cycle regulators, were enriched in stem cells, the overall results were hard to interpret. Additionally, searching of "common" stem cell factors should not benefit the understanding "differential" cell potency. In addition to trans-acting factors, numerous researches were focused on the chromatin features associated with developmental competence. ESCs, the pluripotent stem cell capable of generating every somatic cell type, possess unique chromatin characteristics compatible with their remarkable cell potency, such as higher accessibility, lower DNA methylation, and bivalent marks [28][51]. It has been proposed that the epigenetic features in stem cells are associated with transcriptional competence of lineage specific

genes for future activation [52][53]. Nonetheless, the relationship between gene transcriptional potency and cell developmental potency remains unclear. Importantly, experimental definition and evaluation of gene potency is challenging.

Cell fusion assay of disparate somatic lineages provides a feasible approach to measure gene potency [41-43]. Upon cell fusion, transcriptional regulators from an alternative cell type are provided. Then the transcription potency of silent genes in unfused cells could be measured by whether they can be activated in hybrid cells. Importantly, the expression of the corresponding genes from the fusion partner serves as intrinsic controls to guarantee that the hybrid cells are supportive to their transcription. However, there are two caveats in previous cell fusion assays for gene potency measurement. First, cell lines from different species were utilized in most cases to distinguish transcripts expressed from two genomes in the hybrid cells. Therefore, failure of a gene to be activated is possibly due to species incompatibility rather than lacking transcriptional potency. More importantly, studies were focused on evaluation of gene potency in terminally differentiated cells in order to explore robustness of cell identities. Gene potency in stem cells with broader developmental competence were not systematically assessed, leaving the relationship between gene potency and cell potency unresolved.

2.2 Results

2.2.1 Occluded and activatable genes represent two distinct status of gene potency in cell fusion assay

In order to accurately measure gene potency, we developed a same-species cell fusion system. Mouse strains that are distantly related to widely used lab strains were utilized to derive somatic cell lines. After cell fusion, the abundant genetic variants between fusion partners can be used to distinguish transcripts expressed from the two genomes within the hybrid cells. As a proof

of concept, we constructed an ear fibroblast cell line from mouse SPRET/EiJ, which was immortalized by SV40T antigen and amplified from a single clone for homogeneity (SEF2). Different fluorescence and drug selection markers were introduced to SEF2 and the cell lines to be fused with. Cell fusion were triggered by PEG treatment and hybrid cells were selected according to the presence of both selection markers from two cell lines **(Figure 2.1A)**. Fusion cells were cultured for more than two weeks prior to analysis of gene expression. The benefit of extended culture is two-fold. First, continuous cell division ensures the genomes of fusion partners to be packaged into the same nucleus. Therefore, the comparative gene expression in the same context would more faithfully reflect transcriptional potency. Second, for the genes that were silenced after cell fusion, long-term culture enables elimination of remnant transcripts by degradation and dilution through cell proliferation. Such that the quantification of gene expression within the hybrid cells would be more precise. We first fused SEF2 with C2C12, a myoblast cell line derived from mouse strain C3H. RNA-seq was performed for both unfused and hybrid cells and reads were separated according to the single nucleotide polymorphisms (SNPs) between the two mouse strains. House-keeping genes that are expressed in both SEF2 and C2C12, exemplified by Tubb5, showed reliable expression from both genomes within the fusion cells. In contrast, their transcription was observed exclusively from only one genome in the unfused cells, validating the accuracy of the data processing pipeline **(Figure 2.1B)**. Consistent with previous studies of inter-species cell fusion [41-43], silent genes in unfused cells showed remarkable difference in transcription potency. Arhgap32, a gene that was expressed in C2C12 but not SEF2, was activated in SEF2 genome upon cell fusion. In contrast, Ifitm1 remained silent in SEF2 genome after cell fusion, albeit the same gene was actively transcribed from the C2C12 genome within the hybrid cells **(Figure 2.1B)**. The transcriptional competent genes like Arhgap32 were termed activatable

genes, while those lack gene potency similar to Ifitm1 were termed occluded genes. Measurement of gene potency by cell fusion was possibly confounded by chromosome. Genes on the lost chromosome would never be activated in the hybrid cells, and are likely to be mis-annotated as occluded genes. In order to address this possibility, we quantified strain-specific reads for each chromosome. Reads of the whole chromosome will be assigned to one strain in the cases of chromosome loss, which was not observed in SEF2xC2C12 fusion cells **(Figure 2.1C)**. Similar chromosome loss analysis was performed for all fusion samples in this study and hybrid cells with signs of chromosome loss were excluded. We next quantified strain specific expression of all informative silent genes that are expressed exclusively in SEF2 or C2C12 before cell fusion, and annotated activatable and occluded genes for both cell types **(Figure 2.1D)**. Similar to previous studies, numerous genes became silent in both genomes post cell fusion [43]. The potency of those genes, called extinguished genes, cannot be evaluated by this fusion pair.

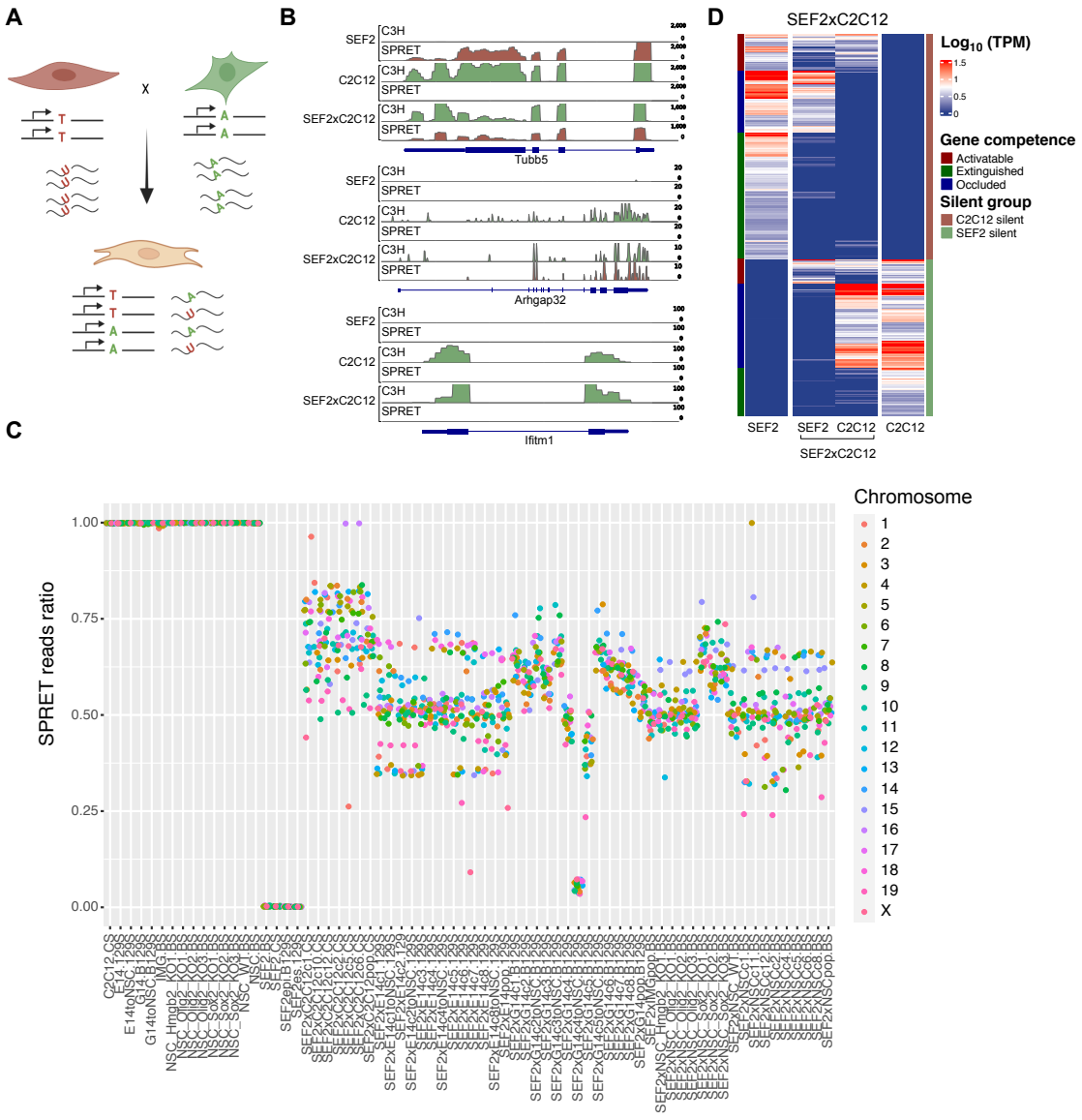

Figure 2.1 Strain-specific RNA-seq of fusion cells measures gene potency.
(**A**) Schematic of cell fusion assay. (**B**) Strain-specific RNA-seq profile of representative constitutive gene (Tubb5), SEF2 activatable gene (Arhgap32) and SEF2 occluded gene (Lfitm1) before and after cell fusion. Only reads that can be confidently separated to SEF2 genome (SPRET) or C2C12 genome (C3H) were reserved. (**C**) Ratio of reads mapping to SEF2 genome (SPRET) for each chromosome in fused and unfused cells used in this study. Fusion clones SEF2xC2C12c1, SEF2xC2C12c5, SEF2xC2C12c6, SEF2xNSCc11, SEF2xE14c7, and NSC like cells differentiated from fusion clone SEF2xG14c4 (SEF2xG14c4toNSC) show pattern of chromosome loss. They are excluded from downstream analysis. (**D**) Heatmap of informative silent gene (n = 738) expression in unfused SEF2 and C2C12 cells, as well as strain-specific expression in fused cells. Informative silent genes in SEF2 or C2C12 were classified into activatable (TPM before fusion < 1, TPM after

(**Figure 2.1, continued**) fusion ≥ 30% of the total TPM from the two strains before fusion, total TPM from the two strains after fusion ≥ 2), occluded (TPM before fusion < 1, TPM after fusion < 10% of the total TPM from the two strains before fusion, total TPM from the two strains after fusion ≥ 2) and extinguished (total TPM from the two strains after fusion < 2) genes based on their expression pattern prior and post cell fusion.

In order to test the robustness of gene potency measurement, we derived another fibroblast cell line from mouse CAST/EiJ (CTF3) and conducted cell fusion with C2C12. Similar to SEF2xC2C12 fusion, activatable and occluded genes in CTF3 and C2C12 were annotated according to their expression prior to and post cell fusion **(Figure 2.2A)**. We compared the potency annotation of C2C12 genes from two fusion pairs. Due to transcriptional difference between SEF2 and CTF3, we did not see much overlap of C2C12 acctivatable genes annotated from two fusions. Nonetheless, of 120 occluded genes identified in SEF2xC2C12 fusion, 28 were annotated as occluded while only one was classified as activatable in CTF3xC2C12 fusion, suggesting high consistency of gene potency measurement **(Figure 2.2B)**. We selected Myf5, a master regulator gene of muscle lineage, for further validation. Sanger sequencing was performed for genomic DNA PCR and RT-PCR product of Myf5 from CTF3xC2C12 fusion cells. When focusing on a SNP between CTF3 and C2C12 genome, we observed equal signal from both strains in genomic DNA PCR product, excluding the possibility of chromosome loss. In contrast, only C2C12 signal was detected in RT-PCR product, suggesting that Myf5 was transcribed exclusively from C2C12 genome in the hybrid cells **(Figure 2.2C)**. Collectively, the above results revealed the reliability of gene potency measurement through same-species cell fusion assay.

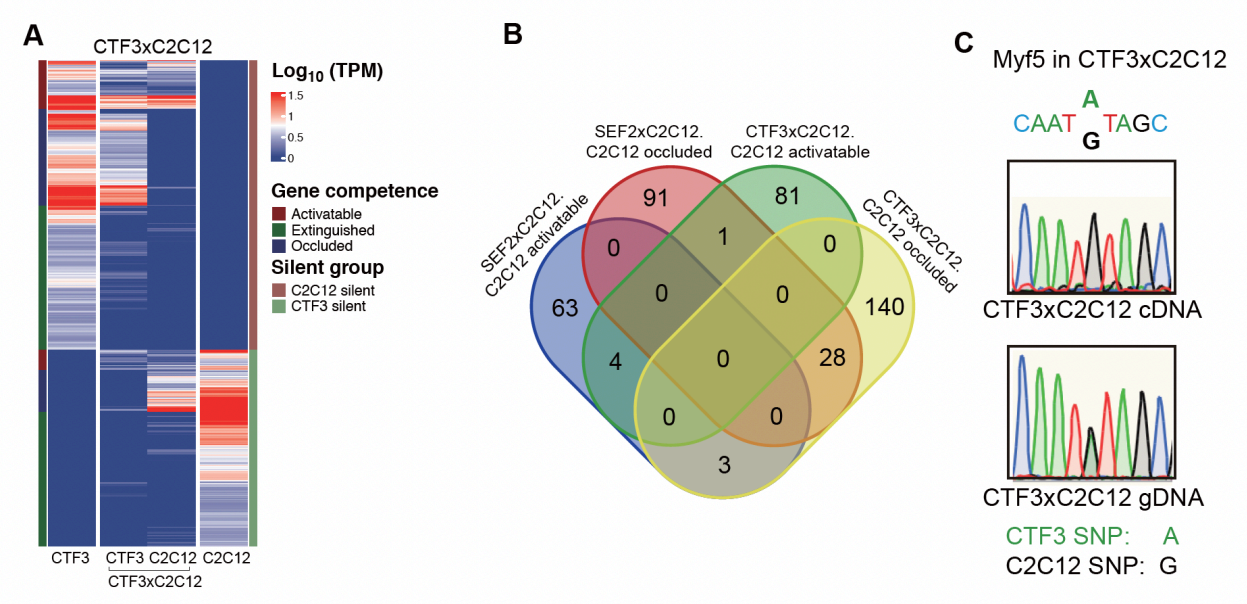

Figure 2.2 CTF3xC2C12 fusion validates the robustness of cell fusion.
(**A**) Heatmap of informative silent gene (n = 858) expression in unfused CTF3 and C2C12 cells, as well as strain-specific expression in CTF3xC2C12 fusion cells. Activatable, occluded and extinguished genes are annotated based on their expression before and after cell fusion. (**B**) Venn diagram of C2C12 activatable and occluded genes annotated by fusion with SEF2 or CTF3. (**C**) Sanger-sequencing of RT-PCR (CTF3xC2C12_cDNA) and genomic DNA PCR (CTF3xC2C12_gDNA) products of Myf5, an occluded gene in CTF3, in CTF3xC2C12 fusion cells. Relative abundance of Myf5 genomic DNA and RNA transcripts in the fusion cells are reflected by the signal at the SNP site.

2.2.2 Restriction of cell potency accompanies decreasing gene potency during differentiation

We next sought to explore the relationship between gene potency and cell potency. Cell lines of various developmental competence were selected to measure their gene potency by cell fusion assay. The embryonic stem cell line E14 cells and epiblast stem cell line G14 cells, originating from inner cell mass of preimplantation embryos and epiblast of postimplantation embryos respectively, represent pluripotency that is capable of generating all somatic lineages. Neural stem cells (NSC), myoblast (C2C12), microglia (IMG) and fibroblast cell lines represent more restricted developmental potential **(Figure 2.3A)**. Remarkably, while the majority of silent

genes in G14 were activatable, more genes were occluded in NSC, C2C12, IMG and CTF3 **(Figure 2.3B&C)**. This observation suggested that the restriction of developmental competence accompanies reduction of gene potency. Notably, the measurements of gene potency in E14 cells were problematic. Due to the reprogramming capacity of embryonic stem cells, the majority of SEF2 specific genes became extinguished within the fusion cells. And the potency of the remaining genes was ambiguous **(Figure 2.3D)**. We reasoned that reprogramming is a gradual process which requires multi-rounds of cell division. Therefore, the potency of E14 genes might be able to be evaluated in early days post cell fusion. We re-analyzed the time-course cell fusion data where rat fibroblast cell line R1A was utilized to hybrid with E14 **(Figure 2.3E)**. Owing to higher efficiency of R1AxE14 fusion, adequate hybrid cells could be obtained to assay for gene expression as early as day 2 post cell fusion. Consistent our hypothesis, while most of gene became extinguished after long-term culture, the potency of E14 genes could be revealed in early-stage fusion samples. Similar to G14 cells, the majority of E14 silent genes were activatable. Importantly, the fraction of E14 genes tend to be underestimated. The reason was twofold. First, the incapability of E14 genes to be activated upon fusion with R1A could be consequent from species incompatibility rather than lack of gene potency. Second, the apparent expression of R1A genes in early-stage hybrid cell did not necessarily indicate a supportive environment for the genes to be transcribed. Alternatively, the R1A genes may already ceased expression while the remnant RNAs were not eliminated. In such cases, differential gene expression within early-stage fusion cells may not faithfully determine the occlusion status of E14 genes.

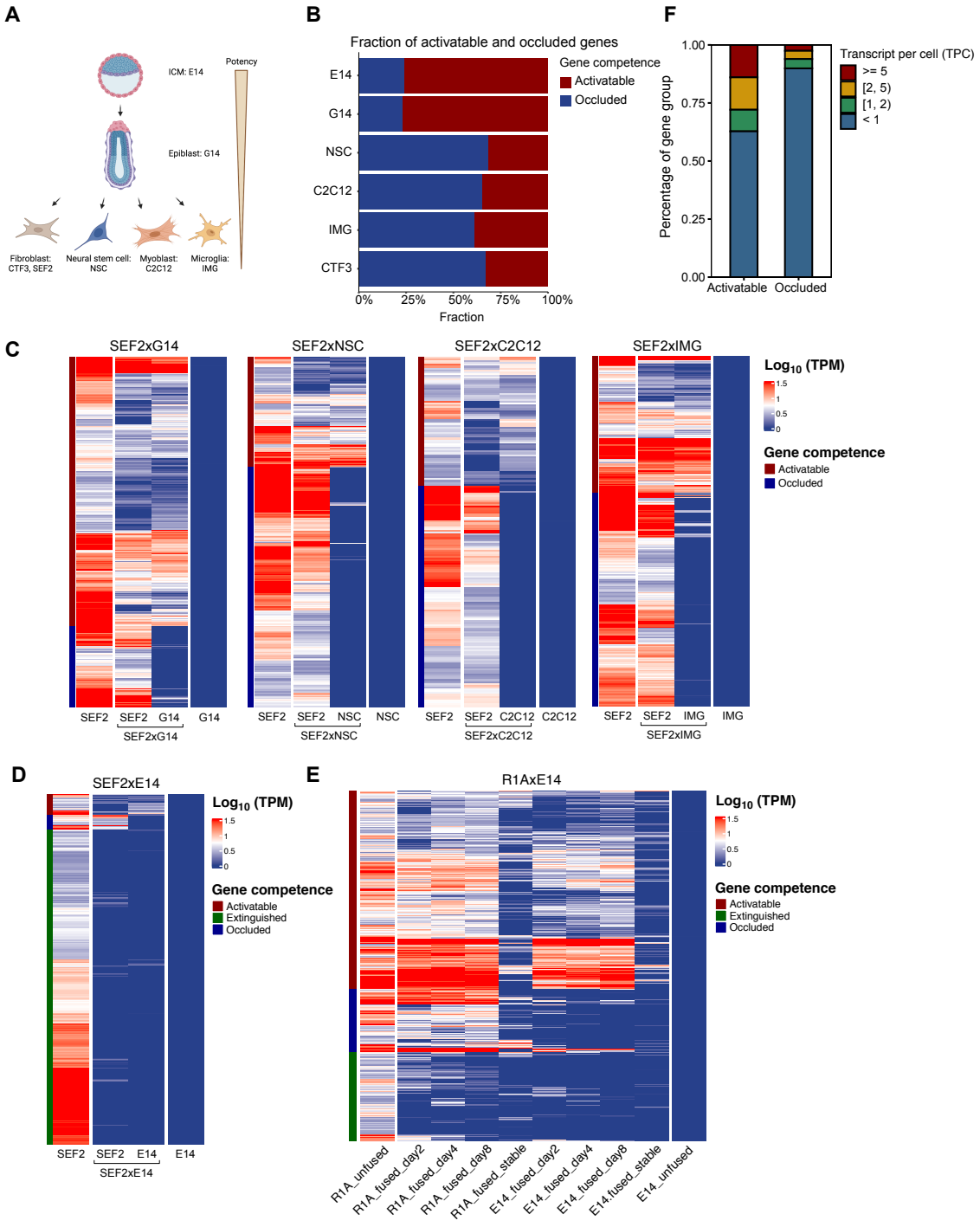

Figure 2.3 Decreasing gene potency associates with cell fate restriction

(**A**) Cell lines of different potency used in cell fusion assay. (**B**) Fraction of activatable and occluded genes in cells with different potency. Informative silent gene potency in G14 (n = 427), NSC (n =426), C2C12 (n = 190) and IMG (n = 652) were measured by fusion with SEF2 cells.

(**Figure 2.3, continued**) Potency of E14 (n = 362) and CTF3 (n = 110) informative silent genes were measured by fusion with R1A and C2C12 cells, respectively. Extinguished genes were not included in the analysis. (**C**) Heatmap of informative silent gene expression in G14 (n = 427), NSC (n = 426), C2C12 (n = 190), IMG (n = 652) unfused cells and their fusion cells with SEF2. (**D**) Heatmap of E14 informative silent gene (n = 788) expression in unfused SEF2 and E14 cells, as well as strain-specific expression in SEF2xE14 fusion cells. (**E**) Time-course expression of E14 informative silent genes (n = 485) in R1A, E14 and R1AxE14 fusion cells after cell fusion. Activatable, occluded and extinguished genes are annotated according to their expression before cell fusion and at day 8 post cell fusion. Only reads that can be confidently separated to mouse or rat genome are reserved to quantify gene expression (transcript per cell, TPC). (**F**) Fraction of activated NSC occluded (n = 198) and activatable (n = 86) genes in Astrocytes differentiated from NSCs.

In order to test the homogeneity of potency measurements by cell fusion assay, we isolated single clones of SEF2xC2C12 and SEF2xNSC fusion cells. The strain-specific expression profile of informative silent genes, that were expressed in only one of the pre-fusion cell types, showed high similarity among fusion clones **(Figure 2.4A)**. Notably, the correlation of expression from the same genome before and after fusion is greater than between fusion partners within fusion cells. However, in SEF2xE14 fusion cells, expression profile of both genomes resembled E14 transcriptome, which is highly consistent in all fusion clones **(Figure 2.4B)**. Collectively, the observations validated the robustness of gene potency measurements. In order to further explore the correlation between gene potency and cell potency, we checked the expression of NSC activatable and occluded genes after differentiation to astrocytes **(Figure 2.3F)**. Intriguingly, while 14.0% (12 out of 86, TPC $\geq$ 5) of NSC activatable genes were activated upon differentiation, only 2.5% (5 out of 198, TPC $\geq$ 5) of NSC occluded genes were turned on. According to this result, we proposed that gene occlusion is irreversible during cell differentiation, which led to restriction of cell potency during developmental process.

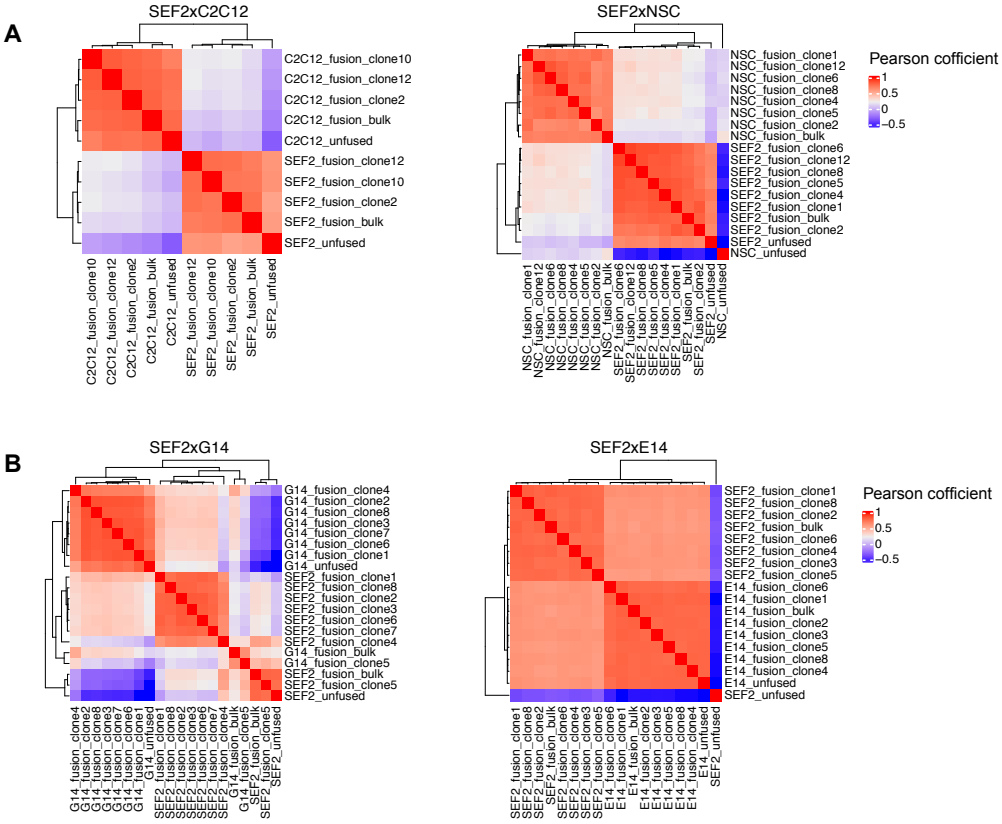

Figure 2.4 Gene expression patterns in fusion clones are similar to fusion populations. Heatmap of pairwise Pearson coefficient of strain-specific gene expression in cells fused by SEF2 and somatic stem cells (**A**) or pluripotent stem cells (**B**). Unfused cells, bulk fusion cells and fusion clones without chromosome loss are hierarchically clustered according to strain-specific expression of informative silent genes (SEF2xC2C12, n = 661; SEF2xNSC, n = 1354; SEF2xG14, n = 1274; SEF2xE14, n = 1488).

2.3 Conclusion and discussion

Compared to the extensive researches on the regulation of gene expression, much less efforts were put on the regulation of gene potency. We reasoned that developmental competency is essentially the capacity of the genes to be specified into expression profiles of downstream lineages, albeit being silent before differentiation. Therefore, transcription potency rather than expression status was speculated to be the basis of cell potency, or stemness. With the aim of more

precisely measuring gene potency, we developed a same-species cell fusion system. The transcriptional potency of silent genes could be evaluated by whether they are responsive to transcriptional activators provided by the fusion partners. Fusion of SEF2 with a serious of cell lines with different potency suggested that restriction of gene potency associates with reduction of gene potency. Importantly, as indicated in the in vitro differentiation experiment, activatable genes but not occluded genes were prone to be activated post differentiation.

Notably, while the same species cell fusion assay provided a relatively faithful measurement for gene potency, there were indeed caveats. First, the minimal unit under regulation of occlusion might be cis-regulator elements rather than genes. It is possible that only a subset of enhancers of a gene were occluded. When fused to different cell lines where transcriptional activators were supporting distinct subgroup of enhancers of this gene, it can be annotated as activatable in one fusion pair while classified as occluded in another fusion pair. Second, occlusion, or lacking of gene potency, is essentially the inability of a cis regulator element to respond to regulators. In addition to enhancers that cooperate with activators, there were abundant silencers that repress gene expression with the assistance of transcriptional repressors [54] [55] [56]. When an active gene within fusion cells possesses occluded silencers that are not responsive to repressors, the silent gene of the fusion partner would be annotated as occluded gene. However, the genes might be activatable when the repressors were downregulated or proper transcriptional activators were provided. Collectively, while an activation of a gene upon cell fusion faithfully indicated its potency, occlusion annotation might be inaccurate. These caveats may provide an explanation for some of the observations in the cell fusion results, such as the small number of inconsistent annotations of C2C12 gene potency between SEF2xC2C12 and CTF3xC2C12 fusion pairs, a few activated “occluded” NSC genes upon differentiation, as well as the small number of “occluded”

genes annotated in pluripotent G14 cells. They are all likely due to false annotation of occluded genes. Nonetheless, the relationship between gene potency and cell potency was clearly revealed.

2.4 Methods

2.4.1 Construction of fibroblast cell lines from mouse SPRET/EiJ and CAST/EiJ

Primary ear and tail fibroblast cells were derived from adult SPRET/EiJ and CAST/EiJ mouse respectively [57]. The primary cells were passaged once and transduced with lentivirus expressing simian virus 40 large T antigen (SV40T) to construct immortalized cell lines. The immortalized cells were sorted into 96-well plates as single cells to derive single clones. SEF2 and CTF3, clones selected for cell fusion assay, were introduced with fluorescence and drug resistance markers by lentivirus. An additional rounds of single clone selection was performed after addition of markers to further enhance homogeneity of the cell lines.

2.4.2 Cell culture and cell fusion

SEF2, CTF3, C2C12 and IMG cells were cultured in DMEM with 10% FBS. E14 cells were cultured under feeder-free conditions in Knockout DMEM supplemented with 10% FBS, non-essential amino acids, sodium pyruvate, penicillin/streptomycin, β-mercaptoethanol, 3uM CHIR99021 and 1uM PD0325091. Epiblast stem cell G14 were cultured in 10%FBS coated plates with DMEM/F12 supplemented with 20% Gibco Knockout Serum Replacement, GlutaMAX™, β-mercaptoethanol, penicillin/streptomycin, 12ng/mL FGF2, 20ng/mL ActivinA, 10uM Y27632 and 2uM IWP-2. NSCs were cultured as monolayers in CELLstart™ substrate coated plates with DMEM/F12 supplemented with N2, B27, GlutaMAX™, penicillin/streptomycin, 20ng/mL FGF2 and 20ng/mL EGF.

Before cell fusion, SEF2 or CTF3 were cultured in conditions of their fusion partners for a week. The cells to be fused were trypsinized and neutralized by trypsin neutralizing solution. Following resuspension with medium, the two cell lines were mixed thoroughly at 1:1 ratio and passage into 6-well plates at high density to enhance cell-cell contacts. The cells were settled for 2 hours to allow them to attach, and then treated with 45.5% PEG1000 that were pre-warmed to 42℃. After 1 minute of PEG treatment, the cells were washed with fresh medium three times and cultured 2 days. The hybrid cells were selected by double-drug selection or fluorescence activated cell sorting.

2.4.3 In-vitro differentiation of NSCs into astrocytes

The NSCs were trypsinized and neutralized by trypsin neutralizing solution. $5x10^5$ cells were seeded on poly-D-lysine coated 10cm dishes and cultured 7 days with astrocytes medium containing: DMEM/F12, N2, B27, penicillin/streptomycin, GlutaMAX™, 1ng/mL EGF and 20ng/mL BMP.

2.4.4 RNA-seq and strain-specific data analysis

Total RNA was extracted by MagNA Pure Compact RNA Isolation kit. Following DNase treatment, mRNA with polyA tails was purified with NEBNext® poly(A) mRNA Magnetic Isolation module. Purified mRNA was reverse transcribed and the generated cDNA was constructed into libraries using illumina primers sets.

Over 30 million high-quality 2x150bp paired-end reads for each sample were obtained. The reads were aligned to N-masked mm10 mouse genome where SNPs between genomes of fusion partners were replaced with the ambiguity base ‘N’. SNPsplit was used to extract reads that

were specific to one strain. TPM (transcripts per million) or TPC (transcripts per cell) was calculated based on the relative amounts of strain-specific reads.

Chapter 3: Irreversible loss of gene potency during lineage differentiation

3.1 Introduction

The development of multicellular organism is a highly orchestrated process during which a wide range of cell types were generated along a branch-work of lineages in a spatiotemporal regulated manner. In humans, hundreds of cell types and over twenty thousand of genes in each single cell need to be properly controlled to build the whole organism. Despite the complexity, the developmental process is strikingly reliable. Restriction of developmental potential during differentiation is natural solution of complicated regulation, exemplified by the induction of neuronal fates in mammals. Signals from the mesoderm induced neuronal differentiation of ectoderm. Remarkably, only ectoderm cells of certain developmental stages respond to the signals [58][59]. The restriction of developmental competence to specific cell groups ensures proper interpretation of signals in the complicated tissue environments and during fast developmental processes.

Measurements of gene potency across cell lines with various developmental potential suggested the correlation between decreasing gene potency and restriction of cell potency. This observation implied that developmental potential is gradually narrowed down during lineage differentiation. However, it was claimed that cell potency was actively acquired in lineage intermediates rather than inherited from parental cells[37][61-63]. During pancreatic and hepatic differentiation process, pioneer transcription factors FOXA1 and FOXA2 were proposed to reshape the chromatin environment. This epigenetic priming process endow target genes with transcriptional potency to be activated, and consequently made endodermal intermediate cell

responsive to developmental signals. Additionally, global profiling of H3K9me3 heterochromatin, which represent stable epigenetic gene silencing, revealed that the abundant compact heterochromatin marks at the germ layer stage were reduced upon lineage commitment at protein coding genes [60]. This observation implied the global acquirement of transcription potency of epigenetically repressed genes during differentiation process. Collectively, in order to understand the regulation of developmental potency, a critical aspect is to determine whether the gene and cell competence is gradually restricted or actively acquired.

Importantly, the active acquirement of gene and cell potency has to occur at the very beginning of embryonic development, when highly specialized zygotic genome undergoes extensive epigenetic changes to be reset into pluripotent status [46]. Embryonic stem cells, consequent from this rebooting event, possess remarkable capacity to open occluded genes to recover gene potency [44]. Notably, pluripotency transits through a series of states, including naïve, formative and primed stages, before the onset of lineage differentiation [64][65]. However, the biological significance of the pluripotency transition remains elusive. Inspired by the decreasing of gene potency along with reduction of cell potency, we speculate that an important aspect of this transition is to shut off the de-occlusion ability prior to the onset of lineage differentiation. Such that occlusion of gene potency becomes irreversible in every possible downstream cell fates. This hypothesis sides with the gradual cell potency restriction theory and is in sharp contrast to the model of developmental competence acquirement in lineage intermediates. Direct test of this speculation would benefit our understanding of potency regulation during development process.

3.2 Results

3.2.1 Naïve but not primed pluripotent stem cells possess de-occlusion ability

In order to test whether de-occlusion ability was shutoff during the pluripotency state transition process, we compared the reprogramming capacity of naïve pluripotent E14 cells and primed pluripotent G14 cells when fused with somatic SEF2 cells. First, we collected all differential expressed genes for all fusion pairs and performed hierarchical clustering according to the strain-specific gene expression profile before and after cell fusion **(Figure 3.1A)**. Consistent with the profile of informative silent genes **(Figure 2.4A)**, when SEF2 was fused to C2C12, IMG or NSC, the transcription pattern of the same genome prior to and post fusion showed higher similarity compared to that between fusion partners within the same hybrid cells. This observation suggested that the cell identities was largely reserved for each lineage after cell fusion. The terminally differentiated cells and multipotent stem cells not only lacked the potency to switch transcriptome to a disparate lineage, but also incapable to reprogram the fusion partner to their own gene expression pattern. E14 cells, in contrast, showed remarkable reprogramming ability. The expression profiles of both E14 and SEF2 after cell fusion clustered together with unfused E14 cells, and are distant from unfused SEF2 cells. Intriguingly, G14 cells, which possessed comparable gene potency to E14 cells, changed their own gene expression to SEF2 profile post cell fusion rather than reprogramming the fusion partner. G14 cells thus represent a unique status when the genome was highly potent while the reprogramming capacity was impaired.

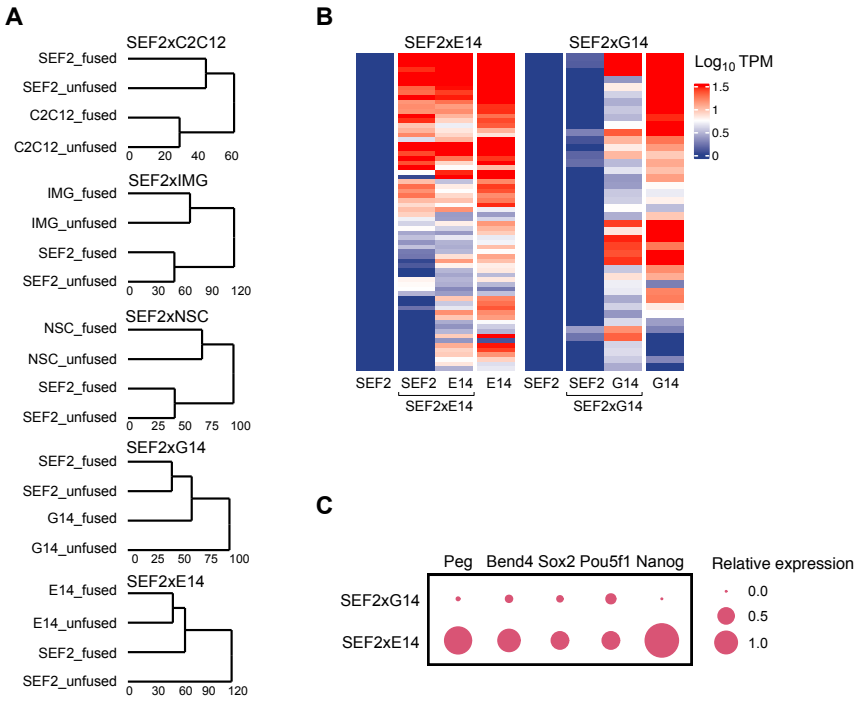

Figure 3.1 Loss of De-occlusion ability at epiblast stage.
(**A**) Hierarchical clustering of unfused and fusion samples according to strain-specific expression of differential expressed genes between fusion partners (TPM ≥ 2, log2FC > 2; SEF2xC2C12, n = 1308; SEF2xIMG, n = 3164; SEF2xNSC, n = 2704; SEF2xG14, n = 3055; SEF2xE14, n = 3434). (**B**) Heatmap of strain-specific expression of SEF2 occluded genes in SEF2xG14 and SEF2xE14 fuison cells. The SEF2 occluded gene list combines potency-annotations from SEF2xC2C12, SEF2xIMG and SEF2xNSC fusion cells. Non-extinguished genes within the combined SEF2 occluded gene list are displayed for each fusion sample (SEF2xG14, n = 42; SEF2xE14, n = 68). (**C**) Relative expression of representing SEF2 occluded genes and pluripotency genes in SEF2xG14 and SEF2xE14 fusion cells. Peg3, Bend4 and Sox2 are identified as SEF2 occluded genes in SEF2xC2C12, SEF2xIMG and SEF2xNSC fusions respectively. Expression levels from the SEF2 genome are scaled to the fusion partners.

In order to directly compare the de-occlusion ability of E14 and G14 cells, we combined the SEF2 occluded genes and activatable genes annotated in SEF2xC2C12, SEF2xIMG and SEF2xNSC fusions. Notably, while the combined activatable genes were readily activated upon cell fusion, the majority of SEF2 occluded genes remained silent in all three fusion pairs **(Figure 3.2)**. The exceptions could be explained by mis-annotation of occluded genes, which has been

discussed in Chapter 2. Importantly, while the occluded genes were activated when SEF2 was fused with E14 cells, they were still repressed in SEF2xG14 fusion **(Figure 3.1B)**. This observation suggested that the de-occlusion capacity was indeed impaired in primed pluripotent stem cells. We checked expression of representative occluded genes that were annotated in SEF2xC2C12, SEF2xIMG and SEF2xNSC fusions. As predicted, Peg3, Bend4 and Sox2 were activated to comparable level to E14 post cell fusion but kept silent in SEF2xG14 fusion **(Figure 3.1C)**. We noticed that Sox2 not only represents a NSC specific gene that was occluded in SEF2, but also stands for pluripotency genes. During the developmental process, the function of pluripotency factors is restricted to specific cell populations in the early embryo. Therefore, we inferred that other pluripotency genes in addition to Sox2, although the potency of which cannot be measured in somatic-somatic fusions, should also be occluded. Indeed, while Pou5f1 and Nanog were activated in SEF2 genome after fusion with E14 cells, their expression was failed to be fully triggered within SEF2xG14 fusion cells **(Figure 3.1C)**.

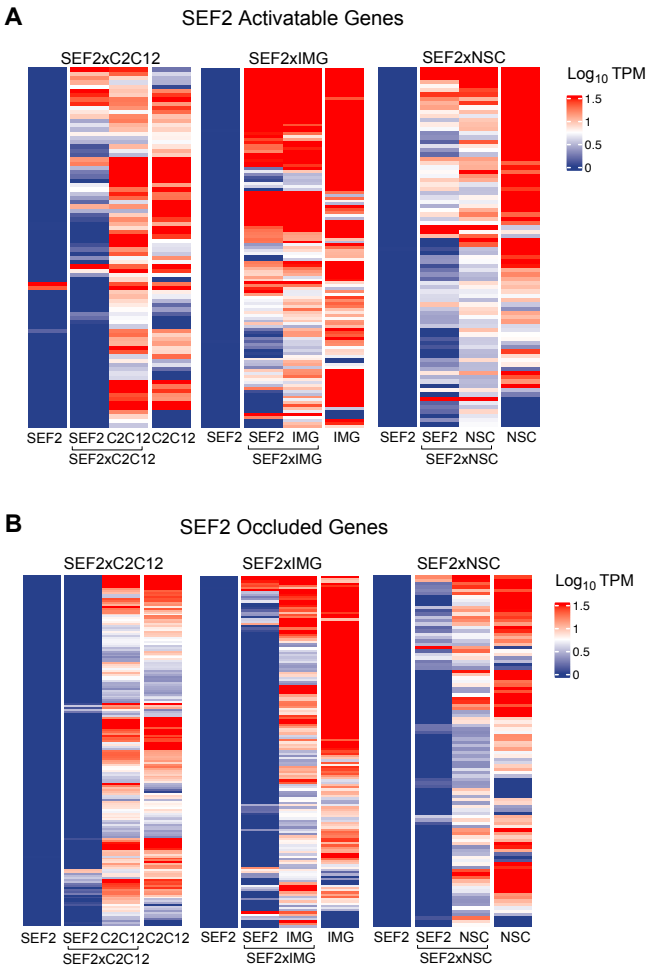

Figure 3.2 Expression of SEF2 activatable and occluded genes within fusion cells.
Heatmap depicting strain-specific expression of SEF2 activatable (**A**) and occluded (**B**) genes in unfused and fusion cells. The SEF2 activatable and occluded gene list combines annotations from SEF2xC2C12, SEF2xIMG and SEF2xNSC fusion cells. Non-extinguished genes within the combined gene list are displayed for each fusion sample (SEF2 activatable genes in SEF2xC2C12, n = 84; SEF2 activatable genes in SEF2xIMG, n = 123; SEF2 activatable genes in SEF2xNSC, n = 84; SEF2 occluded genes in SEF2xC2C12, n = 180; SEF2 occluded genes in SEF2xIMG, n = 157; SEF2 occluded genes in SEF2xNSC, n = 104).

The gene regulatory networks of metazoan are organized as hierarchical architecture, where the top master regulators are more influential than downstream executors [66]. Considering the critical role of gene occlusion in cell identity maintenance, we deduced that occluded genes identified by cell fusion assay should be enriched in master regulators in charge of lineage fates.

In order to test this hypothesis, we conducted gene ontology (GO) analysis for occluded and activatable genes annotated in each fusion pair. Remarkably, SEF2 occluded genes were enriched for key transcription factors and developmental signals of fusion partner's cell fate. In contrast, activatable genes were mostly executors responsible for tissue specific functions of the fusion partner **(Figure 3.1D, Figure 3.3)**.

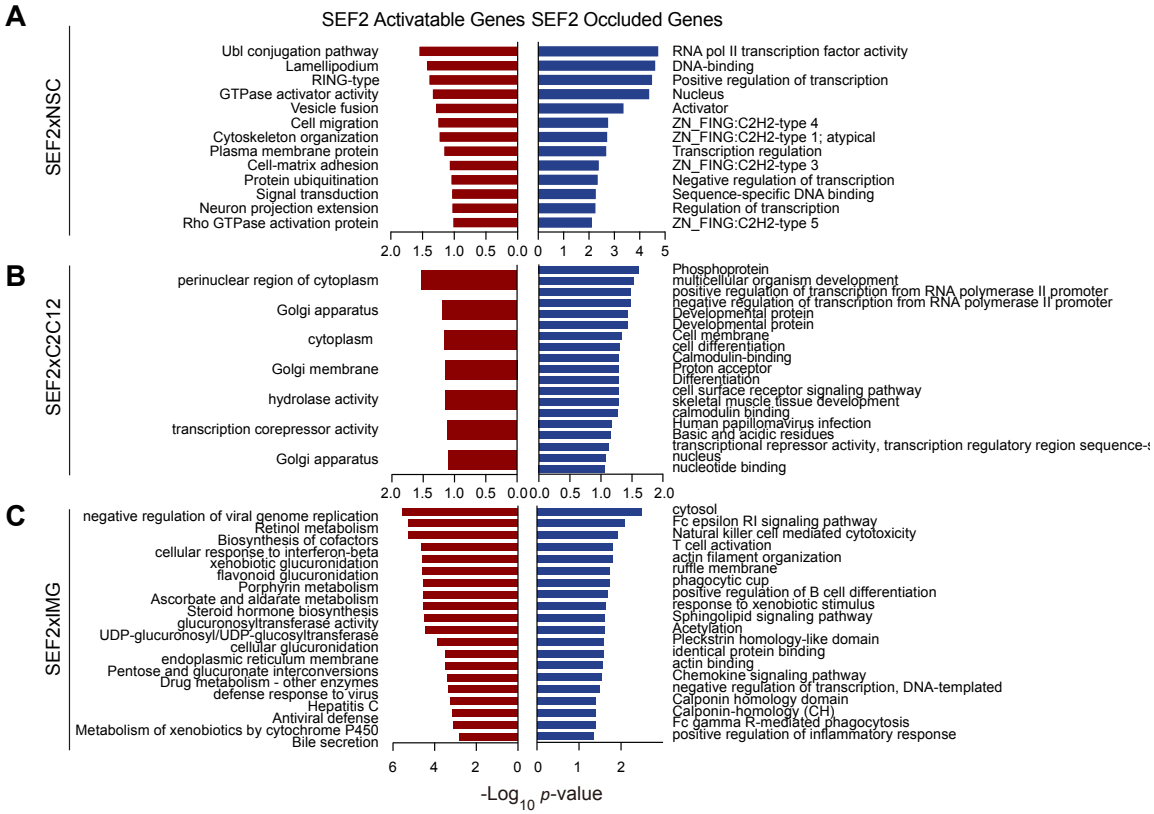

Figure 3.3 SEF2 activatable and occluded genes are enriched for downstream executors and upstream regulators respectively.
GO terms enriched in SEF2 activatable and occluded genes identified in SEF2xNSC (**A**), SEF2xC2C12 (**B**) and SEF2xIMG (**C**) fusion.

3.2.2 Gene occlusion is irreversible during cell differentiation process

Deduced from the impairment of de-occlusion capacity in primed pluripotent stem cells, we speculated that gene occlusion was irreversible after the onset of lineage differentiation.

Consequently, the developmental potency should be gradually and irreversibly restricted in across all cell fates. In order to directly test this hypothesis, we sought to differentiate SEF2xG14 cells into neural stem cells. As opposed to gain-of-potency model of lineage intermediates, according to which NSC genes would acquire potency in both G14 and SEF2 genome post differentiation, irreversible gene occlusion theory predicts that NSC genes would be triggered exclusively in G14 but not in SEF2 genome. As a control, when neural differentiation was induced in SEF2xE14 cells, NSC genes would be turned on in both genomes because the potency of SEF2 occluded genes has been recovered. In order to achieve more homogenous differentiation, two single clones were selected for both SEF2xE14 and SEF2xG14 fusions. The hybrid cells were cultured adherently in N2B27 media with SB431542, a TGF-β pathway inhibitor **(Figure 3.4A)**, to induce neural fates [67][68]. Subsequently, the cells were trypsinized into single cells and cultured in suspension with NSC media supplemented with FGF2 and EGF. Neural stem cells would form easy-to-purify neurospheres [69].

SEF2xE14 cells formed neurospheres efficiently following differentiation **(Figure 3.4B)**. More cell death was observed during differentiation of SEF2xG14 cells. Nonetheless, typical neurospheres were grown from survival cells **(Figure 3.4C)**. In order to check the fidelity of neural differentiation, we checked the expression of pluripotency genes and neural stem cell marker genes before and after differentiation **(Figure 3.4D)**. Downregulation of pluripotency genes and upregulation of NSC genes upon differentiation was observed in both SEF2xE14 and SEF2xG14 cells, suggesting the exiting from pluripotency and specification of neural stem cell fate. Intriguingly, while NSCs genes were activated in both E14 and SEF2 genome following differentiation in SEF2xE14 fusion cells, they were triggered exclusively in G14 genome in

SEF2xG14 fusion cells. Therefore, the tested NSC genes did not acquire potency to be activated during the lineage differentiation process.

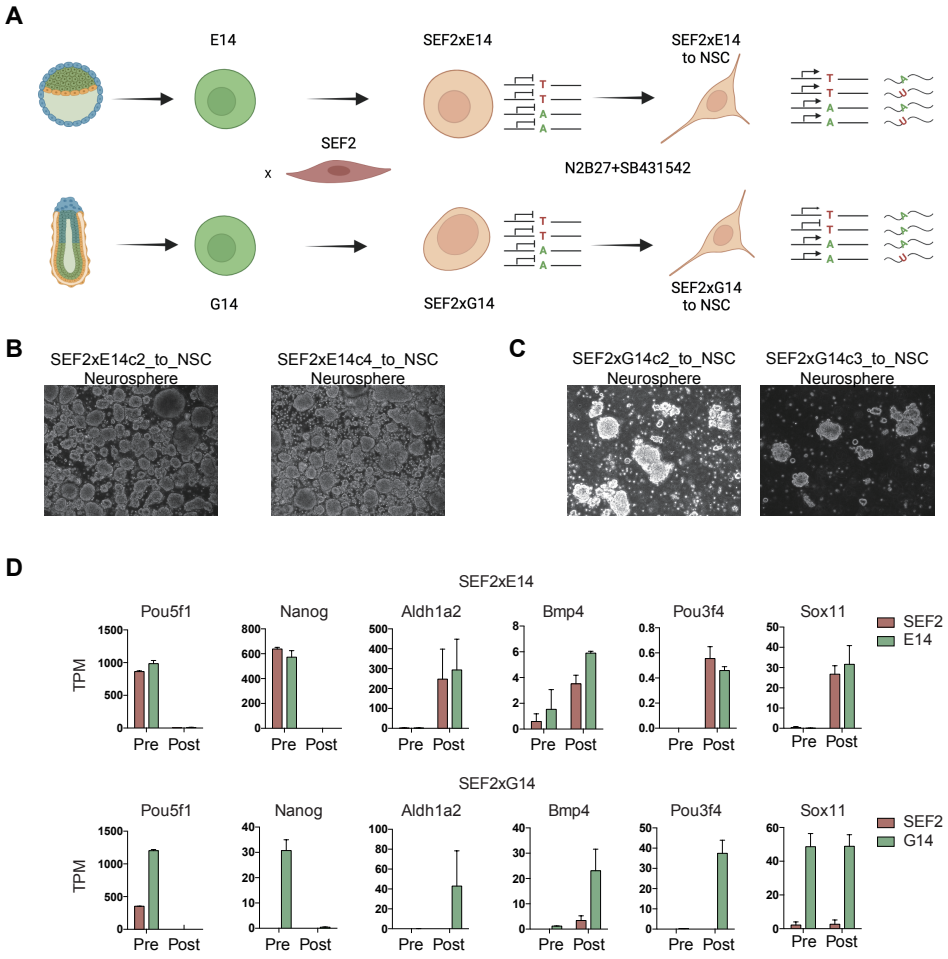

Figure 3.4 NSCLCs differentiated from SEF2xE14 and SEF2xG14 fusion clones display distinct expression of lineage key genes from the SEF2 strain.
(**A**) Schematic of directed differentiation assay of SEF2xE14 and SEF2xG14 fusion cells into neural stem cells. (**B**) Neurospheres formed by NSCLCs differentiated from SEF2xE14 fusion clone 2 and SEF2xE14 fusion clone 4. (**C**) Neurospheres formed by NSCLCs differentiated from SEF2xG14 clone2 and SEF2xG14 clone3. (**D**) Strain-specific expression of pluripotency genes and neural stem cell genes before and after differentiation in SEF2xE14 and SEF2xG14 fusion cells (mean ± SEM of two fusion clones).

We next collected all NSC specific genes that were occluded in SEF2, and checked their strain specific expression profile before and after differentiation in SEF2xE14 and SEF2xG14 cells **(Figure 3.5A)**. In contrast to SEF2xE14 cells where NSC genes were activated to comparable level in both genomes within the hybrid cells following differentiation, SEF2xG14 cells showed impaired upregulation of NSC genes in SEF2 genome. This observation supported the irreversible gene occlusion model, and argued against the acquirement of potency in lineage intermediates. In order to obtain a global view, we check all 889 genes that were upregulated during the differentiation of hybrid cells upon differentiation, irrespective of whether their potency was annotated **(Figure 3.5B)**. Remarkably, while the genes were almost equally activated in both genomes in SEF2xE14 fusion cells, the majority of them were only slightly elevated in SEF2 genome following differentiation of SEF2xG14 cells. Essentially, albeit experienced the same process in the same nucleus, only G14 genome completed the neural differentiation process while SEF2 genome largely kept the fibroblast identity. This result suggested pervasive occlusion of NSC genes in SEF2, which was not reverted during lineage differentiation.

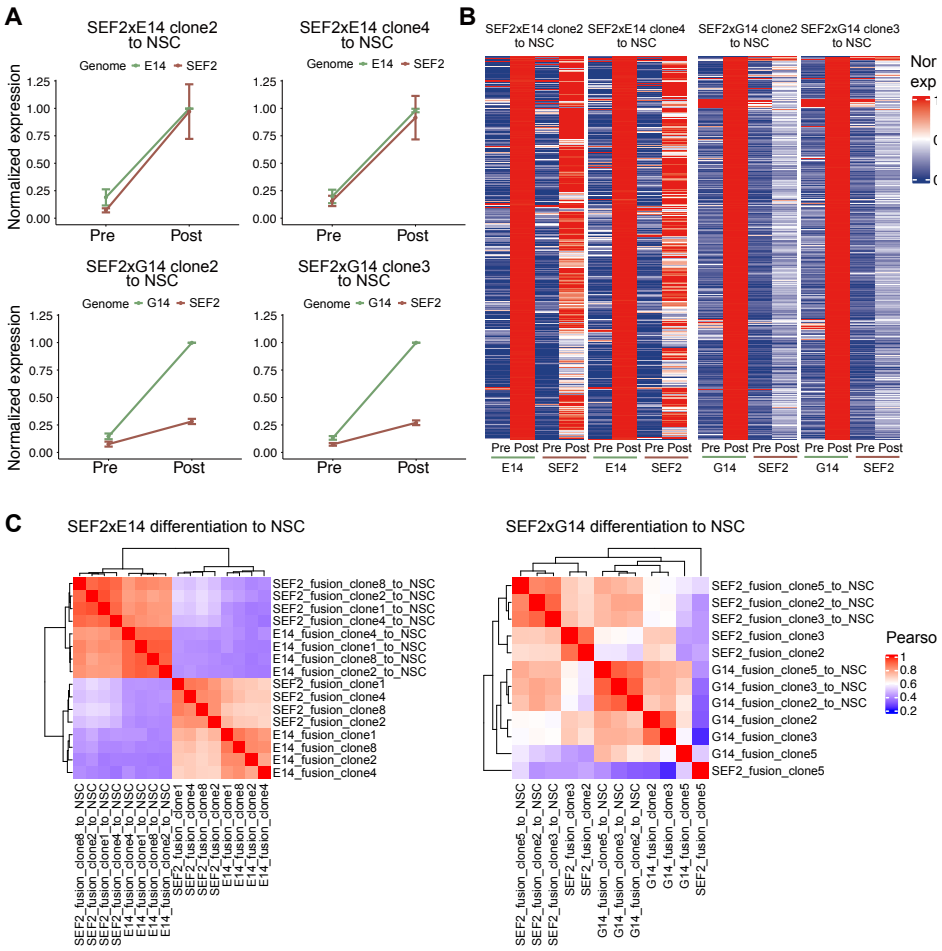

Figure 3.5 SEF2 occluded genes are activated in G14 but not SEF2 genome in SEF2xG14 fusion cells following differentiation.

(**A**) Strain-specific expression of upregulated SEF2 occluded genes in SEF2xE14 and SEF2xG14 fusion clones before and after differentiation. The figure displays genes within SEF2 occluded gene list that are silent prior to differentiation and become activated after differentiation in SEF2xE14 fusion (n = 63) or SEF2xG14 (n = 41) fusion clones. Expression levels are scaled to the post-differentiation expression levels of E14 or G14 genes (mean ± SEM of scaled expression). (**B**) Heatmap of all genes (TPM ≥ 2, log2_FC > 2, n = 889) that are upregulated in both SEF2xE14 and SEF2xG14 fusion clones following differentiation. Strain-specific expression levels are scaled for each gene of each fusion clone. (**C**) Heatmap of pairwise Pearson coefficient of strain-specific gene expression in SEF2xE14 (left, n = 250) and SEF2xG14 (right, n = 269) fusion cells prior to and post differentiation.

In order to address the heterogeneity of differentiation procedure, we differentiated two additional SEF2xE14 and SEF2xG14 fusion clones into neural stem cells. Notably, SEF2xG14 clone 5 was purposely selected because its expression profile clustered separately with all other

SEF2xG14 fusion clones **(Figure 3.5C)**. Intriguingly, less cell death was observed in this fusion clone, leading to growth of more neurospheres **(Figure 3.6A)**. Unexpectedly, G14 chromosomes were lost during differentiation of another SEF2xG14 clone, which was therefore excluded from downstream analysis **(Figure 2.1C)**. Consistent with the other SEF2xE14 clones, the newly selected clones elevated NSC genes to a comparable level in both SEF2 and E14 genomes **(Figure 3.6B)**. While in SEF2xG14 clone5, the activation of NSC genes in SEF2 was impaired **(Figure 3.6C)**. Due to the distinct gene expression pattern, SEF2xG14 clone5 upregulated a different group of NSC specific genes, in which 37.5% (15 out of 50) overlapped with the other SEF2xG14 clones **(Figure 3.6D)**. Nonetheless, the irreversibility of gene occlusion was consistently revealed. Additionally, the activation of the majority of NSC genes was impaired in SEF2 genome post differentiation of SEF2xG14 clone 5, similar to the observations in the other two SEF2xG14 clones **(Figure 3.6E)**.

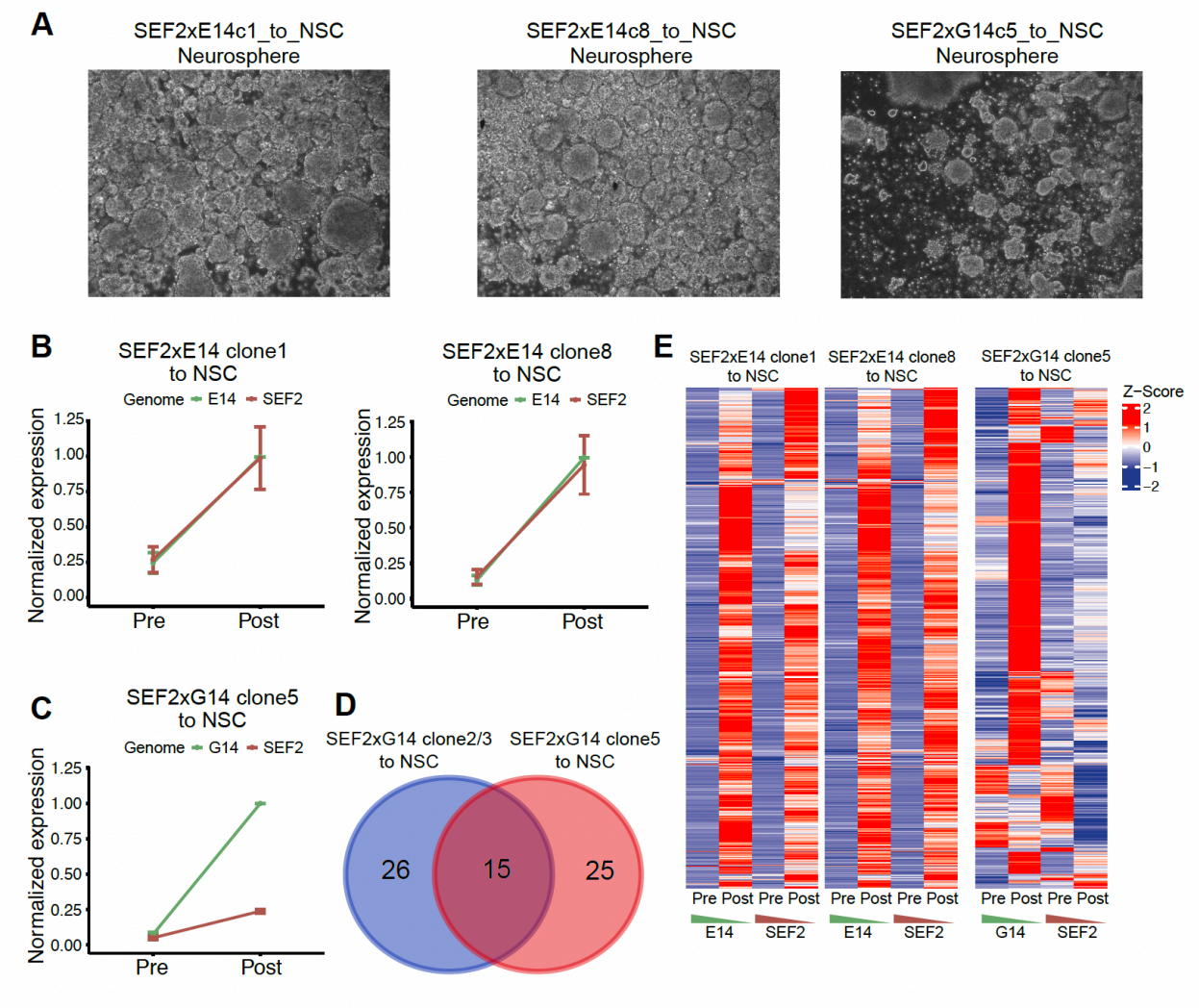

Figure 3.6 Consistent activation in G14 but not SEF2 genome are observed in different SEF2xG14 fusion clones following differentiation.

(**A**) Neurospheres fromed by NSCLCs differentiated from SEF2xE14 clone1, SEF2xE14 clone8 and SEF2xG14 clone5. (**B**) Strain-specific expression of upregulated SEF2 occluded genes in SEF2xE14 fusion clone 1 and clone 8 prior to and post differentiation (n = 63). Expression levels are scaled to the post-differentiation expression levels of E14 genes (mean ± SEM of scaled expression). (**C**) Strain-specific expression of upregulated SEF2 occluded in SEF2xG14 fusion clone 5 before and after differentiation (n = 40, mean ± SEM of scaled expression). (**D**) Venn diagram of SEF2 occluded genes upregulated in SEF2xG14 clone 5 or clone 2&3 following neuronal differentiation. (**E**) Heatmap of all genes (TPM ≥ 2, log2_FC > 2, n = 889) that are upregulated in both SEF2xE14 clones and SEF2xG14 clone 5 following differentiation. Expression levels are scaled for each gene of each clone individually.

3.3 Conclusion and discussion

Evaluation of SEF2 gene expression following fusion with E14 and G14 cells revealed remarkable difference in reprogramming capacity of naïve and primed pluripotent stem cells. Albeit the comparable developmental potential and gene potency, G14 cells failed to activate occluded genes like E14 cells. Indeed, the de-occlusion ability was specific to embryonic stem cells and were absent in all other tested cell types. The shutoff of de-occlusion ability before the onset of lineage differentiation renders gene occlusion irreversible in all possible cell fates. Importantly, the occluded pluripotency genes in somatic cells, which were responsible for the reprogramming capacity of embryonic stem cells, can only be activated by E14 but not G14 cells. This observation suggested that the impairment of de-occlusion capacity is likely to be irreversible itself. The irreversible gene occlusion model leads to the gradual cell potency restriction view of the developmental process. However, numerous studies proposed that developmental competence was actively acquired in lineage intermediates during differentiation, which involved reshaping of epigenetic profile by pioneer transcription factors to endow gene potency. This statement was in sharp contrast to our observation that the lineage intermediates like NSCs failed to endow SEF2 occluded genes with transcription potency following cell fusion, although the pioneer factors were present in the hybrid cells. Indeed, the transcription factors crucial for NSC lineage, such as Sox1, Sox2, Zic1, Zic2 and Pou3f1, were all occluded in SEF2 genome. An important evidence supporting the cell potency acquirement theory was the transdifferentiation experiments, where characteristics of an alternative lineage could be induced by forced expression of master regulators of that cell fate. This could be explained by the activation of a cohort of activatable genes by the regulators, which enables the phenotypical and functional similarities to another cell identity. Indeed, GO analysis suggested that downstream executor genes were enriched in activatable genes.

While the master regulators in charge of the executors were mostly occluded. When the activatable executor genes were activated by ectopic expression of master regulators, the cells are likely to exhibit characteristics mediated by executors of another lineage. Nonetheless, the occlusion of the regulators that determines cell fate was unlikely to be reverted.

Aiming to directly test whether gene occlusion is reversible during the lineage differentiation process, we differentiated SEF2xG14 cells into neural stem cells. Strikingly, while the fusion cells exhibited typical NSC behavior post differentiation, only G14 genome switched the gene expression profile to NSC-like status. While the activation of NSC genes were impaired in SEF2 genome, albeit in the same nucleus. This observation suggested that gene potency cannot be acquired during the differentiation process, and supported the irreversible model of gene occlusion and cell potency restriction.

3.4 Methods

3.4.1 In-vitro differentiation of SEF2xE14 and SEF2xG14 cells into neural stem cells

SEF2xE14 and SEF2xG14 cells were cultured in ESC and EpiSC conditions respectively. One day before differentiation, the cells were passaged to make them 60%-70% confluent the next day. The cells were washed with PBS to completely eliminate molecules supporting growth of ESCs and EpiSCs. Subsequently, fusion cells were trypsinized and neutralized by trypsin neutralization solution. After centrifugation, cell pellets were resuspended by PBS and counted by hemacytometer. Around one million cells were plated on to 10cm dish that was coated with CELLstart™ substrate, and cultured in N2B27 media containing DMEM/F12, Neurobasal Medium, N2, B27, GlutaMax, β-mercaptoethanol and 2uM SB431542 to induce neural fate. SEF2xE14 and SEF2xG14 cells lost pluripotency morphology and became similar to NSCs at day 5 and day 3 respectively. Then the cells were lift over by TrypLE Express Enzyme and neutralized

by trypsin neutralization solution. The resuspended cells were subsequently culture in low attachment plates with NSC media containing FGF2 and EGF. Successfully differentiated cells would form neurospheres, which was purified by gentle centrifugation. The purified neurospheres were trypsinized and cultured as monolayers in CELLstart™ coated plates with NSC medium.

3.4.2 Metagene analysis of SEF2 occluded genes in fusion cells during differentiation

Strain specific gene expression was calculated for both SEF2xE14 and SEF2xG14 fusion cells before and after differentiation. Total TPM, calculated from all reads from both strains of the hybrid cells, were used to select for NSC activated genes (Fold change > 4, total TPM post differentiation >= 4). For each gene, expression from E14 or G14 genome post differentiation was scaled to unit. The expression from E14 or G14 genome prior to differentiation, as well as expression from SEF2 genome before and after differentiation was scaled to post-differentiation E14/G14 expression. The scaled expression of SEF2 occluded genes that were activated following differentiation was plotted together for both stains within the hybrid cells, prior to and post differentiation.

Chapter 4: Default occlusion of unmodified chromatin in somatic cells

4.1 Introduction

Although the occlusion model has been developed to describe gene potency, the mechanism underlying gene occlusion remains elusive. Two hypotheses can be considered: 1) Genes are actively occluded during differentiation process. 2) Occlusion is the default state of the chromatin, while active regulators are necessary to maintain gene potency. The inheritable nature of repressive epigenetic modifications seems to suggest their potentials to actively maintain occlusion, yet previous efforts failed to identify a universal player that diminish gene potency into occlusion status. In the comprehensive analysis of occluded genes in mouse tail fibroblast, characterized by cell fusion with 12 different rat cell lines, no epigenetic modifications displayed a specific profile on the occluded genes [41]. Although occluded genes harbored a slightly higher enrichment of repressive histone modifications (e.g., H3K27me3 and H3K9me3), the subtle differences cannot account for the stark disparity between activatable and occluded genes. Specifically, DNA methylation was the most significant distinguisher between activatable and occluded genes. However, treatment of DNA methyltransferase inhibitor decitabine only activated 11% of the occluded genes. The inability of epigenetic modifications to hallmark occluded genes suggests our second hypothesis: occlusion is the default status of chromatin.

It is widely acknowledged that nucleosomal packaging poses a strong barrier to transcription. When a plasmid is optimally packaged with histones, transcription could not be initiated in the HeLa cell extract [70]. While this barrier can be overcome by preincubation with certain transcription factors before chromatin assembly [71], such a scenario is not observed in

physiological conditions. The mechanisms by which chromatin achieves sufficient accessibility in living cells have been extensively studied. Starting from quiescent chromatin, pioneer factors could initiate chromatin openness due to their capability to bind nucleosomes [72][73]. This allows recruitment of other transcription factors, activators and nucleosome remodelers that can ultimately establish and maintain the accessibility of open chromatin. However, the gene potency in this context has not been addressed. It is worth investigating whether chromatin possesses the potency to be activated when it lacks any modifications or factor bindings other than histone packaging.

4.2 Results

4.2.1 Chromatinized genes without epigenetic modifications were occluded by default in somatic cells

Despite numerous efforts to observe transcription on pre-assembled chromatin in *vitro*, it has never been investigated whether chromatin without factor binding and modifications processes transcriptional potency in living cells. To answer this question, we transfected pre-assembled chromatin into HEK293T cells to measure their transcription activity. We selected four types of promoters, including the Hsp68 promoter, universal strong promoter EF1a, tissue specific promoter of Nanog, which is supposed to be expressed in embryonic inner cell mass, and weak promoter UBC. The corresponding plasmids of the four promoters followed by downstream fluorescence reporters were assembled into chromatin with either HeLa core histones (Chr) or recombinant core histones (rChr). While the former histones were purified from HeLa cell extract that may contain a mixture of all types of modifications in living cells, the latter ones were purified from prokaryotic systems and consequently lacks any eukaryotic post-translational modification. Using recombinant core histones avoided the caveat that silent markers in the HeLa core histones

may prevent activation of the reporter. The chromatin was assembled with the assistance of histone chaperone NAP1 and chromatin remodeler ACF, the latter of which helps generate regularly spaced nucleosomes [74]. To purify the well-assembled chromatin from unchromatinized and partially chromatinized DNA, the system was precipitated by 5 mM $MgCl_2$ and then resuspended in electroporation buffer containing 1 mM EDTA **(Figure 4.1A)**. The quality of the assembled chromatin has been validated by the periodic ladder of DNA upon MNase partial digestion **(Figure 4.1B)**. We chose eletrophoration as the transfection strategy, as lipofectamine would cause dissociation of histones from DNA [75][76]. We optimized the electroporation conditions and achieved successful transfection of chromatinized DNA into living cells. The fluorescent level would indicate the transcription level of corresponding promoters. Intriguingly, the chromatinized reporters with Hsp68 and EF1a promoters showed significantly lower levels of fluorescence compared to naked DNA control **(Figure 4.1C)**. This suggested that well-assembled chromatin was unable to be activated in living cells.

Nevertheless, we recognize the caveat that chromatinized DNA could have lower efficiency of transfection and nuclear entry compared to naked DNA, causing false negative measurement of transcription activity. To avoid this caveat, we conducted more stringent measurement by nuclear purification and qPCR measurement. We removed the untransfected DNA and potential leaked DNA from dead cells by MNase digestion on intact cells, and washed them thoroughly by PBS to remove any residuals. To avoid discrepancy resulted from differential nuclear entry efficiency, we also purified the nuclei before purification of total DNA. Follow this procedure, we were able to quantify the amount of nuclear plasmid DNA in the transfected cells, and the corresponding value was used to normalize RNA expression level measured by RT-qPCR with necessary removal of DNA contamination. After normalization, the expression level of

chromatinized reporters were still negligible compared to naked DNA, and similar behaviors were observed in all four different types of promoters **(Figure 4.1D)**. The consistent silence of transcription suggested that chromatin without any regulators does not have the potency to be activated in HEK293T cells. Remarkably, the fluorescent reporters chromatinized by HeLa core histones remain barely expressed even after 7 days of transfection. This indicated that silent transcription of chromatinized plasmids was not resulted from limited time for activation **(Figure 4.1E)**. Overall, the silence of chromatinized plasmids implied that pre-assembled chromatin does not possess the potency of transcription, corresponding to the occluded status of genes. As genes without any factor binding or epigenetic modifications were already in the state of occlusion, occlusion should be the default status of chromatin in somatic cells.

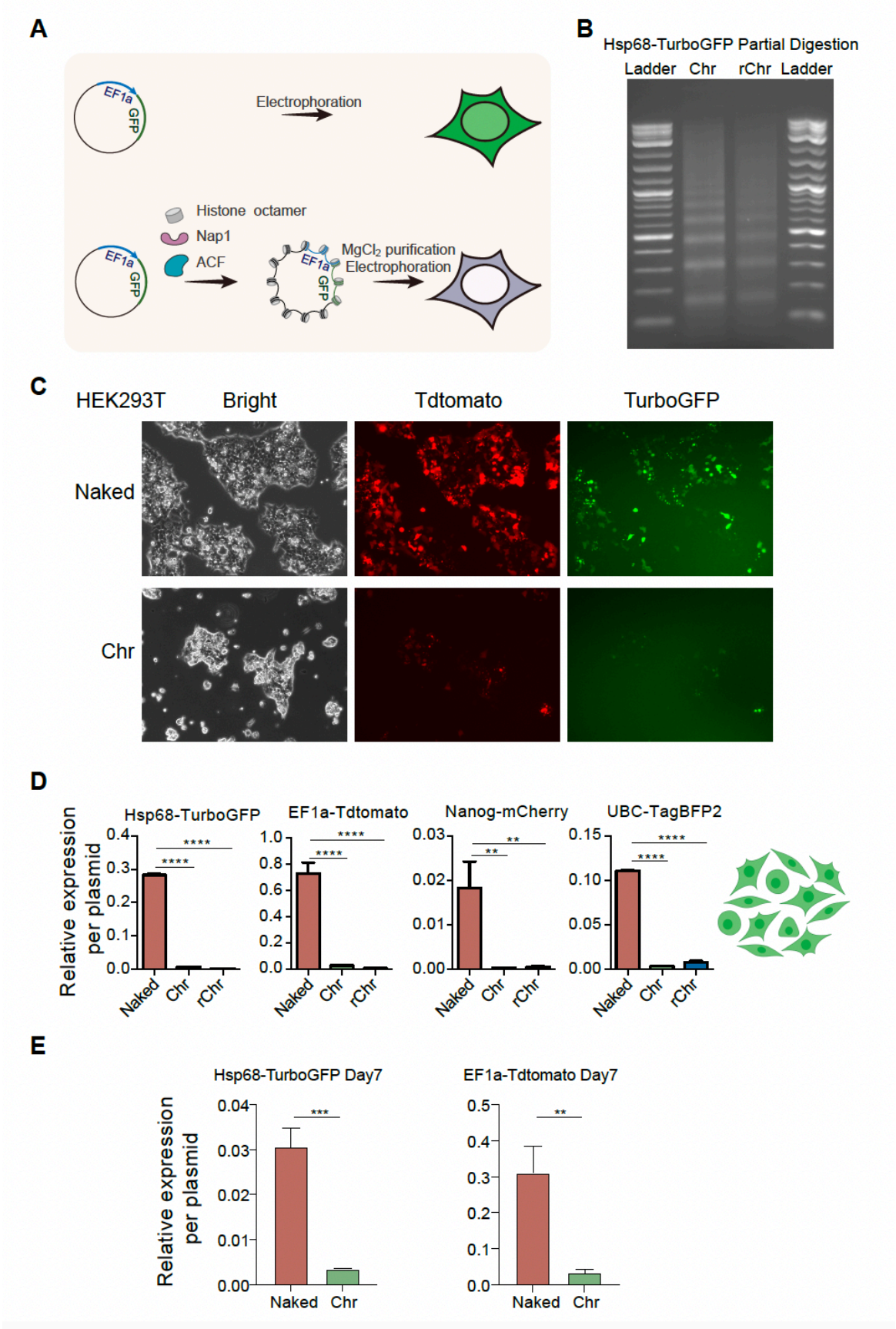
A
EF1a
GFP
Electrophoration
Histone octamer
Nap1
ACF
MgCl2 purification
Electrophoration
B
Hsp68-TurboGFP Partial Digestion
Ladder
Chr
rChr
Ladder
C
HEK293T
Bright
Tdtomato
TurboGFP
Naked
Chr
D
Hsp68-TurboGFP
EF1a-Tdtomato
Nanog-mCherry
UBC-TagBFP2
Relative expression per plasmid
Naked
Chr
rChr
E
Hsp68-TurboGFP Day7
EF1a-Tdtomato Day7
Relative expression per plasmid
Naked
Chr

Figure 4.1 Default occlusion of episomal chromatinized DNA
(**A**) Schematic representation of transcriptional potency test of naked DNA and in-vitro assembled chromatin in living cells. (**B**) Partial digestion analysis of plasmid DNA chromatinized with either HeLa core histones (Chr) or recombinant core histones (rChr) lacking eukaryotic post-translational modifications. (**C**) Fluorescent microscope image of HEK293T cells electorporated with naked or chromatinized plasmids containing EF1a-Tdtomato and Hsp68-TurboGFP (**D**) Comparative evaluation of naked and chromatinized gene expression driven by distinct promoters in HEK293T by qPCR. Abundance of RNA transcripts and nulear plasmid DNA are normalized to endogenous Actb RNA and DNA levels respectively. Relative expressions are calculated by the ratio of normalized RNA levels to normalized DNA levels (mean ± SD, n = 3, ordinary one-way ANOVA test). (**E**) Evaluation of naked and chromatinized gene expression during long-term culture. HEK293T cells were electroporated with naked or chromatinized genes and cultured for 7 days without passaging. Relative expressions are calculated by the ratio of normalized RNA levels to normalized DNA levels (mean ± SD, n = 3, ordinary one-way ANOVA test). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

4.2.2 Chromatinization mediated occlusion could be reprogramed in ESCs

In Chapter 3, we showed that naïve pluripotent stem cells had the capability of de-occlusion. Gene identified as occluded in SEF2 fusion with other cell types remain actively expressed at a similar level to the E14 copy after SEF2xE14 fusion. Such a de-occlusion effect should also exert on the chromatinized fluorescent reporters. To validate our hypothesis, we transfected chromatinized reporters into E14 cells. Reporters with Hsp68 or EF1a promoters showed comparable expression levels between chromanized and naked DNA group, suggesting an expected de-occlusion ability of E14 **(Figure 4.2A&B)**. The similar behaviors of chromatinized plasmids and SEF2 occluded genes also suggested that SEF2 occluded gene may undergo a similar mechanism of default occlusion: a nucleosome packaged chromatin without any modifications. Notably, the chromatinized Nanog and UBS promoters remains occluded in E14 **(Figure 4.2B)**. The failure of transcription may due to an essential role of replication in de-occlusion.

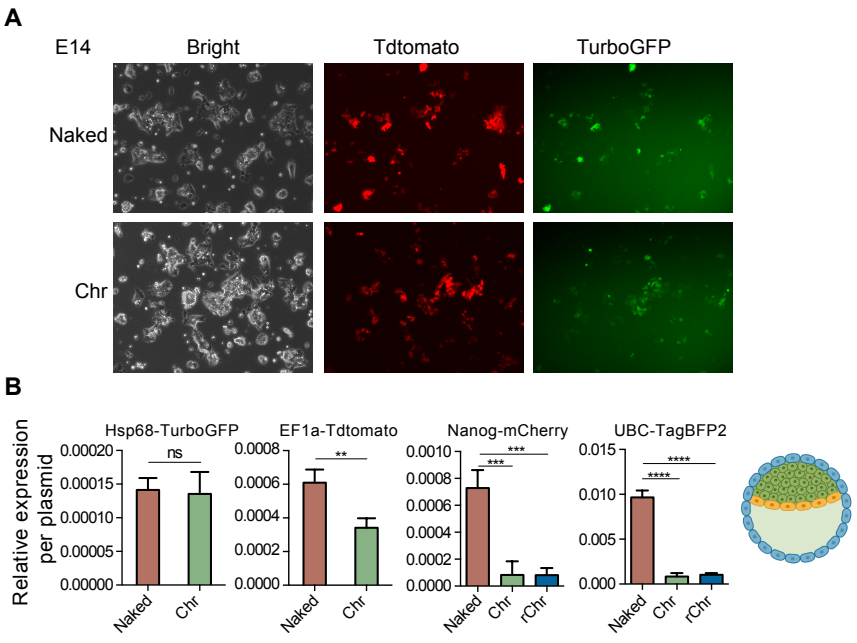

Figure 4.2 Chromatinized DNA are de-occluded in ESCs.
(**A**) Fluorescent microscope image of E14 cells electorporated with naked or chromatinized plasmids containing EF1a-Tdtomato and Hsp68-TurboGFP. (**B**) Comparative evaluation of naked and chromatinized gene expression driven by distinct promoters in ESCs by qPCR. Abundance of RNA transcripts and nulear plasmid DNA are normalized to endogenous Actb RNA and DNA levels respectively. Relative expressions are calculated by the ratio of normalized RNA levels to normalized DNA levels (mean ± SD, n = 3, ordinary one-way ANOVA test). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. ns, not significant.

4.2.3 Chromatinization mediated occlusion is stable in somatic cells during cell proliferation

Replication is a major obstacle for dividing cells to maintain the transcriptional state. In the context of occlusion, replication requires disassembly of the chromatinized genome, posing challenges to preserve the occluded status. Therefore, we investigated whether default occlusion of chromatin is inheritable after DNA replication. Testing inheritability requires integration of the transfected plasmids into the genome. Unfortunately, the chromatin assembly caused severe difficulties in genome integration. We employed multiple strategies, including random integration,

transposon-mediated integration by piggyBAC or sleeping beauty transposon system, and non-homologous end joining (NHEJ)-mediated integration. While the naked DNA integrated successfully into the genome, the efficiency of chromatinized DNA was insufficient for practical experimentation. To resolve this limitation, we reasoned that promoter is the major responder of occlusion. Chromatinization of the promoter region, but not the rest part of the plasmid, might still allow default occlusion of the reporter, while the genome integration cassettes could remain naked and function with a better efficiency. Based on this assumption, a half-chromatinization assay was developed **(Figure 4.5A)**. We designed two linear fragments with compatible sticky ends that could ligate into a circular DNA. This circular DNA contained expression cassettes of two fluorescent reporters, mCherry and GFP, located head-to-tail, as well as the transposon elements of piggyBAC (ITR). The two ligation loci were in the middle of the mCherry and GFP coding sequences. Before ligation, one fragment (the GFP fragment) contained the 3' half of mCherry reporter, together with the 5' half of the GFP reporter including its promoter. The other fragment (the mCherry fragment) contained the 3' half of the GFP coding sequence, a complete set of ITR, and the 5' half of the mCherry reporter including the promoter. Note that the two fluorescent reporters were both driven by EF1a promoters. With such a design, only properly ligated circular molecules could be successfully integrated into the genome. By chromatinizing the GFP fragment but not the mCherry fragment, the promoter of GFP reporter would be occluded, while the promoter of mCherry reporter remain naked, consequently able to transcribe. At the same time, ITR was in the mCherry fragment and remained naked, so the integration efficiency should not be impaired. With the half-chromatinization assay, we could measure the potency of chromatinized EF1a promoter by monitoring the GFP expression after rounds of DNA replication.

Notably, the distance between GFP and mCherry reporters are much closer than most genes in the genome. It is still unknown whether the default occlusion of chromatin persist when adjacent to an active gene. To test, we chromatinized either GFP fragment or mCherry fragment, then ligated with the other naked fragment and conducted electroporation to HEK293T without piggyBAC transposase **(Figure 4.3A)**. Partial digestion by MNase showed that the regular spacing of nucleosomes were impaired after ligation **(Figure 4.3B)**. However, flow cytometry still detected enrichment of the fluorescent signals from the naked fragments compared to the chromatinized one **(Figure 4.3C)**. qPCR measurement detected comparable amount of properly ligated GFP and mCherry DNA from the single-positive cell population, indicating that the enrichment of fluorescent signals was not resulted from failure of DNA ligation, but truly transcriptional differences **(Figure 4.3D)**. Upon normalization to the DNA levels, RT-qPCR showed that RNA expressed from the chromatinized promoters was significantly lower than the RNA from naked promoters **(Figure 4.3E)**. The result suggested that chromatinization still caused occlusion despite adjacent to an active gene.

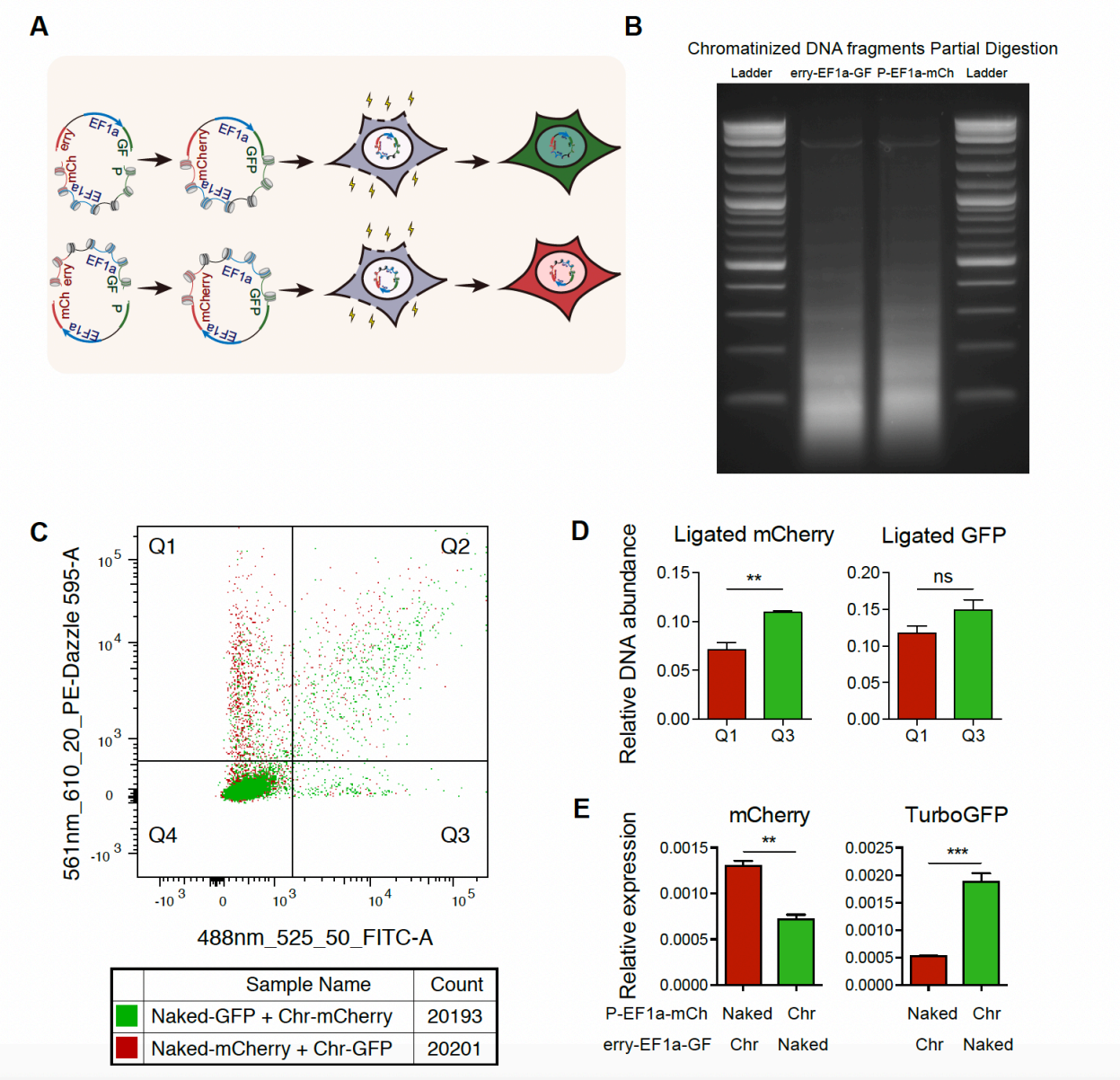

Figure 4.3 Half-chromatinized DNA are occluded at the chromatinized part when transfected under episomal condition.

(**A**) Schematic of half-chromatinization assay. Expression cassettes of GFP and mCherry are split in the middle of coding sequence. Fragments carrying GFP promoter (GFP fragments) or mCherry promoter (mCherry fragments) are chromatinized or kept naked. Chromatinied fragments and naked fragments are ligated together, reconstituting complete GFP and mCherry coding sequence, and then electroporated into HEK293T cells. (**B**) Partial digestion analysis of half-chromatinized DNA following ligation. (**C**) Flow cytometric analysis of 293T electroporated with half-chromatinized DNA. Green dots represent cells containing ligation product of chromatinized mCherry fragments and naked GFP fragments. Red dots represent cells containing ligation product of chromatinized GFP fragments and naked mCherry fragments. (**D**) The abundance of ligated complete GFP and mCherry DNA in red-only and green-only cells (mean ± SEM, n = 3, unpaired

t-test). Ligated DNA abundance are evaluated by qPCR and normalized Actb. Primers are designed across the ligation junctions. (**G**) Relative expression of mCherry and GFP from ligated half-chromatinized DNA. RNA levels are measured by RT-qPCR and normalized Actb (mean ± SEM, n = 3, unpaired t-test). Primers are designed across the ligation junctions. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. ns, not significant.

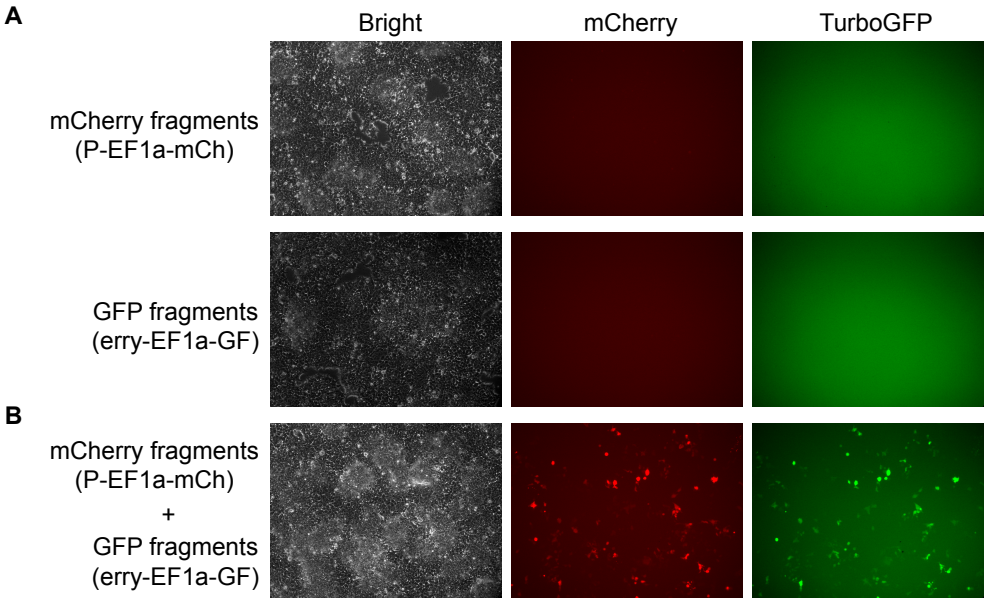

Figure 4.4 DNA fragments with matching sticky ends are ligated in HEK293T cells.
Fluorescent microscope of 293T electorporated with DNA fragments. Expression cassettes of GFP and mCherry are split in the middle of coding sequence, producing GFP fragments composed of 3' of mCherry CDS, EF1a promoter and 5' of GFP CDS, and mCherry fragments composed of 3' of GFP CDS, EF1a promoter and 5' of mCherry CDS. The two fragments are electroporated into 293T cells either individually (**A**) or simultaneously (**B**).

In the half chromatinization assay, we accidently found that the two fragments without ligation could be ligated in the HEK293T cells. When two naked fragments were co-electroporated, the fluorescence of GFP and mCherry would be spontaneously reconstituted **(Figure 4.4)**. This finding suggested an optimized strategy of the half chromainization assay. The chromatinized GFP fragment and naked mCherry fragment could be co-electroporated without ligation. Bypassing the ligation step avoided impairing nucleosomal spacing. Furthermore, the success of ligation in cell

could be selected by mCherry fluorescent signals in cells, as failed ligation generated neither intact coding sequence for expression, nor circular DNA for integration. Consequently, we chromatinized the GFP fragment with HeLa core histones or recombinant core histones, and validated the nucleosomal spacing by partial digestion assay **(Figure 4.5B)**. Then, we co-electroporated the chromatinized GFP fragment with naked mCherry fragment and piggyBAC transposase for genome integration in HEK293T cells. The co-electroporation of both naked GFP and mCherry fragment, as well as piggyBAC transposase, was conducted as a control **(Figure 4.5A)**. We sorted cells with red fluorescence, which indicated successful electroporation and fragment reconstitution. The enriched cells were cultured with normal proliferation for over a month, ensuring enough rounds of DNA replication before they were subjected to fluorescence measurement. Upon selection for red fluorescent signals, which indicated success of genome integration, the majority of cells with chromatinized GFP fragment remained negative of green fluorescence **(Figure 4.5 C-E)**. On the contrary, a significant percentage of cells with naked GFP fragment showed positive GFP signals. The differential expression was further validated by RT-qPCR with normalization on DNA amount **(Figure 4.5F)**. Notably, much less expression differences between chromatinized and naked samples was observed compared with whole-plasmid chromatinization. We speculated that genome integration still prefers fragments that were less optimally chromatinized, contributing to the smaller differences. Overall, the half chromatinization assay proved that default occlusion of chromatin is inheritable during DNA replication, not in need of marking by the epigenetic modifications.

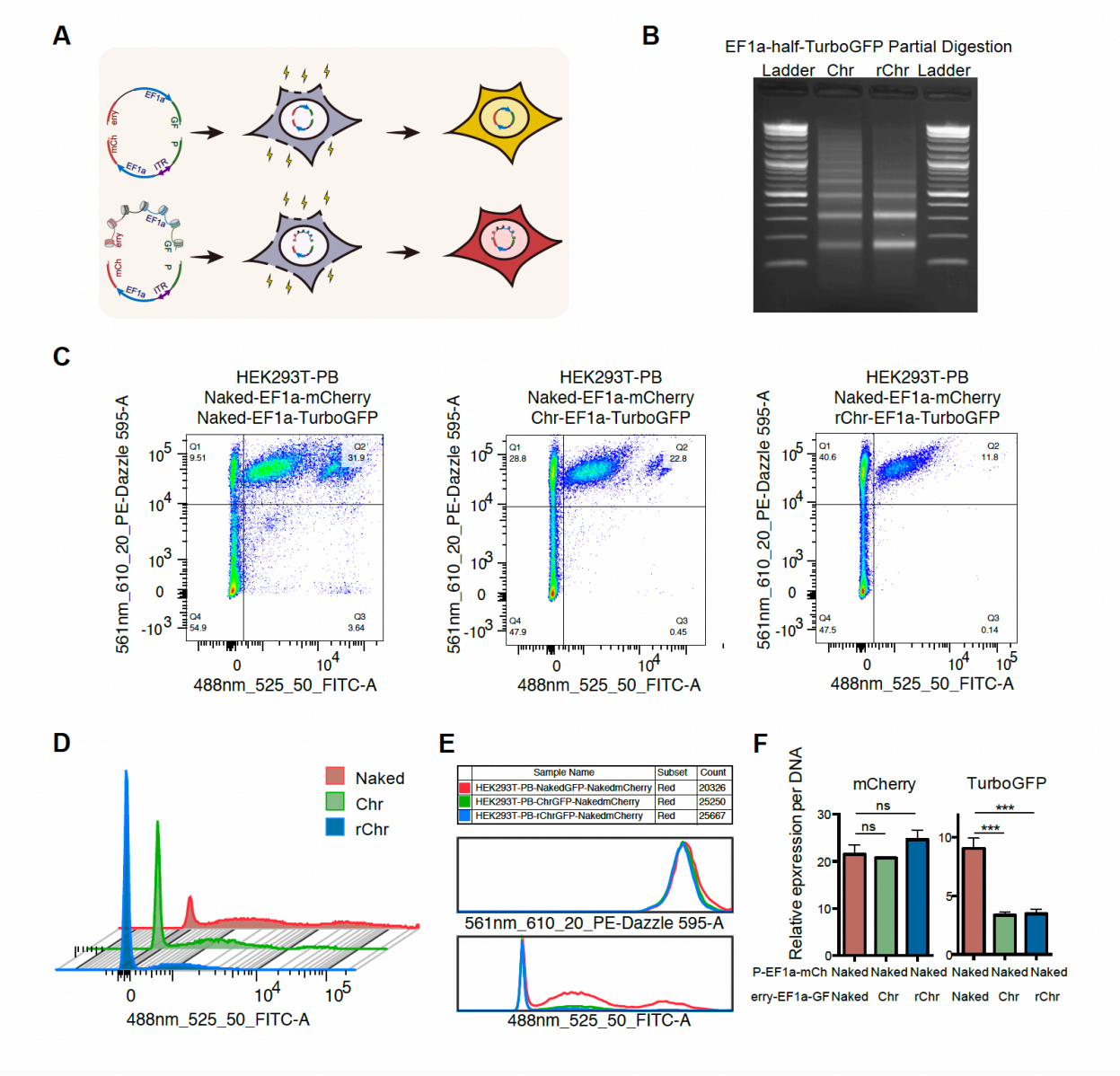

Figure 4.5 Integrated chromatinized DNA are stably occluded during DNA replication. (**A**) Schematic representation of half-chromatinization assay. Expression cassettes of GFP and mCherry are split in the middle of coding sequence. Fragments carrying GFP promoter (GFP fragments) are chromatinized and co-electroporated into HEK293T cells with fragments carrying mCherry promoter (mCherry fragments). Two DNA fragments are ligated within the cell, reconstituting complete coding sequence of GFP and mCherry simultaneously. The PiggyBac ITRs in the naked mCherry fragments enable the ligated, half-chromatinized DNA to be integrated into HEK293T genome. (**B**) Partial digestion analysis of chromatinized GFP fragments. (**C**) Flow cytometric analysis of HEK293T cells carrying mCherry fragments ligated with naked, Chr or rChr GFP fragments. The piggyBac ITRs in the mCherry fragments enable the ligated DNA to be stably integrated into HEK293T genome with reduced bias. The cells have been continuously passaged for over a month at the point of analysis. (**D**) Histogram of HEK293T cells carrying

(Figure 4.5 continued) mCherry fragments ligated with naked, Chr or rChr GFP fragments. The mCherry positive cells are selected and their GFP fluorescence intensity are displayed. (**E**) Statistics and histogram of red and green fluorescence intensity of samples from (C) with modal normalization of y axis. (**F**) Relative expression of mCherry and GFP from integrated ligated DNA. DNA and RNA abundance are normalized Actb levels. Relative expressions are calculated by the ratio of normalized RNA levels to normalized DNA levels (mean ± SEM, n = 3, ordinary one-way ANOVA test). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. ns, not significant.

4.3 Conclusion and discussion

Despite the extensive investigation of transcriptional regulation, how the gene potency is regulated was unexplored. We demonstrated that pre-assembled chromatin was occluded when transfected into cells, indicating that occlusion is the default status of chromatin. Specifically, half-chromatinization assay suggested that default occlusion of chromatin is resistant to DNA replication. Overall, our data argue against the hypothesis that occlusion is actively acquired. Instead, occlusion is the default state. Genes that are programed to be activated in downstream lineages requires additional regulators to maintain their potency.

Upon transfection, naked DNA would be assembled into chromatin, independent of its DNA replication [77][78]. The differing potency between naked and chromatinized DNA is not attributable to the absence of nucleosomal barriers on the plasmid molecules when it is transfected in the naked form. However, the timing of nucleosomal packaging is different. Chromatinized DNA is packaged with nucleosomes from the beginning. Conversely, naked DNA is likely to be occupied by activating factors and transcription machineries before nucleosomal packaging. The early establishment of active transcription prior to chromatin assembly would help sustain an active environment, thereby preserving potency of the gene. Consequently, we could observe the remarkable differences in transcription activity. Notably, there are still chances that nucleosomal packaging occurrs before the occupation of activating factors and transcription machineries on

naked DNA. In such cases, even naked DNA could become occluded upon transfection. Intriguingly, in the half chromatinization assay, mCherry positive cells co-transfected with naked GFP fragments contained a significant population with negative GFP signals (Figure 4.5A). This population corresponded to the instances where nucleosomal packaging occurred prior to transcriptional activation. Such population could exist in any transfection, but it could only be detected by the optimized half chromatinization procedure, as the two fragments could be ligated after they acquired different potency of transcription. This could also explain the clustered distribution of GFP signals in the half chromatinization assay (Figure 4.5A). When different copies of naked GFP fragments were transfected to the same cell, they could acquire different transcriptional potency, while their mCherry patterner might remained consistent. Due to the all-or-none nature of this potency discrepancy, the fluorescence profile within the population would exhibit clustering rather than a dispersed distribution.

It is important to note that default occlusion in the absence of epigenetic modifications does not imply that occluded genes are not subject to epigenetic modifications. The pre-assembled chromatin might acquire additional modifications and undergo epigenetic regulation upon transfected into cells. Repressive histone modifications have been detected in occluded genes [41], with levels comparable to silent genes that are activatable. This suggests a complex interplay between occlusion and repressive epigenetic modifications. Occluded genes could be modified by repressive marks for regulation other than gene potency, including chromatin folding [79] and phase separation [80]. Conversely, genomic regions lacking epigenetic modifications are not necessarily occluded. These genes may still be occupied by transcription factors for maintenance of gene potency. The intricate relationship underscores the importance of further investigation into functions of epigenetic modifications, with due consideration for gene potency.

The default occlusion mechanism suggests that genes properly packaged with nucleosome cannot be reactivated. It has been reported that pioneer factors can bind to quiescent chromatin due to their affinity to nucleosome-embedded DNA, enabling subsequent recruitment of activating factors to initiate transcription [72][73]. Our model of default occlusion seems contradictory to the consensus view of pioneer factors. However, it could be determined by gene potency whether the nucleosomal binding of pioneer factors could overcome the energy barriers of chromatin to induce transcription activation. The fact that pioneer factors can bind to or even disentangle nucleosomes does not necessarily mean that pioneer factors alone are sufficient for transcription activation. For example, binding sites of the paradigm pioneer factor Foxa2 are epigenetically primed in mESC [81]. Additional factors are required in the activation of silent genes, which were described as gene potency in our study. However, the molecular mechanism of gene potency still needs further study. Notably, the mechanism is not necessarily dependent on epigenetic modifications.

Our default occlusion model necessitates that activatable genes maintain their gene potency. The underlying mechanism is expected to be have several features. Firstly, it should be resistant to DNA replication. Nucleosome packaging of newly synthesized DNA can easily cause gene occlusion. Second, the functional molecules must be consistently present at the activatable genes throughout the differentiation lineage. As cells lack the de-occlusion capability since primed pluripotent stem cells, any intermediate cell type that misses the required molecules on an activatable gene will result in loss of its potency. Although it is a consensus that epigenetic modifications maintain the transcriptional status of genes, their capacity to maintain gene potency is unclear. Specifically, active epigenetic modifications are not self-sustaining, raising questions about their resistance to DNA replication. These modifications, even when present at activatable genes, may require else factors for their establishment, such as transcription factors or even

retained Pol II. While repressive epigenetic modifications H3K27me3 and H3K9me3 are self-replicative, their distributions do not distinguish between activatable and occluded genes [41]. Furthermore, the specific distributions of epigenetic modifications in different lineages require another layer of regulation. Overall, the molecular mechanism underlying gene potency may depend on additional factors. Our hypothesis will be explored in Chapter 5.

4.4 Methods

4.4.1 Cell culture and electroporation

HEK293T cells were maintained in DMEM with 10% FBS, supplemented with penicillin/streptomycin. E14 cells were cultured as described in Chapter 2, with LIF and 2i. The cells were passaged one day prior to electroporation to make them 70% confluent at the day of transfection. For electroporation, the cells were lift over by trypsinization and subsequently neutralized by trypsin neutralization solution. The cells were washed with PBS and counted by hemacytometer. One million cells were resuspended in Gene Pulser Electroporation Buffer containing 1mM EDTA and 1ug naked/chromatinized DNA. Electroporation was conducted with Gene Pulser II Electroporation System using recommended parameters. MNase was added to the medium one-day post electroporation to remove naked or chromatinized DNA molecules that did not enter the cells. Reporter expression was tracked by fluorescence microscope. To assay for promoter strength, the electroporated cells were separated in two batches, one of which was used to extract RNA for RT-qPCR and the other was used to isolate nuclear plasmid DNA for quantification and normalization.

4.4.2 Plasmid construction

Plasmids used in the chromatinization and half-chromatinization assay were constructed by VectorBuilder Inc. VB171220-1258gxn, VB200911-1183yfq, VB210716-1111vfu and

VB210719-1129pad were used for chromatinization assay. For half chromatinization assay, VB220428-1064cuh or VB221020-1032pzd for piggyBac mediated integration, and VB220428-1062bsm were digested by BbsI to create mCherry fragments and GFP fragments with compatible sticky ends, which was subsequently gel-purified for chromatin assembly assay.

4.4.3 Chromatin assembly

Chromatinization was performed by Chromatin Assembly Kit (Cat:53500, Active Motif). The HeLa core histones in the kit were replaced with equal amounts of recombinant histones from NEB (M2508S and M2509S) for rChr samples. Notably, high salt buffer and low salt buffer were mixed together during the incubation step of h-NAP-1 and core histone. This modification of the manual was suggested to improve the efficiency and reproducibility of chromatin assembly reaction. Following chromatin reconstitution, 5mM $MgCl_2$ was added to the solution and centrifuge for 15 minutes to precipitate assembled chromatin. The pellet was resuspended with Gene Pulser Electroporation buffer with 1mM EDTA. Unresolved chromatin pellet was removed by centrifugation before electroporation.

4.4.4 Partial digestion assay

Following purification of assembled chromatin, 3ul of 0.1M $CaCl_2$ was added to 100ul of resuspended chromatin. Subsequently, 1000 units of MNase was added to the solution and incubated at room temperature for 30 seconds. The digestion was stopped by addition of 34ul 4X Enzymatic Stop Solution in the Chromatin Assembly Kit. Following purification of digested DNA by QIAquick PCR Purification Kit (QIAGEN), gel electrophoresis was used to visualize the pattern of nucleosome array.

Chapter 5: Identification of place-holding factors maintaining gene potency in stem cells

5.1 Introduction

The default occlusion of chromatin suggests the existence of a class of molecules that sustains the transcriptional potency of activatable genes. We refer to these molecules as "placeholders". As discussed in Chapter 4, placeholders are expected to be consistently present at the activatable loci. It is worth noting that, for the same gene or cis-regulatory element, placeholders could consist of the same group of molecules across all types of cells. Alternatively, they could comprise different groups of molecules in different cell types, where they act in a relay on the same activatable genes. Interestingly, it has been reported that transcription factors could premark binding sites of other transcription factors in the downstream lineage [82]. ES transcription factors Esrrb, Nanog, Sox2 and Oct4 would bind to the macrophage-restricted enhancers without installation of active histone modifications. The premarking is essential for proper function of the enhancers when cells are differentiated into macrophages. This finding suggested the potential of transcription factor relay to sustain gene potency.

In neuronal lineages, multiple transcription factors play significant roles in the establishment of expression profiles. Sox2, famous as a Yamanaka factor, is highly expressed in stems cells from the neuronal lineages [83]. As a master regulator, Sox2 is essential for the neurogenesis function of NSCs [84][85]. Specifically, Sox2 deficiency impairs astrocyte maturation [86], yet it is barely expressed in quiescent astrocytes [87]. It is worth investigation whether Sox2 is included in maintenance of gene potency, aside from its function of lineage-specific gene activation. Another important transcription factor, Olig2, is also specifically

expressed in neuronal tissues and important for neuronal differentiation [88]. Intriguingly, it has been reported that NSC differentiation to astrocytes requires translocation of Olig2 to cytoplasm in the glial precursors [89][90]. The investigation of Olig2 in gene potency maintenance could broaden our understanding of potency regulation due to its unique molecular behaviors.

5.2 Results

5.2.1 Deletion of Sox2 in NSCs impaired potency of target activatable genes

We hypothesized that factors we termed "placeholders" bind to the activatable genes to sustain their gene potency. We assume such binding could maintain a more open environment compared to chromatinized occluded status. Indeed, analysis of published ATAC-seq data [91] showed that activable genes are more accessible than occluded genes in NSCs **(Figure 5.1A)**, suggesting the presence of "placeholders" on the activatable genes. As discussed above, placeholders should have binding preference towards lineage specific genes. Besides, its activity is necessary in stem cells but not terminally differentiated cells. Consequently, we hypothesized that transcription factors specifically expressed in stem cells could function as placeholders. To identify placeholders in NSCs, we investigated the expression changes of transcription factors during the differentiation from NSC to astrocytes, and picked up three candidates that were downregulated after differentiation **(Figure 5.1B)**. The two genes, Sox2 and Olig2, are well-known master regulators in NSC, while Hmgb2 is a transcription factor that has higher expression levels in bone marrow and thymus.

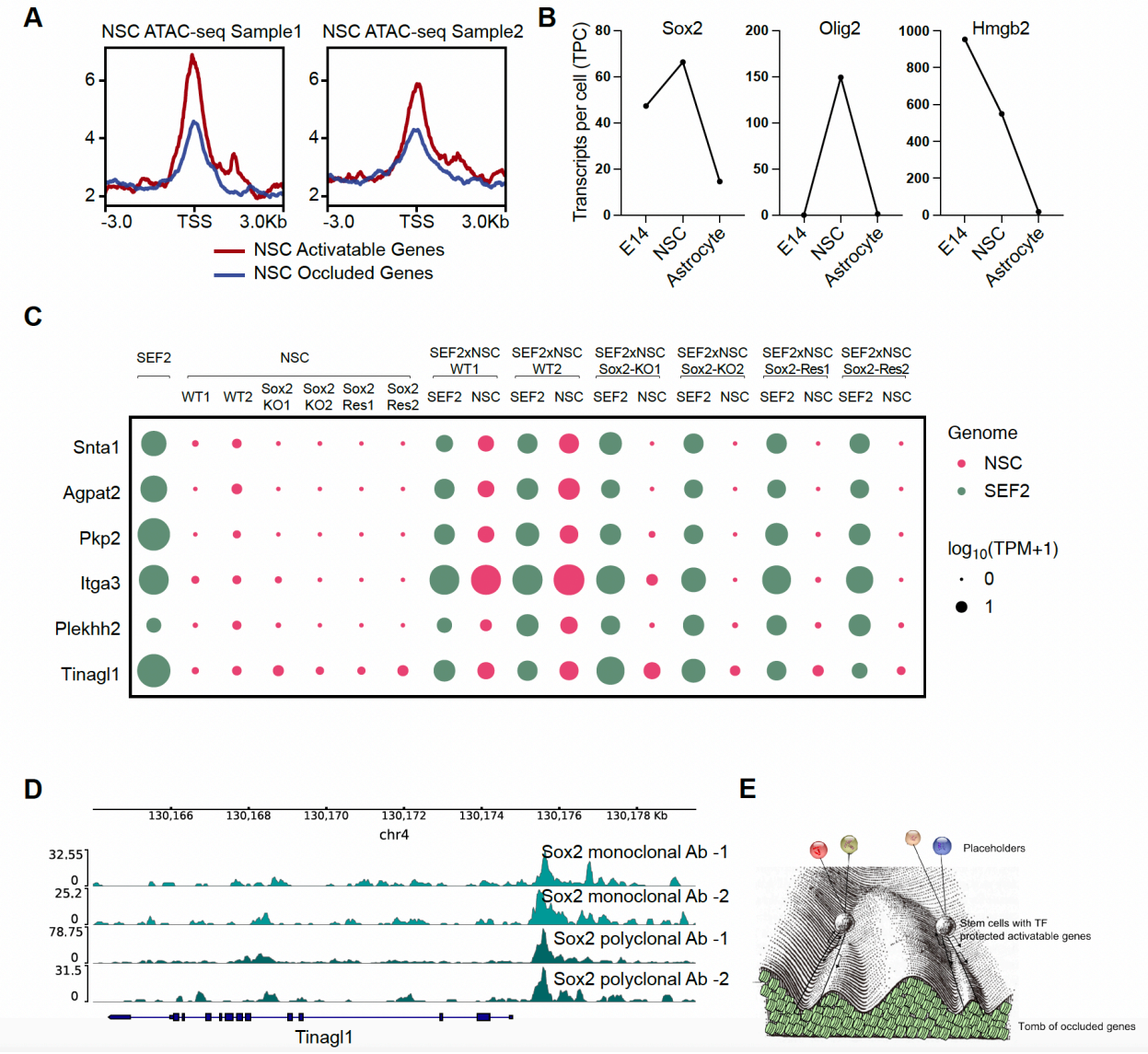

Figure 5.1 Sox2 binding prevents target genes from being occluded in NSCs.
(**A**) ATAC-seq profile of activatable (n = 85) and occluded (n = 186) genes in NSCs. (**B**) Expression of Sox2, Olig2 and Hmgb2 in E14, NSCs and Astrocytes differentiated from NSCs. (**C**) Bubble plot depicting strain-specific expression of unfused SEF2, NSC and fusion cells. Two NSC clones homozygous for Sox2 deletion are fused to SEF2 either directly (SEF2xNSC-Sox2-KO1 and SEF2xNSC-Sox2-KO2), or rescued for Sox2 expression with lentivirus vector before cell fusion (SEF2xNSC-Sox2-Res1 and SEF2xNSC-Sox2-Res2). Genes that are activatable in wildtype NSCs while occluded after Sox2 knockout are displayed. (**D**) Sox2 binding profile of a representing NSC activatable gene that becomes occluded following Sox2 deletion. Sox2 binding are measured by CUT&Tag with both monoclonal and polyclonal antibodies. Two replicates are shown for each antibody. (**E**) Schematic representation of placeholder-occlusion model in regulation of developmental process. Placeholder binding keeps lineage-specific genes activatable in stem cells. Following differentiation, placeholders disappear

(Figure 5.1 continued) in response to developmental signal, rendering lineage-inappropriate gene occluded by chromatinization.

We first knockout the whole coding sequence of Sox2 in NSC utilizing CRISPR/Cas9, and fused them with SEF2 to measure the potency of NSC genes. Remarkably, 6 activatable genes were occluded upon Sox2 knockout in both NSC clones **(Figure 5.1C)**, indicating that Sox2 was necessary for their transcriptional potency. In order to test whether Sox2 directly bind to the genes, we conducted CUT&Tag using both monoclonal and polyclonal antibodies against Sox2. Remarkably, all newly occluded genes except Agpat2 exhibited reliable Sox2 binding signals **(Figure 5.1D, Figure 5.2)**. This result supported the placeholder function of Sox2.

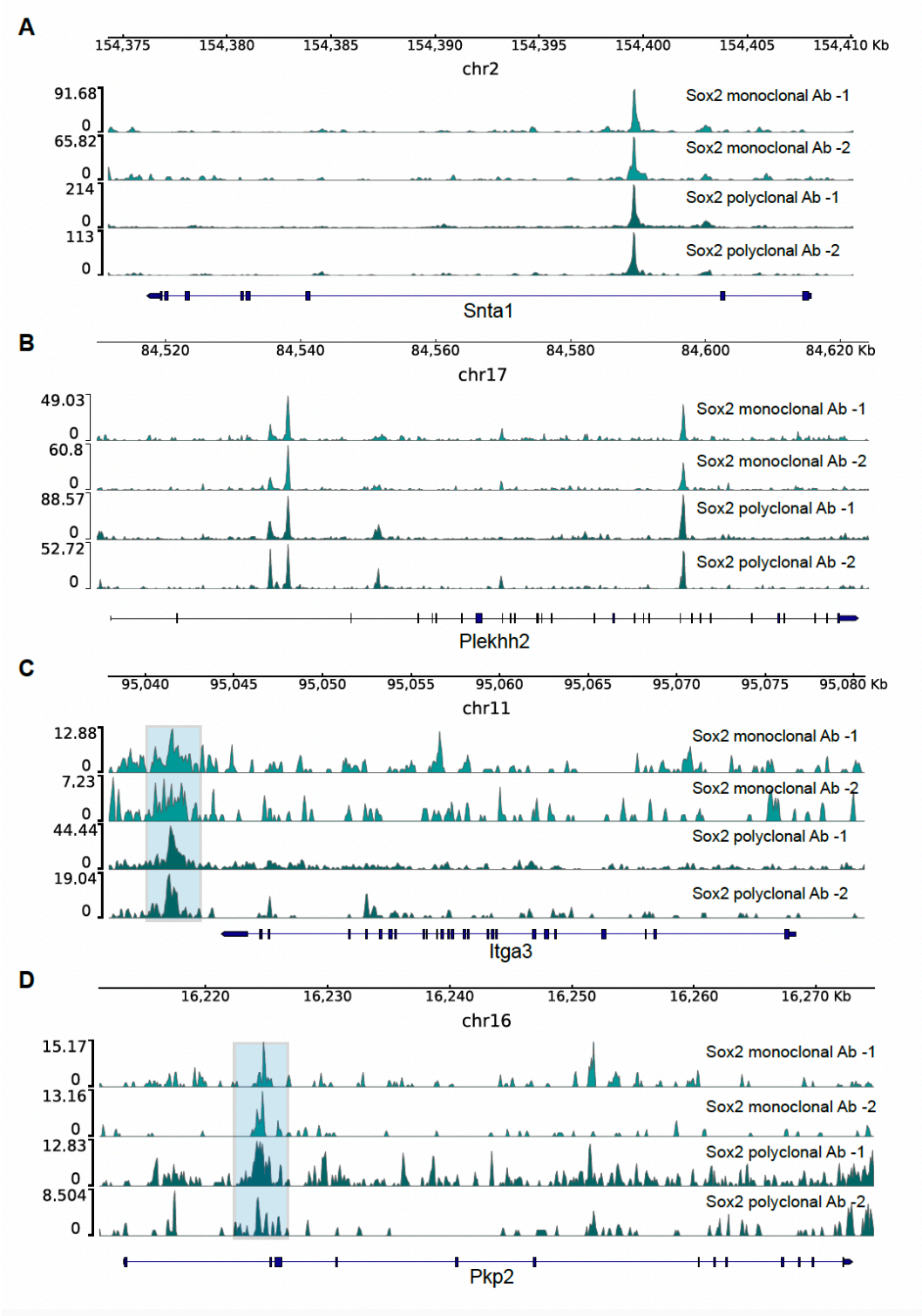

Figure 5.2 Genes that become occluded after Sox2 deletion are bound by Sox2.
Sox2 binding profile of NSC activatable genes that become occluded following Sox2 deletion. Sox2 binding are measured by CUT&Tag with both monoclonal and polyclonal antibodies. Two replicates are shown for each antibody.

5.2.2 Deletion of Sox2 or Olig2 impaired potency of target expressed genes

We next focused on expressed genes that definitely possess transcriptional potency because they are already actively expressed. Similar to activatable genes, we speculated that the potency of expressed genes can also be impaired following deletion of placeholders. Indeed, three genes that were silenced upon Sox2 deletion became occluded consistently in two NSC clones, as suggested by cell fusion with SEF2 **(Figure 5.3A)**. Intriguingly, although we did not find activatable-to-occluded switch in NSC after Olig2 knockout, 22 actively expressed genes became occluded **(Figure 5.3B)**. These results suggested that Sox2 and Olig2 served as not only transcriptional activators but also placeholders.

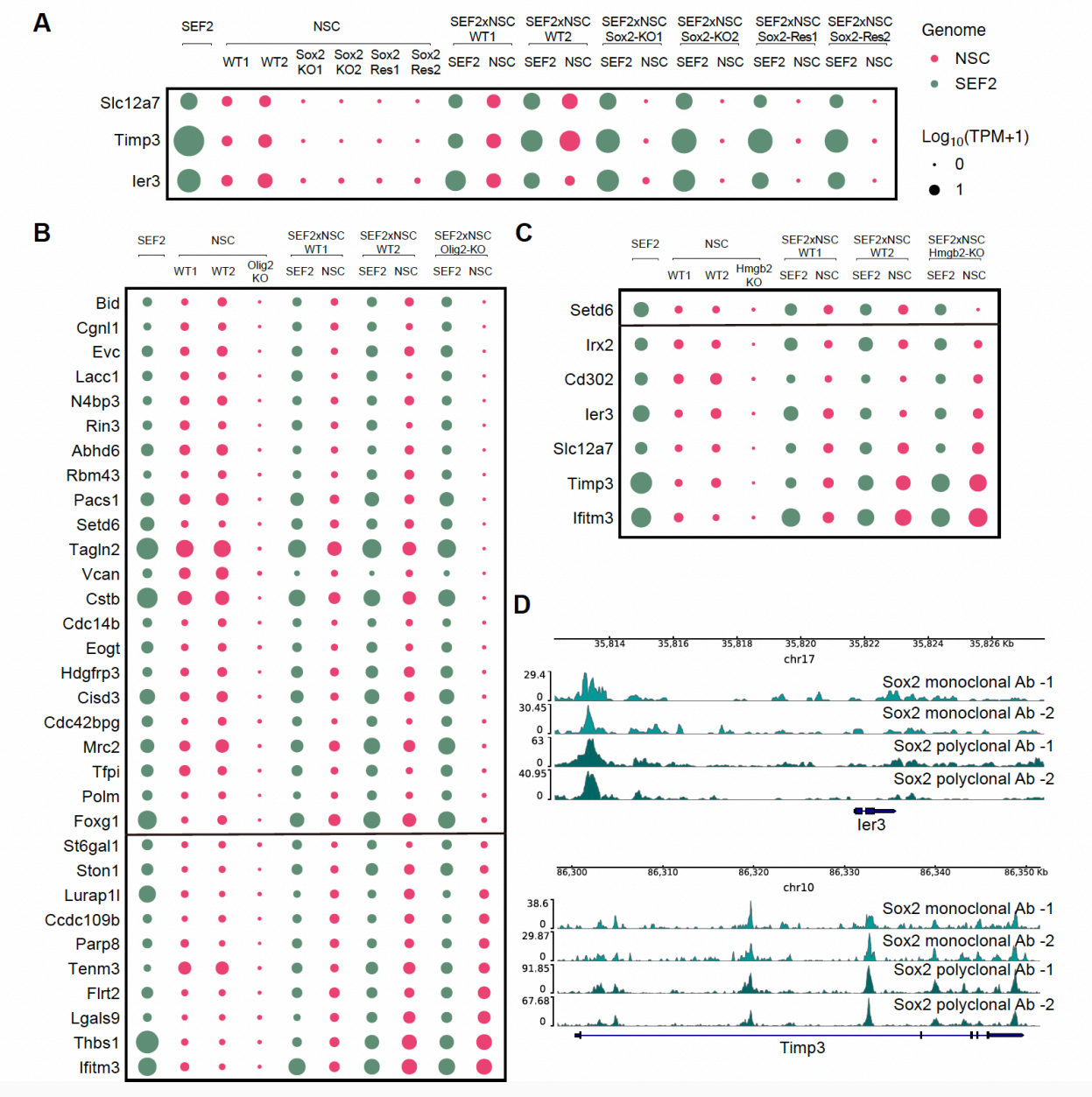

Figure 5.3 Expressed genes become occluded following Sox2 or Olig2 deletion.
(**A**) Bubble plot depicting strain-specific expression of unfused SEF2, NSC and fusion cells. Two NSC clones homozygous for Sox2 deletion are fused to SEF2 either directly (SEF2xNSC-Sox2-KO1 and SEF2xNSC-Sox2-KO2), or rescued for Sox2 expression with lentivirus vector before cell fusion (SEF2xNSC-Sox2-Res1 and SEF2xNSC-Sox2-Res2). Genes that are expressed in wildtype NSCs while silenced after Sox2 knockout are displayed. (**B**) Bubble plot depicting strain-specific expression of unfused SEF2, wildtype NSC, a NSC clone homozygous for Olig2 deletion and SEF2xNSC fusion cells. Genes that are expressed in wild type NSCs while silenced after Olig2 knockout are displayed. (**C**) Bubble plot depicting strain-specific expression of unfused SEF2, wildtype NSC, a NSC clone homozygous for Hmgb2 deletion and SEF2xNSC fusion cells. Genes that are expressed in wild type NSCs while silenced after Hmgb2

(Figure 5.3 continued) knockout are displayed. (**D**) Sox2 binding profile of representing NSC expressed genes that become occluded following Sox2 deletion. Sox2 binding are measured by CUT&Tag with both monoclonal and polyclonal antibodies. Two replicates are shown for each antibody.

Unlike Sox2 and Olig2, depletion of Hmgb2 did not induce potency loss of many genes. Only Setd6 became silenced and occluded following Hmgb2 knockout, which was observed in Olig2 knockout NSCs as well. The majority of genes that were repressed upon Hmgb2 deletion reserved transcriptional potency to be activated in fusion cells **(Figure 5.3C)**. Notably, the expression of Timp3 and Ier3 diminished in both Sox2 and Hmgb2 knockout NSCs. However, they became occluded only after Sox2 depletion. Consistently, Sox2 binding was detected around the two genes by CUT&Tag **(Figure 5.3D)**. The striking difference in gene potency change between Sox2 knockout and Hmgb2 knockout NSCs suggested that Hmgb2 might serve as transcriptional activators but not placeholders of target genes. Alternatively, redundant placeholders may work cooperatively to maintain the gene potency. Depletion of Sox2 but not Hmgb2 destroyed the cooperation.

5.2.3 Adding Sox2 back to NSC failed to rescue gene potency

According to the observations in Chapter 3, faithful gene occlusion during cell differentiation should be irreversible. In order to test the irreversibility of gene occlusion induced by Sox2 deletion, we added Sox2 back to the NSC knockout clones by lentivirus transduction, and checked the potency of target genes by cell fusion assay. As expected, reintroduction of Sox2 failed to rescue transcription potency of occluded genes, irrespective of whether they were switch from activatable or expressed genes **(Figure 5.1C, Figure 5.3A)**. Although Sox2 has been reported as a pioneer factor that is capable of binding to nucleosomal DNA and inducing chromatin openness [92][93], occluded genes became irresponsive to reintroduced Sox2. This result further

supported the irreversible gene occlusion model of developmental potency restriction during lineage differentiation.

5.3 Conclusion and discussion

The default occlusion of chromatinized DNA suggested that factors were required to maintain gene potency before differentiation. We therefore hypothesized that stem cell specific transcription factors might serve as placeholders preventing chromatinization and occlusion of target genes. Indeed, deletion of Sox2 and Olig2 impaired transcription potency of both activatable and expressed genes, which was not rescued by reintroduction of the factors. Based on the observations, we proposed a placeholder model of gene potency regulation during lineage differentiation process. In multipotent stem cells, binding of placeholders at target genes sustains their transcriptional potency by preventing chromatinization. Upon differentiation, stem cell specific placeholders are downregulated in response to developmental signals. Simultaneously, different groups transcriptional regulators are upregulated in various downstream lineages, taking over a subset of placeholder bound genes and leading to cell fate specification. Genes of alternative lineages lost protection from placeholder and were not taken over by new transcription factor. Consequently, they were spontaneously occluded by chromatinization, leading to irreversible restriction of developmental potential.

Although the placeholder model provided a simple and elegant explanation for gene potency regulation during cell differentiation, there are still critical questions need to be further explored. First, the function of placeholders is continuously challenged by DNA replication. It remains unclear how the gene potency was maintained in this process. It has been published that RNA pol II associates with expressed genes during DNA replication [34]. An appealing hypothesis is that placeholders are retained at activatable genes similar to RNA pol II, probably in a

cooperative manner. Alternatively, epigenetic modifications might be involved to facilitate stable maintenance of gene potency. Second, we proposed that transcription activators could take over placeholder's place upon differentiation. A critical question to ask is whether the relay between placeholders and activators has selectivity. In a simple scenario, placeholders maintain gene potency by sustain chromatin accessibility, and they do not possess selectivity for downstream transcriptional regulators themselves. In contrast, if interaction between placeholders and lineage specific transcription factors is required for the relay, the placeholder should have intrinsic selectivity for downstream TFs. Third, as discussed in Chapter 2, the minimal unit under regulation of placeholders and occlusion might be cis-regulatory elements rather than genes. It is crucial to identify which elements were required for potency maintenance of target genes.

5.4 Methods

5.4.1 CRISPR/Cas9 mediated gene knockout

Guide RNAs were designed at 5' and 3' UTRs of Sox2, Olig2 and Hmgb2 genes to delete the full coding sequences. Plasmids containing double sgRNAs, including VB220823-1355ufr, VB220823-1356hth and VB220823-1358fhs, were constructed by VectorBuilder™. Following transfection, the NSCs were selected by 30ug/mL blasticidin for 2 days, and then sorted into CELLstart™ coated 96-well plates as single cells. The single clones were genotyped by primers across the deletion site to screen for successful deletion, and primers within deletion fragment to screen for homogenous knockout. The knockout clones were further validated by RNA-seq results, wherein the coding sequences of targeted genes were depleted of reads.

5.4.2 Lentivirus transduction

The lentivirus containing Sox2 driven by PGK promoter (VB230911-1130rye) were produced by VectorBuilder™. The Sox2 knockout NSC clones were transduced at MOI of 5 and

selected by G418 for 7 days. The rescue of Sox2 expression can be viewed in RNA-seq results, in which the coverage of the coding sequence was recovered.

5.4.3 CUT&Tag

CUT&Tag was performed following the instruction of NovoNGS® CUT&Tag® 3.0 High-Sensitivity Kit. Both monoclonal (Cat: 23064, cell signaling) and polyclonal (Cat: ab97959, abcam) antibodies against Sox2 were used to detect the binding sites.

LIST OF REFERENCES

[1] Evans, M. J., & Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. nature, 292(5819), 154-156.

[2] Brons, I. G. M., Smithers, L. E., Trotter, M. W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S. M., ... & Vallier, L. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature, 448(7150), 191-195.

[3] Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., ... & McKay, R. D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature, 448(7150), 196-199.

[4] Huang, Y., Osorno, R., Tsakiridis, A., & Wilson, V. (2012). In vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation. Cell reports, 2(6), 1571-1578.

[5] Joo, J. Y., Choi, H. W., Kim, M. J., Zaehres, H., Tapia, N., Stehling, M., ... & Schöler, H. R. (2014). Establishment of a primed pluripotent epiblast stem cell in FGF4-based conditions. Scientific reports, 4(1), 7477.

[6] Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., & Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. science, 282(5391), 1145-1147.

[7] Gafni, O., Weinberger, L., Mansour, A. A., Manor, Y. S., Chomsky, E., Ben-Yosef, D., ... & Hanna, J. H. (2013). Derivation of novel human ground state naive pluripotent stem cells. Nature, 504(7479), 282-286.

[8] Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2014). Molecular Biology of the Cell. Garland Science, 1227.

[9] Wagner, D. E., Wang, I. E., & Reddien, P. W. (2011). Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. Science, 332(6031), 811-816.

[10] Yamanaka, S., & Blau, H. M. (2010). Nuclear reprogramming to a pluripotent state by three approaches. Nature, 465(7299), 704-712.

[11] Hou, P., Li, Y., Zhang, X., Liu, C., Guan, J., Li, H., ... & Deng, H. (2013). Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. Science, 341(6146), 651-654.

[12] Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. cell, 126(4), 663-676.

[13] Weintraub, H., Tapscott, S. J., Davis, R. L., Thayer, M. J., Adam, M. A., Lassar, A. B., & Miller, A. D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. Proceedings of the National Academy of Sciences, 86(14), 5434-5438.

[14] Tachibana, M., Takeda, K., Nobukuni, Y., Urabe, K., Long, J. E., Meyers, K. A., ... & Miki, T. (1996). Ectopic expression of MITF, a gene for Waardenburg syndrome type 2, converts fibroblasts to cells with melanocyte characteristics. Nature genetics, 14(1), 50-54.

[15] Xie, H., Ye, M., Feng, R., & Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. Cell, 117(5), 663-676.

[16] Vierbuchen, T., Ostermeier, A., Pang, Z. P., Kokubu, Y., Südhof, T. C., & Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. Nature, 463(7284), 1035-1041.

[17] Hotchkiss, R. D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. Journal of Biological Chemistry, 175(1), 315-332.

[18] Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & development, 16(1), 6-21.

[19] Smith, Z. D., & Meissner, A. (2013). DNA methylation: roles in mammalian development. Nature Reviews Genetics, 14(3), 204-220.

[20] Guibert, S., Forné, T., & Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome research, 22(4), 633-641.

[21] Stewart, K. R., Veselovska, L., & Kelsey, G. (2016). Establishment and functions of DNA methylation in the germline. Epigenomics, 8(10), 1399-1413.

[22] Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A., & Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature, 484(7394), 339-344.

[23] Zhang, Y., Xiang, Y., Yin, Q., Du, Z., Peng, X., Wang, Q., ... & Xie, W. (2018). Dynamic epigenomic landscapes during early lineage specification in mouse embryos. Nature genetics, 50(1), 96-105.

[24] Métivier, R., Gallais, R., Tiffoche, C., Le Péron, C., Jurkowska, R. Z., Carmouche, R. P., ... & Salbert, G. (2008). Cyclical DNA methylation of a transcriptionally active promoter. Nature, 452(7183), 45-50.

[25] Hellman, A., & Chess, A. (2007). Gene body-specific methylation on the active X chromosome. science, 315(5815), 1141-1143.

[26] Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., ... & Zhao, Y. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell, 146(6), 1016-1028.

[27] Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., ... & Lander, E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell, 125(2), 315-326.

[28] Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., ... & Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature, 448(7153), 553-560.

[29] Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., ... & Schübeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Molecular cell, 30(6), 755-766.

[30] Yu, J., Xiong, C., Zhuo, B., Wen, Z., Shen, J., Liu, C., ... & Li, G. (2020). Analysis of local chromatin states reveals gene transcription potential during mouse neural progenitor cell differentiation. Cell reports, 32(4), 107953.

[31] Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & development, 16(1), 6-21.

[32] Ming, X., Zhang, Z., Zou, Z., Lv, C., Dong, Q., He, Q., ... & Zhu, B. (2020). Kinetics and mechanisms of mitotic inheritance of DNA methylation and their roles in aging-associated methylome deterioration. Cell research, 30(11), 980-996.

[33] Reinberg, D., & Vales, L. D. (2018). Chromatin domains rich in inheritance. Science, 361(6397), 33-34.

[34] Escobar, T. M., Oksuz, O., Saldaña-Meyer, R., Descostes, N., Bonasio, R., & Reinberg, D. (2019). Active and repressed chromatin domains exhibit distinct nucleosome segregation during DNA replication. Cell, 179(4), 953-963.

[35] Stewart-Morgan, K. R., Reverón-Gómez, N., & Groth, A. (2019). Transcription restart establishes chromatin accessibility after DNA replication. Molecular cell, 75(2), 284-297.

[36] Bar-Ziv, R., Brodsky, S., Chapal, M., & Barkai, N. (2020). Transcription factor binding to replicated DNA. Cell Reports, 30(12), 3989-3995.

[37] Wang, A., Yue, F., Li, Y., Xie, R., Harper, T., Patel, N. A., ... & Sander, M. (2015). Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. Cell stem cell, 16(4), 386-399.

[38] Fenstermaker, T. K., Petruk, S., Kovermann, S. K., Brock, H. W., & Mazo, A. (2023). RNA polymerase II associates with active genes during DNA replication. Nature, 620(7973), 426-433.

[39] Davidson, R. L. (1974). Gene expression in somatic cell hybrids. Annual review of genetics, 8(1), 195-218.

[40] Blau, H. M. (1989). How fixed is the differentiated state?: Lessons from heterokaryons. Trends in Genetics, 5, 268-272.

[41] Lee, J. H., Bugarija, B., Millan, E. J., Walton, N. M., Gaetz, J., Fernandes, C. J., ... & Lahn, B. T. (2009). Systematic identification of cis-silenced genes by trans complementation. Human molecular genetics, 18(5), 835-846.

[42] Gaetz, J., Clift, K. L., Fernandes, C. J., Mao, F. F., Lee, J. H., Zhang, L., ... & Lahn, B. T. (2012). Evidence for a critical role of gene occlusion in cell fate restriction. Cell research, 22(5), 848-858.

[43] Looney, T. J., Zhang, L., Chen, C. H., Lee, J. H., Chari, S., Mao, F. F., ... & Lahn, B. T. (2014). Systematic mapping of occluded genes by cell fusion reveals prevalence and stability of cis-mediated silencing in somatic cells. Genome Research, 24(2), 267-280.

[44] Foshay, K. M., Looney, T. J., Chari, S., Mao, F. F., Lee, J. H., Zhang, L., ... & Lahn, B. T. (2012). Embryonic stem cells induce pluripotency in somatic cell fusion through biphasic reprogramming. Molecular cell, 46(2), 159-170.

[45] Lahn, B. T. (2011). The “occlusis” model of cell fate restriction. Bioessays, 33(1), 13-20.

[46] Paranjpe, S. S., & Veenstra, G. J. C. (2015). Establishing pluripotency in early development. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1849(6), 626-636.

[47] Lee, J. H., Gaetz, J., Bugarija, B., Fernandes, C. J., Snyder, G. E., Bush, E. C., & Lahn, B. T. (2009). Chromatin analysis of occluded genes. Human molecular genetics, 18(14), 2567-2574.

[48] Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., ... & Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell, 153(2), 307-319.

[49] Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., ... & Young, R. A. (2013). Super-enhancers in the control of cell identity and disease. Cell, 155(4), 934-947.

[50] Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R. C., & Melton, D. A. (2002). " Stemness": transcriptional profiling of embryonic and adult stem cells. Science, 298(5593), 597-600.

[51] Meshorer, E., & Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nature reviews Molecular cell biology, 7(7), 540-546.

[52] Ram, E. V. R., & Meshorer, E. (2009). Transcriptional competence in pluripotency. Genes & development, 23(24), 2793-2798.

[53] Szutorisz, H., & Dillon, N. (2005). The epigenetic basis for embryonic stem cell pluripotency. Bioessays, 27(12), 1286-1293.

[54] Pang, B., & Snyder, M. P. (2020). Systematic identification of silencers in human cells. Nature genetics, 52(3), 254-263.

[55] Doni Jayavelu, N., Jajodia, A., Mishra, A., & Hawkins, R. D. (2020). Candidate silencer elements for the human and mouse genomes. Nature communications, 11(1), 1061.

[56] Pang, B., van Weerd, J. H., Hamoen, F. L., & Snyder, M. P. (2023). Identification of non-coding silencer elements and their regulation of gene expression. Nature Reviews Molecular Cell Biology, 24(6), 383-395.

[57] Khan, M., & Gasser, S. (2016). Generating primary fibroblast cultures from mouse ear and tail tissues. JoVE (Journal of Visualized Experiments), (107), e53565.

[58] Linker, C., & Stern, C. D. (2004). Neural induction requires BMP inhibition only as a late step, and involves signals other than FGF and Wnt antagonists.

[59] Storey, K. G., Crossley, J. M., Robertis, E. M. D., Norris, W. E., & Stern, C. D. (1992). Neural induction and regionalisation in the chick embryo. Development, 114(3), 729-741.

[60] Nicetto, D., Donahue, G., Jain, T., Peng, T., Sidoli, S., Sheng, L., ... & Zaret, K. S. (2019). H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. Science, 363(6424), 294-297.

[61] Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes & development, 25(21), 2227-2241.

[62] Zaret, K. S., Watts, J., Xu, J., Wandzioch, E., Smale, S. T., & Sekiya, T. (2008, January). Pioneer factors, genetic competence, and inductive signaling: programming liver and pancreas progenitors from the endoderm. In Cold Spring Harbor symposia on quantitative biology (Vol. 73, pp. 119-126). Cold Spring Harbor Laboratory Press.

[63] Larson, E. D., Marsh, A. J., & Harrison, M. M. (2021). Pioneering the developmental frontier. Molecular cell, 81(8), 1640-1650.

[64] Smith, A. (2017). Formative pluripotency: the executive phase in a developmental continuum. Development, 144(3), 365-373.

[65] Hoogland, S. H., & Marks, H. (2021). Developments in pluripotency: a new formative state. Cell research, 31(5), 493-494.

[66] Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K. K., Cheng, C., ... & Snyder, M. (2012). Architecture of the human regulatory network derived from ENCODE data. Nature, 489(7414), 91-100.

[67] Ying, Q. L., Stavridis, M., Griffiths, D., Li, M., & Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nature biotechnology, 21(2), 183-186.

[68] Yu, Y., Wang, X., Zhang, X., Zhai, Y., Lu, X., Ma, H., ... & Li, L. (2018). ERK inhibition promotes neuroectodermal precursor commitment by blocking self-renewal and primitive streak formation of the epiblast. Stem Cell Research & Therapy, 9, 1-13.

[69] Soares, R., Ribeiro, F. F., Lourenço, D. M., Rodrigues, R. S., Moreira, J. B., Sebastião, A. M., ... & Xapelli, S. (2021). The neurosphere assay: an effective: in vitro: technique to study neural stem cells. Neural regeneration research, 16(11), 2229-2231.

[70] Knezetic, J. A., & Luse, D. S. (1986). The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. Cell, 45(1), 95-104.

[71] Workman, J. L., & Roeder, R. G. (1987). Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. Cell, 51(4), 613-622.

[72] Zaret, K. S. (2020). Pioneer transcription factors initiating gene network changes. Annual review of genetics, 54, 367-385.

[73] Luzete-Monteiro, E., & Zaret, K. S. (2022). Structures and consequences of pioneer factor binding to nucleosomes. Current opinion in structural biology, 75, 102425.

[74] Ito, T., Bulger, M., Pazin, M. J., Kobayashi, R., & Kadonaga, J. T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. Cell, 90(1), 145-155.

[75] Lundberg, D., Berezhnoy, N. V., Lu, C., Korolev, N., Su, C. J., Alfredsson, V., ... & Nordenskiold, L. (2010). Interactions between cationic lipid bilayers and model chromatin. Langmuir, 26(15), 12488-12492.

[76] Berezhnoy, N. V., Korolev, N., & Nordenskiöld, L. (2014). Principles of electrostatic interactions and self-assembly in lipid/peptide/DNA systems: applications to gene delivery. Advances in colloid and interface science, 205, 221-229.

[77] Cereghini, S., & Yaniv, M. (1984). Assembly of transfected DNA into chromatin: structural changes in the origin-promoter-enhancer region upon replication. The EMBO journal, 3(6), 1243-1253.

[78] Reeves, R., Gorman, C. M., & Howard, B. (1985). Minichromosome assembly of non-integrated plasmid DNA transfected into mammalian cells. Nucleic acids research, 13(10), 3599-3615.

[79] Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Beliveau, B. J., Fudenberg, G., ... & Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature, 529(7586), 418-422.

[80] Wang, L., Gao, Y., Zheng, X., Liu, C., Dong, S., Li, R., ... & Li, P. (2019). Histone modifications regulate chromatin compartmentalization by contributing to a phase separation mechanism. Molecular cell, 76(4), 646-659.

[81] Cernilogar, F. M., Hasenöder, S., Wang, Z., Scheibner, K., Burtscher, I., Sterr, M., ... & Schotta, G. (2019). Pre-marked chromatin and transcription factor co-binding shape the pioneering activity of Foxa2. Nucleic Acids Research, 47(17), 9069-9086.

[82] Kim, H. S., Tan, Y., Ma, W., Merkurjev, D., Destici, E., Ma, Q., ... & Rosenfeld, M. G. (2018). Pluripotency factors functionally premark cell-type-restricted enhancers in ES cells. Nature, 556(7702), 510-514.

[83] Mercurio, S., Serra, L., & Nicolis, S. K. (2019). More than just stem cells: functional roles of the transcription factor Sox2 in differentiated glia and neurons. International journal of molecular sciences, 20(18), 4540.

[84] Ferri, A. L., Cavallaro, M., Braida, D., Di Cristofano, A., Canta, A., Vezzani, A., ... & Nicolis, S. K. (2004). Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain.

[85] Favaro, R., Valotta, M., Ferri, A. L., Latorre, E., Mariani, J., Giachino, C., ... & Nicolis, S. K. (2009). Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. Nature neuroscience, 12(10), 1248-1256.

[86] Wang, Y., Zhang, S., Lan, Z., Doan, V., Kim, B., Liu, S., ... & Guo, F. (2022). SOX2 is essential for astrocyte maturation and its deletion leads to hyperactive behavior in mice. Cell reports, 41(12).

[87] Bani-Yaghoub, M., Tremblay, R. G., Lei, J. X., Zhang, D., Zurakowski, B., Sandhu, J. K., ... & Sikorska, M. (2006). Role of Sox2 in the development of the mouse neocortex. Developmental biology, 295(1), 52-66.

[88] Meijer, D. H., Kane, M. F., Mehta, S., Liu, H., Harrington, E., Taylor, C. M., ... & Rowitch, D. H. (2012). Separated at birth? The functional and molecular divergence of OLIG1 and OLIG2. Nature Reviews Neuroscience, 13(12), 819-831.

[89] Setoguchi, T., & Kondo, T. (2004). Nuclear export of OLIG2 in neural stem cells is essential for ciliary neurotrophic factor–induced astrocyte differentiation. The Journal of cell biology, 166(7), 963-968.

[90] Magnus, T., Coksaygan, T., Korn, T., Xue, H., Arumugam, T. V., Mughal, M. R., ... & Mattson, M. P. (2007). Evidence that nucleocytoplasmic Olig2 translocation mediates brain-injury-induced differentiation of glial precursors to astrocytes. Journal of neuroscience research, 85(10), 2126-2137.

[91] Lattke, M., Goldstone, R., Ellis, J. K., Boeing, S., Jurado-Arjona, J., Marichal, N., ... & Guillemot, F. (2021). Extensive transcriptional and chromatin changes underlie astrocyte maturation in vivo and in culture. Nature communications, 12(1), 4335.

[92] Dodonova, S. O., Zhu, F., Dienemann, C., Taipale, J., & Cramer, P. (2020). Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. Nature, 580(7805), 669-672.

[93] Vanzan, L., Soldati, H., Ythier, V., Anand, S., Braun, S. M., Francis, N., & Murr, R. (2021). High throughput screening identifies SOX2 as a super pioneer factor that inhibits DNA methylation maintenance at its binding sites. Nature Communications, 12(1), 3337.