TITLE

Mechanistic basis of lineage restriction

AUTHORS

Bohou Wu[1,3], Jae Hyun Lee[1,2], Kara M. Foshay[1], Li Zhang[1], Croydon J. Fernandes[1,2], Boyang Gao[4], Xiaoyang Dou[4], Chris Z. Zhang[5], Guoping Fan[6], Becky X. Xiao[2], Bruce T. Lahn[1,2,3,7]

AFFILIATIONS

[1]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA

[2]VectorBuilder, Chicago, IL 60609, USA

[3]Future Institute of Gene Delivery Research, Guangzhou, 510663, China

[4]Department of Chemistry, Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637, USA

[5]Lantu Biopharma, Guangzhou, 510663, China

[6]Department of Human Genetics, Broad Stem Cell Research Center, David Geffen School of Medicine, UCLA, Los Angeles, CA 90095, USA.

[7]To whom correspondence should be addressed. Email: brucelahn@vectorbuilder.net

ABSTRACT

Lineage restriction, the biological phenomenon whereby developing cells progressively lose fate potency for all but their adopted lineages, is foundational to multicellular lifeforms as it secures the functional identities of the myriad cell types in the body. The mechanisms of lineage restriction remain enigmatic. We previously defined occlusion as a mode of gene silencing wherein affected genes lack the transcriptional potency to be activated by their cognate transcription factors (TFs). Here, we present a comprehensive mechanistic basis of lineage restriction as driven by gene occlusion. Specifically, we show that genes can become occluded simply by the default action of chromatinization in the absence of TF binding, that naive pluripotent stem cells establish full developmental potency via their capacity to erase occlusion, that primed pluripotent cells shut down this deocclusion ability in preparation for differentiation, that differentiating cells become increasingly restricted in their fate potency by the irreversible occlusion of lineage-inappropriate genes, and that stem cells employ placeholder factors (PFs) to protect silent genes needed for later activation from premature occlusion. Collectively, these mechanisms drive lineage restriction whereby the transcriptionally potent portion of the genome shrinks progressively and irreversibly during differentiation, rendering the fate potency of developing cells to also dwindle progressively.

ONE-SENTENCE SUMMARY

Mechanistic link between lineage potency of developing cells and transcriptional potency of their genomes

KEYWORDS

lineage restriction, cell fate restriction, occlusion, occlusis, chromatinization, naive pluripotency, primed pluripotency, nucleosome-dominated state, factor-dominated state, Esrrb, Sox2

INTRODUCTION

The hallmark of multicellular life is the presence, within a single organism, of myriad cell types bearing the same genome but disparate transcriptional profiles. This is achieved when a single zygote proliferates and progressively differentiates along a branchwork of lineages to eventually give rise to a multitude of specialized cell types. A defining feature of this process is lineage restriction (aka cell fate restriction), whereby differentiating cells progressively lose their fate potency for all but their adopted lineages (Fig. 1A) (*1*). Lineage restriction is a prerequisite for multicellularity as it ensures that the myriad cell types in the body, once created during development, would faithfully maintain their committed functional identities over the entire lifespan of the organism (*1, 2*). As yet, the mechanisms of lineage restriction are essentially unknown and remain one of the most important open questions in biology (*2, 3*).

Our previous studies hinted that the progressive restriction of fate potency of differentiating cells might stem from shrinking transcriptional potency of their genomes (*4-8*). Specifically, we developed a cell fusion assay to determine the transcriptional potency of genes based on whether they can be activated by their cognate TFs availed by the fusion. This assay uncovered a class of genes known as occluded genes, which have lost the potency to be activated by their TFs. We hypothesized that during development, as cells differentiate progressively down specific lineages, lineage-inappropriate genes that drive alternative cell fates become progressively and irreversibly occluded. This restricts the fate potency of cells to within their adopted lineages because key genes outside these lineages can no longer be activated. However, our previous studies did not convincingly link cell fate potency to genome transcriptional potency, nor was any molecular mechanism of occlusion offered.

In this study, we discovered that genes not protected by their TFs can, by default, become occluded simply by chromatinization into the nucleosomal form. We systematically analyzed genome-wide transcriptional potency of multiple cell types representing varying degrees of developmental potency, and uncovered several key mechanisms that link declining fate potency of developing cells with reduced transcriptional potency of their genomes, including:

1. Naive pluripotent stem cells in early development possess the ability to erase occlusion globally in order to establish full transcriptional potency of the genome.
2. As naive cells progress into the primed pluripotent stage, such deocclusion capacity is abolished in preparation for differentiation, which is achieved via the occlusion of Esrrb, a key component of the deocclusion machinery. From this stage onward, genes in the genome can undergo occlusion in ensuing differentiation where needed.
3. In stem cells beyond the naive stage, silent genes whose activation is needed in later differentiation are protected from occlusion by PFs such as Sox2.
4. As differentiation proceeds, lineage-inappropriate genes extraneous to the adopted lineages, especially master regulators for alternative fates, undergo irreversible occlusion by the default action of chromatinization when their TFs and/or PFs disappear from cells.

Collectively, these mechanisms drive lineage restriction in a process that we hypothesized previously and termed “occlusis” (*5*), whereby the portion of the genome that still retains transcriptional potency shrinks progressively and irreversibly during differentiation, rendering the fate potency of developing cells to dwindle progressively as well.

RESULTS

Assessing transcriptional potency of genes by intraspecies cell fusion

We previously developed a cell fusion assay to ascertain the transcriptional potency of silent genes (*4, 8*). It entails fusing two different cell types followed by examining expression patterns of the two genomes separately in fused cells. The assay led to the discovery of two classes of silent genes: activatable and occluded. Albeit indistinguishably silent before fusion, activatable and occluded genes show distinct transcriptional potency in response to their cognate TFs made available by the fusion. Activatable genes are readily activated in hybrid cells, indicating that they are potent to respond to TFs availed by the fusion. By contrast, occluded genes remain silent after fusion even though their orthologs in the fusion partner's genome are actively expressed. This indicates their lack of potency to be activated by their TFs, which are clearly available in hybrid cells as evidenced by the active expression of their orthologs in the same hybrids.

However, our previous fusion studies were mostly performed between cell lines of different species in order to distinguish transcripts between the two genomes of hybrid cells. Differential expression of genes between partner genomes could be due to interspecies incompatibility – i.e., the inability of TFs from one species to activate their orthologous target genes from the other species – rather than occlusion. Moreover, there were no systematic comparisons of gene potency across different cell types of varying developmental potency. To address these drawbacks, we performed fusion between cells both of mouse origin but from different strains, followed by whole transcriptome analysis that exploited sequence polymorphisms between strains to assign transcripts to the correct fusion partner (Fig. 1B). Furthermore, we selected a range of cell types representing different developmental potency for the fusions.

We first established two mouse fibroblast clonal lines, SEF2 and CTF3, from SPRET/EiJ and CAST/EiJ strains, respectively. We then fused either of them to cell lines from common laboratory mouse strains (i.e., C57BL/6, 129 or C3H), followed by transcriptome analysis by RNA-Seq. SPRET/EiJ and CAST/EiJ are inbred wild mouse strains bearing sufficient polymorphisms from common lab strains to allow transcripts from most genes in hybrid cells to be assigned to the correct fusion partner. Fluorescence and drug selection marker genes were integrated into these cell lines to facilitate the isolation of hybrid cells. As a proof of concept, we fused SEF2 with the mouse myoblast cell line C2C12, and cultured the resulting SEF2-C2C12 hybrid cells for over two weeks to ensure that SEF2 and C2C12 genomes were combined into the same nucleus after multiple cell divisions, and that any mRNA remaining from the pre-fusion parental cells were mostly eliminated. RNA-Seq was then performed on the cells to quantify strain-specific gene expression. In either SEF2 or C2C12 parental cells prior to fusion, housekeeping genes, exemplified by Tubb5, were shown by our polymorphism-based transcriptome mapping to only produce transcripts matching the strain of the cell line, whereas in SEF2-C2C12 hybrid cells, transcripts matching both strains were detected (Fig. 1C, top). This profile validated our intraspecies fusion assay and data analysis pipeline.

We next focused on genes differentially expressed in SEF2 and C2C12 before fusion, and examined whether their post-fusion expression changes were consistent with their being occluded or activatable. In line with previous interspecies fusion studies (*4, 8*), this intraspecies fusion revealed that silent genes can vary dramatically in transcriptional potency, with occluded genes lacking the potency to be activated by their TFs whereas activatable genes possessing such potency. An example of occluded genes is Ifitm1 in SEF2. It was silent in SEF2 but active in C2C12 prior to fusion, and this differential expression persisted in hybrid cells (Fig. 1C, middle). A contrasting pattern was observed for activatable genes, exemplified by Arhgap32. Like Ifitm1, Arhgap32 was silent in SEF2 but active in C2C12 before fusion. Yet, upon fusion, unlike Ifitm1, Arhgap32 became actively expressed from both SEF2 and C2C12 genomes of

hybrid cells (Fig. 1C, bottom). The SEF2-C2C12 fusion thus allowed us to systematically identify occluded and activatable genes in both cell lines without the confound of interspecies incompatibility (Fig. 1D). It is worth noting that a given fusion assay is not informative about the transcriptional potency of all the silent genes for several reasons. One is when a gene is silent in both cell lines before and after fusion. The other scenario, known as extinction, is when a gene differentially expressed in the two pre-fusion parental cell lines becomes silent in both genomes after fusion. Genes become extinguished in hybrid cells presumably due to the dilution or disappearance of relevant TFs upon fusion, or the action of transcriptional repressors (*8, 9*). Additionally, expression levels of a gene in the fusion assay may not cross the statistic threshold needed to call whether it is occluded or activatable. We will refer to silent genes in a cell type whose transcriptional potency can be ascertained by the fusion assay as being informative.

We next fused C2C12 with CTF3 to test whether the measurement of occlusion is consistent between different fusion pairs (Fig. S1A). Of the 120 C2C12 occluded genes identified in the SEF2-C2C12 fusion, 29 were informative in the CTF3-C2C12 fusion, and of these, 28 were also found to be occluded. This result, together with additional experiments described later, demonstrates that the annotation of occluded genes in a cell line is remarkably consistent when fused to different partner cell lines (except when fused to naive pluripotent stem cells as described later). That a minority of genes might appear occluded in one fusion but activatable in another could be due to several sources of uncertainty in the assay (see Supplementary Text).

We selected Myf5, a myogenic master regulator occluded in CTF3 but expressed in C2C12 for further validation. PCR products from gDNA or cDNA of CTF3-C2C12 hybrid cells were sequenced (Fig. S1B). While similar levels of Myf5 gDNA from SEF2 and C2C12 were detected based on polymorphism, Myf5 mRNA was only detected from C2C12 in the fusion sample, indicating that only the C2C12 copies but not the CTF3 copies of Myf5 was expressed in hybrid cells.

Restriction of cell lineage potency is associated with reduction in genome transcriptional potency

We next selected a series of mouse cell lines representing varying degrees of developmental potency and assayed transcriptional potency of their genomes by cell fusion (Fig. 1E). For pluripotent stem cells, we used the mouse embryonic stem cell (ESC) line E14 derived from the inner cell mass (ICM) of pre-implantation blastocyst and the mouse epiblast stem cell (EpiSC) line G14 derived from the epiblast of post-implantation embryo (Fig. 1E). ESCs are said to possess naive pluripotency that represents an early stage of development not yet committed to differentiation, whereas EpiSCs are said to possess primed pluripotency that represents a later stage of development already primed to differentiate (Fig. 1A) (*10*). RNA-Seq on E14 and G14 confirmed the expression of characteristic naive and primed markers, respectively. For somatic cells with more restricted fate potency, we chose the mouse neural stem cell (NSC) line B6NSC, microglia cell line IMG, and the above described C2C12 and CTF3. SEF2 was selected for fusion with most of these cell lines to assess their genome potency because of its abundant sequence polymorphisms with them.

To examine the transcriptional potency of E14, G14, B6NSC, C1C12 and IMG genomes, informative silent genes in these cells were classified as occluded or activatable based on fusion with SEF2. Similar analysis was also performed on CTF3 based on its fusion with C2C12. For the four somatic cell types, B6NSC, C1C12, IMG, and CTF3, more silent genes were occluded than activatable (Fig. 1F, Fig. S2A). For the pluripotent G14, by contrast, far more silent genes were activatable than occluded (Fig. 1F, Fig. S2B). We performed the same analyses on clonal hybrid lines derived from the above bulk fusions, and obtained similar results (data not shown).

Measuring genome potency of E14 cells in SEF2-E14 fusion was problematic because the vast majority of SEF2-specific genes (i.e., active in SEF2 but silent in E14) became extinguished in the SEF2 genome after fusion (Fig. S2C), presumably due to ESC's strong reprogramming ability that robustly repressed SEF2-specific genes (see more description later). We reasoned that reprogramming is a gradual process and as such, E14 genome potency can perhaps be measured in early-stage fusion samples before reprogramming has taken full effect. We therefore reanalyzed time-course fusion data from our previous study where the rat fibroblast cell line R1A was fused with E14 (*7*). Unlike SEF2, R1A fusion efficiency with E14 was high enough to produce sufficient amounts of early-stage hybrid cells for RNA-Seq analysis. Similar to SEF2-E14 fusion, most R1A-specific genes became extinguished in R1A-E14 hybrid cells after long-term cell culture. However, large-scale gene extinction was not obvious for day 2, 4, and 8 post-fusion, enabling measurement of E14 genome potency in these early-stage fusion samples (Fig. S2D). Similar to G14 cells, most informative silent genes in the pluripotent E14 cells were activatable (Fig. 1F, Fig. S2D). We note that our fusions might overestimate the fraction of occluded genes, especially in E14, while underestimate activatable genes (see Supplementary Text).

It has been proposed that the fate potency of cells is actively acquired during development by pioneer transcription factors, rather than gradually restricted (*11-13*). To address this notion, we examined the expression of activatable and occluded genes in B6NSC cells after their differentiation into astrocytes (Fig. 1G). While 14% (12 out of 86) of B6NSC activatable genes are activated, only 2.5% (5 out of 198) of B6NSC occluded genes showed expression in differentiated astrocytes, and the difference is statistically significant ($p<0.0005$ by Fisher's exact test). This observation argues that the loss of gene potency is essentially irreversible during somatic differentiation, and that the fate potency of differentiating cells is irreversibly restricted rather than actively acquired. The small number of occluded B6NSC genes that turned on after differentiation could be due to uncertainty in the measurement of occlusion (see Supplementary Text). Additional evidence for the irreversibility of occlusion in somatic cells is described later.

Occluded genes are enriched for developmental master regulators

During lineage differentiation, the activation of lineage-specific genes follows a hierarchical order, with master regulator genes (MRGs) that specify the target lineage – typically TFs – turning on first, which in turn activate downstream effector genes to produce physiological properties characteristic of the resulting cell type (*14*). We reasoned that if gene occlusion indeed serves the purpose of restricting the fate potency of a differentiated cell type, then it might act on MRGs more than effectors, as occlusion of MRGs should be sufficient to shut down entire lineage programs including their target effectors. Gene ontology (GO) analysis is in line with our hypothesis. Genes occluded in SEF2 identified by the SEF2-B6NSC fusion were highly enriched in MRGs of neural fates, including C2H2-type zinc finger proteins such as Zic1, Zic2 and Zic5, high mobility group (HMG) box domain proteins such as Sox1, Sox2 and Sox11, as well as homeodomain proteins such as Pou3f1 and Pou4f1. By contrast, activatable genes were enriched for downstream effector functions, such as neuronal projection and extension, GTPase activation, and vesicle fusion (Fig. 1H). Similarly, occluded SEF2 genes identified in SEF2-C2C12 and SEF2-IMG fusions were enriched in transcriptional regulators and developmental signals (Fig. S3).

The above observation also sheds light on the phenomenon of transdifferentiation by ectopic expression of MRGs or cell fusion (*15, 16*). On the surface, such transdifferentiation seems to contradict the irreversibility of gene occlusion or cell fate restriction (*17*). However, there is no contradiction if the forced introduction of ectopically expressed MRGs by either transgenes or

cell fusion are only turning on their target effectors that are activatable, which is sufficient to confer phenotypic similarities to alternative cell fates, while leaving occluded genes still occluded. Indeed, we repeated the classic experiment of converting fibroblasts into myotubes by ectopically expressing the myogenic master regulator Myf5 (*18*), and noticed that muscle-related genes activated in fibroblasts by the transgene were themselves activatable, whereas occluded genes, including endogenous Myf5, remained silent (*6*) (unpublished data).

Deocclusion ability is present in naive pluripotency but shut off in primed pluripotency

Early mammalian embryos possess pluripotent stem cells capable of differentiating into all somatic lineages and the germline. Their pluripotency is proposed to progress from naive to primed stages before embarking on lineage differentiation (*10, 19*). We previously reported that naive pluripotent cells possess the capacity to erase occlusion across the genome, and this deocclusion ability is lost in differentiated somatic cells (*7*). However, our previous study was limited by interspecies fusion and did not address whether primed pluripotent cells were capable of deocclusion or not. We therefore analyzed signs of deocclusion in the same set of intraspecies fusions involving SEF2 as described above. Focusing on genes differentially expressed between pre-fusion partner cells, we employed hierarchical clustering to quantify changes in expression patterns before and after fusion. Strikingly, it showed that in somatic-somatic fusions (i.e., SEF2 fused to C2C12, IMG or B6NSC), the correlation of expression patterns from the same genome before and after fusion was much greater than that between the two fusion partner genomes in the same fused cells (Fig. 2A). This is mainly attributable to the presence of occluded genes from either of the two partner genomes, which remained similarly differentially expressed post-fusion as pre-fusion. By contrast, in the SEF2-E14 fusion, expression pattern of the SEF2 genome was reprogrammed to become quite similar to E14 (Fig. 2A). This is due to genes differentially expressed between pre-fusion SEF2 and E14 mostly adopting expression patterns similar to E14 after fusion. Importantly, the SEF2-G14 fusion showed a very distinct pattern from all the above fusions. Unlike E14, G14 didn't reprogram the gene expression pattern of SEF2 post-fusion, but rather changed its own expression pattern closer to SEF2 (Fig. 2A). This is consistent with G14's high genome potency as described earlier, and low reprogramming ability as described below.

We next interrogated whether SEF2 genes found to be occluded in fusions with other somatic cells would undergo deocclusion when SEF2 was fused to the pluripotent E14 or G14. SEF2 occluded genes annotated in SEF2-C2C12, SEF2-IMG and SEF2-NSC fusions were combined, and the subset with expressed orthologs in SEF2's pre-fusion partners were selected for analysis. Occluded genes in the combined list were rarely activated in any of the three somatic-somatic fusion pairs, confirming the consistency of gene potency measurement across different somatic fusions (Fig. S4). By contrast, upon fusion with E14, most SEF2 occluded genes were activated, indicating the remarkable deocclusion ability of ESCs (Fig. 2B). Strikingly, when fused to G14, SEF2 occluded genes did not undergo massive activation. Instead, they mostly remained occluded similar to when SEF2 was fused to other somatic cells (Fig. 2B). This indicates that G14 does not possess the robust deocclusion capacity as E14.

We examined several representative genes in detail, including Peg3, Bend4, and Sox2. These genes in SEF2 were annotated as occluded based on fusions to other somatic cells, whereas they were highly expressed in E14 and G14. In the SEF2-E14 fusion but not the SEF2-G14 fusion, the SEF2 copies of these genes were robustly deoccluded, displaying expression levels comparable to their E14 orthologs (Fig. 2C). Prompted by this observation, we checked the expression of two pluripotency genes, Nanog and Pou5f1, whose transcriptional potency in SEF2 could not be ascertained by fusions with other somatic cells because these two genes are silent in somatic cells. Nanog expression from the SEF2 genome resembled occluded genes,

being activated upon fusion with E14 but not G14 (Fig. 2C). Pou5f1 was slightly activated from the SEF2 genome in SEF2-G14 hybrids, but to a much lesser extent as compared to the activation seen in the SEF2-E14 fusion (Fig. 2C). We performed the same analyses using clonal hybrid lines derived from the above bulk fusions, which produced similar results (data not shown). Together, the above data demonstrate convincingly that naive pluripotent cells, represented by E14, possess the deocclusion capacity, which is largely shut off when cells enter the primed pluripotent stage, represented by G14, before embarking on differentiation.

Esrrb is responsible for the contrasting deocclusion ability between naive and primed pluripotency

While naive and primed pluripotent states share very similar gene expression patterns, there are a number of genes expressed in naive but silent in primed cells. We speculated that some of these genes, especially those implicated in pluripotency, might encode key components of the deocclusion machinery present in the naive state but absent in the primed state. We focused on two such candidate genes, Klf4 and Esrrb, both highly expressed in naive cells but silent in primed cells (*20, 21*), and which were previously implicated in promoting naive pluripotency (*22-25*). We fused either E14 or G14 with the rat neuroblastoma cell line B35 due to the ease to manipulate the resulting hybrid cells including lentiviral transduction. RNA-Seq on the G14-B35 fusion sample uncovered a set of robustly occluded genes in B35 that showed little expression from the B35 genome but high expression from the G14 genome of hybrid cells, and the same pattern persisted in two clones derived from the bulk fusion (Fig. 3A, left). In the E14-B35 fusion, by contrast, these occluded B35 genes showed noticeable activation by day 8 post-fusion, and even greater activation in two hybrid clones (Fig. 3A, middle). This is consistent with the presence of robust deocclusion capacity in naive but not primed pluripotent cells as described earlier.

We then transduced the hybrid clone G14-B35(clone9) with lentivirus expressing either Klf4 or Esrrb, and performed RNA-Seq after prolonged culture. In the Klf4-transduced hybrid cells, the occluded B35 genes remained similarly differentially expressed between the B35 and G14 genomes, whereas in the Esrrb-transduced sample, the occluded B35 genes became robustly activated, reaching expression levels comparable to their G14 orthologs, just like in E14-B35 hybrid cells (Fig. 3A, left). Thus, expression of ectopic Esrrb but not Klf4 is sufficient to confer deocclusion capability to primed pluripotent cells that otherwise lack this ability.

Next, we examined whether Esrrb is necessary for the deocclusion capacity in naive cells. We utilized an Esrrb knockout ESC line in which exon 2 was deleted on both copies of the gene, referred to hereon as ESC(EsrrbKO) (*25*). RNA-Seq analysis confirmed that this cell line has a characteristic ESC transcriptome profile. ESC(EsrrbKO) was fused with B35, cultured for 8 days and harvested for RNA-Seq. For the occluded B35 genes, their behavior in this fusion was very similar to that of the G14-B35 fusion, namely they remained highly differentially expressed between the ESC(EsrrbKO) and B35 genomes of the hybrid cells just as in parental cells before fusion (Fig. 3A, right). This stands in sharp contrast to the fusion between B35 and the wildtype ESC line E14, demonstrating that Esrrb is necessary for the deocclusion capacity present in naive pluripotency. Collectively, the above data argue that the silencing of Esrrb in the primed state is responsible for the loss of deocclusion capacity therein.

Two of the occluded B35 genes are noteworthy. One is the pluripotency gene Pou5f1, whose behavior is similar to that seen in SEF2 when fused to E14 or G14 as described earlier. The other is Tert, the telomerase gene known to be expressed in pluripotent stem cells and many cancers, but silent in most normal somatic cells (*26*). Our data suggest that Tert is not only silent, but occluded in somatic cells.

Esrrb is occluded in primed pluripotent cells

We hypothesized that the silencing of Esrrb in primed pluripotent cells is itself the consequence of occlusion. To test this, we fused ESC(EsrrbKO) to G14 to examine if the G14 copies of Esrrb would show signs of occlusion in fused cells. We used ESC(EsrrbKO) rather than a wildtype ESC line in the fusion for two reasons. First, any occluded genes in G14 would have their occlusion erased upon fusion with wildtype ESCs, given that the latter possess the deocclusion ability. ESC(EsrrbKO) cells on the other hand, have lost their deocclusion ability due to Esrrb knockout as described above. Second, the Esrrb knockout allele in ESC(EsrrbKO) cells lacks exon 2 but retains other exons. It is therefore still expressed except its mRNA lacks exon 2. This allows the mapping of Esrrb transcripts in hybrid cells to either the ESC(EsrrbKO) or the G14 genome based on whether it contains exon 2. We performed RT-PCR on the ESC(EsrrbKO)-G14 fusion sample after extended culture. Primers designed to only amplify the wildtype Esrrb transcripts did not produce any product, whereas primers designed to amplify both wildtype and knockout Esrrb transcripts produced strong products (Fig. 3B). This indicates that in hybrid cells, the wildtype Esrrb copies in the G14 genome is silent, whereas the knockout Esrrb copies in the ESC(EsrrbKO) genome is expressed, consistent with the occluded status of Esrrb in G14.

Irreversibility of gene occlusion during somatic development

We reasoned that in order for occlusion to play the role of restricting the fate potency of somatic cells, it needs to act irreversibly even when cells encounter developmental cues for alternative fates. To directly test this, we selected two clonal lines from each of the SEF2-E14 and SEF2-G14 fusions, and differentiated them into NSCs. Despite the existence of a fibroblast genome, all four hybrid clones successfully generated neurospheres (Fig. 4A). Less cell death occurred during the differentiation of the SEF2-E14 clones as compared to the SEF2-G14 clones, likely because the SEF2 genome in the SEF-E14 fusion but not the SEF-G14 fusion was extensively reprogrammed toward the pluripotent state as described earlier. In either case, typical neurospheres formed readily, which were examined by RNA-Seq.

We first confirmed that the pluripotency markers Pou5f1 and Nanog turned off completely in all hybrid clones upon differentiation, indicating full exit from pluripotency (Fig. 4B). We then checked several NSC markers known to be upregulated when differentiating pluripotent cells into NSCs (Fig. 4B). In the two SEF2-E14 hybrid clones, these markers turned on from both SEF2 and E14 genomes in neurospheres, consistent with the reprogramming of the SEF2 genome into pluripotent state including the deocclusion of occluded genes therein. Remarkably, in the two SEF2-G14 hybrid clones, whereas the NSC markers in the G14 genome were properly expressed in neurospheres as would be expected of successful NSC differentiation, their SEF2 orthologs failed to turn on. Their failure to turn on during NSC differentiation even when their G14 orthologs became properly expressed indicates a permanent loss of transcriptional potency for these genes in SEF2. These results support the model that differentiating cells lose fate potency via the irreversible occlusion of lineage-inappropriate genes, and are at odds with the notion that fate potency is actively acquired during development (*11-13*).

We next examined the overall expression of occluded SEF2 genes whose E14 or G14 orthologs were upregulated in the hybrids upon NSC differentiation (Fig. 4C). As predicted by the irreversible gene occlusion model, expression of occluded SEF2 genes were only slightly elevated in SEF2-G14 hybrid clones post differentiation, which stands in sharp contrast to their full activation in SEF2-E14 hybrid clones. To get a global view, we examined all the genes upregulated in the E14 genome of the SEF2-E14 fusion or the G14 genome of the SEF2-G14 fusion upon differentiation, irrespective of whether their potency was ascertained (Fig. 4D). In SEF2-E14 hybrid clones, many of these genes were upregulated to comparable levels from

both SEF2 and E14 genomes upon differentiation into neurospheres. By contrast, in SEF2-G14 hybrid clones, the great majority of genes showed full activation only in the G14 but not SEF2 genome. Essentially, albeit the two partner genomes in SEF2-G14 hybrid cells resided in the same nucleus, went through the same proliferation, and experienced the same differentiation signals, only the G14 genome underwent proper differentiation. This result further contradicts the notion that cells actively acquire fate potency during development (*11-13*).

Evidence that occlusion can be the default state of chromatinized DNA

It is generally assumed that eukaryotic genes possess transcriptional potency by default, and their stable silencing requires additional repressive chromatin modifications (*27, 28*). Accordingly, we previously probed the mechanism of occlusion by searching for chromatin marks associated with occluded genes. But this approach yielded little success (*8, 29*), prompting us to consider an opposing model that certain genes by default are transcriptionally inert – namely they are occluded – unless protected by some means from entering this default state. The eukaryotic nuclear genome is packaged by histones into nucleosome arrays as the default state, except in regions protected by DNA-binding factors such as TFs. DNA undergoing such packaging is said to be chromatinized. We therefore explored whether chromatinization alone was sufficient to render genes occluded by default.

Previous studies have shown that in vitro transcription is inhibited by chromatinization of the DNA template (*30, 31*). But how chromatinization affects transcription in living cells has not been systematically elucidated. To examine this, we selected four promoters, EF1A, Hsp68, UBC, and Nanog, and made four reporter plasmids each bearing one of the promoters driving a different fluorescence reporter (Fig. 5A). Plasmid DNA was chromatinized in vitro with HeLa core histones, or recombinant core histones lacking eukaryotic posttranslational modifications. Partial digestion of chromatinized DNA with MNase revealed regularly spaced nucleosome arrays, confirming successful chromatinization (Fig. 5B). We then transfected naked or chromatinized DNA constructs into 293T cells followed by three days of culture. Remarkably, while robust fluorescence was observed in cells transfected with naked constructs, only sparse and dim signals were detected in cells transfected with chromatinized constructs (Fig. S5). We quantified mRNA expression levels from the fluorescence reporters by RT-qPCR, and normalized it to the amount of plasmid DNA reaching the nuclei of transfected cells as measured by qPCR on DNA extracted from nuclei. Consistent with fluorescence data, chromatinized constructs exhibited negligible expression compared to naked constructs (Fig. 5C). This result revealed that simply by chromatinization into the nucleosomal form, genes can become unresponsive to their cognate TFs. Crucially, the silencing effect of chromatinization holds true even when using recombinant histones devoid of any eukaryotic posttranslational modifications. This argues that chromatinization can trigger gene silencing without any DNA or histone modifications, though it is reasonable to assume that modifications can be subsequently added to the silent chromatin (see further discussion below). We took these data as evidence that occlusion can be the default state of chromatinized DNA in somatic cells.

One caveat of the above experiment is that the expression difference observed between chromatinized and naked plasmids might be due to the former being more prone to getting trapped in areas of the nucleus unconducive to transcription. Additionally, it does not speak to the resolution of chromatinization's repressive effect on gene expression, namely, whether it impacts individual genes or expansive regions spanning multiple genes. To address these questions, we devised a half-chromatinization assay. We prepared two reciprocal linear DNA fragments with compatible sticky ends to facilitate their ligation into a circular construct (Fig. S6A). The sticky ends were designed to only permit ligation between the two fragments in the correct orientation, and did not support circularization of either fragment, or ligation between the

two fragments in the wrong orientation. One fragment carried the 3' half of the mCherry reporter with a polyadenylation signal (PolyA), followed by the EF1A promoter driving the 5' half of the TurboGFP reporter. It was termed the TurboGFP fragment because its promoter activity would be reflected by TurboGFP expression. The other fragment carried the 3' half of TurboGFP with a PolyA, followed by the EF1A promoter driving the 5' half of mCherry, and was termed the mCherry fragment. These two reciprocal DNA fragments complement each other in that, individually, neither could express any functional TurboGFP or mCherry due to their truncation, but when the two fragments were ligated together, the result was a circular reporter construct containing complete expression cassettes for both TurboGFP and mCherry. We chromatinized either the TurboGFP or mCherry fragment with Hela histones, and ligated it to the naked version of the reciprocal fragment (Fig S6A). Recombinant histones were not used here because their chromatinized samples were hard to dissolve, leading to insufficient materials after the additional ligation step. In theory, each ligated DNA molecule should be half-chromatinized and half naked. Partial digestion with MNase showed that the pattern of regularly spaced nucleosome array was still visible, but diluted by a background smear as would be expected from the presence of naked DNA and possibly also nucleosome sliding (Fig. S6B). To examine ligation fidelity, we performed PCR across the ligation junctions and sequenced the PCR products, which produced the correct sequences expected from proper ligation. We then transfected the ligated samples into 293T cells, and observed that the ligation product between chromatinized TurboGFP and naked mCherry gave rise to an enrichment of mCherry single-positive cells, whereas the opposite was true for the ligation product between naked TurboGFP and chromatinized mCherry (Fig. S6C). RT-qPCR was used to quantify mRNA expression levels from the TurboGFP and mCherry reporters, and results were normalized to the amount of ligated DNA reaching the nuclei of transfected cells as measured by qPCR on DNA extracted from isolated cell nuclei. PCR primers were designed to flank the ligation junctions in order to only interrogate successfully ligated DNA. Consistent with fluorescence data, the ligated construct bearing chromatinized TurboGFP fragment showed reduced expression than the construct carrying the naked TurboGFP fragment, and the same is true for the mCherry fragment (Fig. S6D). The expression difference between chromatinized and naked DNA seen in this assay is not as dramatic as that seen in Fig. 5C, possibly due to nucleosome sliding on the ligated construct, which could partly degrade the difference in nucleosome loading between the chromatinized half and the naked half of the circular DNA molecule (*32*). This notwithstanding, the result lends additional support to our hypothesis that chromatinization alone can lead to gene occlusion by default. It also argues that the silencing effect of chromatinization can occur at the resolution of individual promoters.

Chromatinization-mediated gene occlusion can occur at the resolution of individual genes and is heritable through DNA replication

As yet, the observation that chromatinization alone could trigger gene silencing by default was made with episomal DNA that cannot replicate. To address if DNA replication could alter this effect, we sought to integrate the half-chromatinized DNA into the host genome. In the course of our work, we noticed that the two reciprocal DNA fragments used in the half-chromatinization assay could be transfected into cells without prior in vitro ligation, and upon cell entry, they would correctly ligate with each other to form a complete reporter construct, likely due to DNA repair mechanisms in cells (Fig. S7). We therefore co-transfected into 293T cells the naked mCherry fragment that also carried a pair of piggyBAC inverted terminal repeats (ITRs) (Fig. 5D), plus the TurboGFP fragment in either naked form or chromatinized with Hela or recombinant core histones, along with a plasmid expressing piggyBac transpose to drive genome integration of the construct formed by in-cell ligation of the two reciprocal fragments. Partial digestion by MNase confirmed proper nucleosome array formation of the chromatinized

TurboGFP fragment used for transfection (Fig. S8A). A week after transfection, mCherry-positive cells, enriched for piggyBac-mediated integration of in-cell ligated dual-fluorescence construct, were isolated by cell sorting. Cells were then cultured for over a month with continuous proliferation and assayed for fluorescence reporter expression (Fig. S8B). At this point, cells that remained mCherry positive should carry stably integrated reporter construct in the genome. We observed that when the TurboGFP fragment used in transfection was in the naked form, the great majority of mCherry-positive cells were also positive for TurboGFP. By contrast, when the TurboGFP fragment was chromatinized, the majority of mCherry-positive cells were negative for TurboGFP, and the effect was particularly pronounced when recombinant histones devoid of any posttranslational modifications were used (Fig. 5E). Furthermore, the expression patterns of the reporters persisted in culture over time. These results further support our hypothesis that chromatinization alone could render genes occluded by default in somatic cells, and that either the occluded or expressed state, once established, is stably inherited through cell division. Inferring from this, we propose that during somatic differentiation, genes can become occluded simply by the disappearance of their TFs that leaves them no longer protected from the default action of chromatinization.

We note that in the chromatinized TurboGFP samples, a minority of mCherry-positive cells were TurboGFP positive, which could be due to nucleosome sliding that partially eroded the chromatin difference between the TurboGFP and the mCherry halves of the ligated construct. Interestingly, the opposite was also observed in the naked TurboGFP sample, namely, a minority of mCherry-positive cells were TurboGFP negative. This suggests that the naked reporter could occasionally undergo occlusion. We speculate that there is competition between the transcription machinery and the nucleosome assembly machinery. It has been reported that naked DNA are readily assembled into chromatin when transfected into cells (*33*). We suggest that after our naked reporter construct enters cells, the transcription machinery tends to win the competition as TFs can bind to their target promoter on the construct with faster kinetics than nucleosome assembly. But occasionally, nucleosome assembly occurs first before productive TF binding could take place, leading to occlusion of the reporter. However, when the reporter construct is already chromatinized before entering cells, even if incompletely due to technical reasons or erosion by nucleosome sliding, the balance would tip heavily against TF binding and toward chromatinization-mediated occlusion.

Evidence for placeholder factors that protect genes from default occlusion

Our data thus far support the model that during somatic differentiation, fate potency is progressively restricted through the irreversible occlusion of lineage-inappropriate genes. Our data also point to a possible mechanism of occlusion, namely, genes can, upon the disappearance of their cognate TFs, become occluded simply through the default action of chromatinization. However, our model presents a dilemma for lineage-specific genes silent in early development but turned on later to drive the differentiation of specific lineages. If occlusion is irreversible in somatic cells as we postulated, then such genes must not undergo occlusion when silent during early development, such that they can undergo either activation or occlusion in later differentiation. How then do lineage-specific genes retain transcriptional potency while silent? To resolve this dilemma, we postulated that stem cells beyond the naive pluripotent stage (i.e., primed pluripotent cells and somatic stem cells) utilize PFs to protect silent genes needed for later activation from premature occlusion. PFs should be similar to TFs in that they can bind target genes to shield them from chromatinization-mediated default occlusion, but they differ from TFs in that they do not drive transcription by themselves. Identifying such PFs would fill a critical missing piece in our model.

We reasoned that should PFs exist, they might confer a more open chromatin configuration around their binding sites by displacing nucleosomes in a manner similar to TFs (*34, 35*). Consequently, activatable genes might show greater chromatin openness if they are indeed enriched for PF binding. We tested this in NSCs by analyzing published ATAC-Seq data (*36*), which, when combined with our own gene potency data, allowed the profiling of chromatin accessibility of NSC activatable and occluded gene (Fig. 6A). While both types of genes were similarly silent in NSCs, activatable genes indeed showed greater openness, suggesting the presence of PF binding. Actively expressed genes showed much greater openness as expected.

We reasoned that PFs should be abundantly expressed in NSCs where they bind to – and keep activatable – certain silent genes whose activation are required for later differentiation, but they should also turn off upon differentiation to allow their target genes to become either active or occluded. We therefore focused on three genes encoding DNA binding proteins previously implicated in neural development, Sox2, Olig2 and Hmgb2 (*37-41*), which are highly expressed in B6NSC but downregulated upon differentiation into astrocytes (Fig. 6B). We used CRISPR to delete coding regions of these genes in B6NSC, and fused the knockout cells with SEF2 to assay for changes in gene potency. Notably, in Sox2 knockout, six silent genes in B6NSC switched their potency from activatable to occluded status (Fig. 6C), indicating the requirement of Sox2 in maintaining the transcriptional potency of these silent genes. Importantly, reintroduction of Sox2 into knockout cells by lentivirus failed to revert the occluded status of these genes as measured by fusion with SEF2 (Fig. 6C), which is consistent with the irreversible nature of occlusion. These results argue that Sox2 act as a PF to protect these silent genes from becoming prematurely occluded in B6NSC. Moreover, the data suggest that genes becoming occluded after losing protection from their cognate PFs would become permanently unresponsive to such PFs, the same way that genes become permanently unresponsive to their cognate TFs once occluded. Interestingly, three genes originally active in B6NSC cells became silent as well as occluded after Sox2 knockout (Fig. S9A), suggesting that Sox2 is required not only for their expression, but also their transcriptional potency.

To examine whether Sox2 protein directly binds to these genes, we performed CUT&Tag (*42*), using several Sox2 antibodies. Of the above nine newly occluded genes following Sox2 knockout, reliable Sox2 binding peaks were detected in the vicinity of seven genes in wildtype B6NSC (Fig. 6D; Fig. S9D, Fig. S10).

We note that the real number of genes that lost transcriptional potency following Sox2 knockout should be much greater than what we observed. This is because ascertaining the transcriptional potency of silent genes in B6NSC, with or without Sox2 knockout, requires that the SEF2-B6NSC fusion is informative, namely, the genes are expressed from the SEF2 genome both before and after fusion. For genes silent in SEF2 to begin with, or active in SEF2 before fusion but become extinguished after fusion, the assay cannot speak to the potency of their B6NSC orthologs. The same applies to the other two NSC-related TFs analyzed below.

In Olig2 knockout B6NSC, we did not find reliable examples of activatable genes becoming occluded. However, 22 genes originally active in B6NSC lost both their expression and potency (Fig. S9B). We interpret this as Olig2 acting as a key TF to drive the expression of these genes, and once Olig2 was deleted, these genes not only became silent, but in the absence of Olig2 binding, underwent chromatinization-mediated default occlusion. There were also 10 genes originally active in B6NSC that turned off but remained activatable after Olig2 knockout.

Hmgb2-deleted B6NSC cells also did not show good examples of activatable-to-occluded switch. Of the seven genes originally expressed in B6NSC and turned off following Hmgb2 knockout, one became occluded while six remained activatable (Fig. S9C). Notably, Ier3 and Timp3 were silenced in both Sox2 and Hmgb2 knockout B6NSC. But their potency was lost only

after Sox2 deletion (compare Fig. S9A with S9C). Consistent with this, Sox2 binding was detected in both genes (Fig. S9D), suggesting that while both Sox2 and Hmgb2 were required for the expression of these two genes, Sox2 likely served as a PF to maintain transcriptional potency of these genes even when they became silent upon Hmgb2 knockout.

Collectively, the above data support a placeholder model whereby primed pluripotent cells and somatic stem cells use PFs to sustain the transcriptional potency of silent genes that need to turn on in later differentiation. When these cells receive developmental cues to differentiate down a particular lineage, PFs disappear from cells and consequently, the genes they protect become either activated if TFs for these genes emerge, or occluded if otherwise. This model may underlie bivalent genes in stem cells, namely, silent genes thought to be poised for later activation that possess both active and repressive histone modifications (*43*). Mechanistically, PF binding may well be the causal agent of bivalency. Indeed, Sox2 binds to many inactive genes in NSCs that turn on during differentiation, and many Sox2-bound promoters are bivalent (*38, 44, 45*). Importantly, the placeholder model is relevant to the long-explored but poorly understood concept of stemness (*46-49*), arguing that whereas the source of stemness in naive pluripotent cells lies in their deocclusion capacity, the source of stemness in stem cells beyond the naive stage lies in PFs that hold silent genes needed for later activation in the activatable state.

DISCUSSION

Our study provides a comprehensive mechanistic account of lineage restriction. At the onset of development, naive pluripotent stem cells possess the deocclusion machinery that establishes full transcriptional potency of the genome, and in doing so, confers full fate potency to cells. This puts cells on a developmental blank slate, from which cells of more restricted transcriptional potency – and hence fate potency – can be sculpted during subsequent differentiation. As naive pluripotent cells advance into primed pluripotency, this deocclusion capacity is shut off to prepare cells for lineage differentiation, and this is done via the occlusion of Esrrb, a key component of the deocclusion machinery. At this stage, the genome still retains full transcriptional potency, and cells retain full fate potency, but genes can now undergo occlusion in ensuing differentiation. In primed pluripotent stem cells and also somatic stem cells, silent genes needed to turn on in later differentiation are protected from occlusion by PFs. As differentiation proceeds, developmental cues that drive cells to differentiate toward specific lineages would accomplish two things. One is turning on lineage-specific genes needed to specify target lineages. The other is occluding lineage-inappropriate genes no longer needed in the adopted lineages, especially MRGs for alternative fates. The latter comes about when cognate TFs or PFs of these genes disappear from cells. Upon losing protection from such factors, genes undergo irreversible occlusion via the default action of chromatinization. The overall outcome is a process that we previously hypothesized and termed “occlusis” (*5*), whereby the portion of the genome that retains transcriptional potency shrinks progressively and irreversibly during differentiation, pushing the fate potency of the cells to dwindle progressively as well. Several outstanding questions about this process warrant further discussion.

One essential aspect of our model is that genes in somatic cells can undergo occlusion by the default action of chromatinization when their cognate TFs or PFs are not available to protect them, and furthermore, once the occluded state is established, it is not reversible even when relevant TFs or PFs reappear in cells. How might DNA-histone and DNA-factor interactions facilitate such a behavior? We propose that genes can be bistable in that they can assume one of two energetically stable states (Fig. 7A). One is the factor-dominated state where genes are stably bound by trans-acting factors available from the cellular milieu (i.e., TFs or PFs). Genes in this state can be either active if the factors are TFs, or activatable if the factors are PFs. The

other is the nucleosome-dominated state that corresponds to occlusion, where genes are stably packaged into the nucleosomal form. From a physicochemical perspective, these two states occupy two stable energy wells separated by a high energy barrier that prevents easy transition between the states (Fig. 7A). From an intuitive standpoint, DNA tightly wound in nucleosome arrays is shielded from factor binding, whereas DNA bound by an ensemble of factors on its cis-regulatory elements is shielded from nucleosome assembly.

A large body of literature has implicated chromatin modifications in transcriptional regulation (*50*). Some modifications are associated with silent genes, such as DNA CpG methylation and histone H3K9 methylation, and are referred to as silent, or repressive, marks. Other modifications are associated with active genes, such as H3K4 methylation and H3K9 acetylation, and are referred to as active marks. What role might these marks play in gene occlusion? We argue that chromatin marks are not themselves the causal agent in establishing either the factor-dominated or nucleosome-dominated state of genes, as this task is accomplished by the appearance and disappearance of TFs and PFs during development as described above. Instead, chromatin marks serve the supporting role of reenforcing the stability of either state once established. For genes in the factor-dominated state, TFs (and potentially also PFs) can recruit active marks (i.e., H3K4 methylation and H3K9 acetylation) and remove silent marks (i.e., DNA methylation and H3K9 methylation), and in doing so, make DNA-factor binding energetically more favorable while DNA-histone interaction less favorable (*51, 52*) (Fig. 7B). Conversely, for genes in the nucleosome-dominated state, TFs are no longer present to recruit active marks, allowing silent marks to be added by a baseline machinery in the cell, which makes DNA-factor binding energetically less favorable while DNA-histone interaction more favorable (*53, 54*) (Fig. 7C). Thus, once either the factor-dominated or nucleosome-dominated state is established for a gene, that state can further induce the addition or removal of relevant chromatin marks to reenforce its stability, which in essence creates a deeper energy well that raises the barrier of transition to the opposite state. The end result of having these marks is reduced likelihood of genes undergoing spurious, unintended state transition. Indeed, we argue that the bistability genes between nucleosome-dominated state and factor-dominated state, and the role of chromatin marks in modulating this bistability, is the foundation of many chromatin-based epigenetic phenomena. We also note that when chromatin marks are forcibly altered by artificial means, either globally or for specific genes (*8, 27*), the barrier between the two states are reduced, which could potentially allow some genes to undergo state transition.

Our study revealed that PFs such as Sox2 can keep silent genes in somatic cells activatable. It invites the question, do all activatable genes require PFs to stay activatable? We envision two scenarios. In the first one, many genes are “occludable”, meaning that they would undergo occlusion by chromatinization if not protected by their cognate TFs or PFs, but there are also genes that are “unoccludable”, meaning that they would not undergo occlusion even when chromatinized in the absence of factor binding. Unoccludable genes need not require PFs to stay activatable because their cognate TFs can overcome the energy barrier of nucleosomes to activate them. Unoccludable genes could include two classes of genes that need not ever undergo occlusion. One is ubiquitously expressed housekeeping genes. The other is effector genes needed for activation in terminally differentiated cells. As discussed earlier, effectors don’t need to undergo occlusion because they will stay faithfully silent in lineages where they are not needed as long as their upstream MRGs are silenced by occlusion in those lineages. In the second scenario, all the genes in somatic cells require PFs to stay activatable even if they don’t ever need to undergo occlusion. In this case, housekeeping genes don’t require PFs because their TFs are present in all cell types at all times to drive their expression. But effector genes would need PFs to keep them activatable in stem cells whose later differentiation requires these genes to turn on. All considered, being unoccludable would seem to be a more parsimonious solution for effectors.

In summary, out study brings unprecedented clarity to the understanding of lineage restriction, a phenomenon foundational to multicellular lifeforms. As such, our study has potential implications for many fields of biology, such as development, stem cell biology, gene regulation and epigenetics, and may also contribute to the understanding of disease processes such as developmental disorders, cancer and aging.

MATERIALS AND METHODS

Cell culture and fusion

Primary ear and tail fibroblasts were derived from adult SPRET/EiJ and CAST/EiJ mice, respectively, as previously described (*55*). Primary cells were passaged once and transduced with lentivirus expressing simian virus 40 large T antigen (SV40-T) to generate immortalized cells. They were then sorted into 96-well plates as single clones, giving rise to SEF2 and CTF3 clonal lines from SPRET/EiJ and CAST/EiJ, respectively. The mouse EpiSC cell line G14 (aka 1117E3) was derived from an E5.5 F1 embryo of a cross between C57BL/6 male and 129Jae female using published method and culture conditions (*20, 21*). Fluorescence and drug resistance markers were introduced into cells by lentivirus, after which another round of single-clone selection was performed to obtain the final clonal lines for cell fusion.

Cell lines and culture conditions were as described previously for C2C12 (from C3H mouse), B6NSC (C57BL/6 mouse) and, IMG (C57BL/6 mouse) (*56*), E14 (129 mouse) (*7*), ESC(EsrrbKO) (129 mouse) (*25*), and B35 (rat) (*8*). Briefly, SEF2, CTF3, C2C12, IMG and B35 were cultured in DMEM with 10% FBS. E14 and ESC(EsrrbKO) were cultured under feeder-free conditions in Knockout DMEM with 10% FBS, non-essential amino acids, sodium pyruvate, penicillin/streptomycin, β-mercaptoethanol, 3 uM CHIR99021 and 1 uM PD0325091. G14 was cultured on 10% FBS coated plates in DMEM/F12 supplemented with 20% Gibco Knockout Serum Replacement, GlutaMAX™, β-mercaptoethanol, penicillin/streptomycin, 12 ng/mL FGF2, 20 ng/mL ActivinA, 10 uM Y27632 and 2 uM IWP-2. B6NSC was cultured as monolayer in CELLstart substrate coated plates with DMEM/F12 supplemented with N2, B27, GlutaMAX, penicillin/streptomycin, 20 ng/mL FGF2 and 20 ng/mL EGF.

SEF2 or CTF3 were cultured in conditions of their fusion partners before fusion for a week as well as after fusion. The two cell lines to be fused were trypsinized, resuspended in medium, mixed thoroughly at 1:1 ratio, and plated into 6-well plates at high density to enhance cell-cell contact. Cells were settled for 2 hours to allow attachment to the plate, and treated with 45.5% PEG1000 pre-warmed to 42°C. After 1 minute of PEG treatment, cells were washed with fresh medium three times and cultured for 2 days. Hybrid cells were selected by dual drug selection or dual fluorescence-activated cell sorting. Chromosome loss could occur upon fusion, especially when clonal lines are derived from bulk hybrid cells (*8*), and genes on lost chromosomes could be wrongly annotated as occluded. To address this, we quantified strain-specific RNA-Seq reads for each chromosome in hybrid cells, and excluded fusion samples with signs of chromosome loss.

In vitro differentiation

To induce astrocyte differentiation, B6NSC was seeded on poly-D-lysine coated 10 cm dishes at $5x10^5$ cells per plate, and cultured for 7 days with astrocyte differentiation medium containing DMEM/F12, N2, B27, penicillin/streptomycin, GlutaMAX™, 1 ng/mL EGF and 20 ng/mL BMP.

To induce NSC differentiation, SEF2-E14 or SEF2-G14 fusion cells were cultured in E14 or G14 conditions, respectively. One day before differentiation to NSCs, hybrid cells were passaged to allow 60%-70% confluency the next day. Upon differentiation, cells were washed with PBS to

completely eliminate factors supporting ESC or EpiSC growth. Subsequently, hybrid cells were plated on 10 cm dishes coated with CELLstart at $1x10^6$ cells per plate, and cultured in N2B27 media containing DMEM/F12, Neurobasal Medium, N2, B27, GlutaMax, β-mercaptoethanol and 2 uM SB431542 to induce the neural fate. SEF2-E14 and SEF2-G14 fusion cells lost typical pluripotency morphology and acquired NSC morphology at day 5 and day 3, respectively. Cells were then dissociated by TrypLE Express Enzyme, and cultured in low attachment plates with NSC media containing FGF2 and EGF. Successfully differentiated cells would form neurospheres, which was purified by gentle centrifugation. Purified neurospheres were trypsinized and cultured as monolayers in CELLstart coated plates with NSC medium.

RNA-Seq and strain-specific data analysis

Total RNA was extracted by MagNA Pure Compact RNA Isolation kit. Following DNase treatment, mRNA with polyA tail was purified with NEBNext poly(A) mRNA Magnetic Isolation module. Purified mRNA was reverse transcribed and the resulting cDNA made into libraries using Illumina primer sets following vendor's protocol.

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 fusion partner genomes were replaced with the ambiguity base 'N'. SNPsplit was used to extract reads specific to each strain. Transcripts per million (TPM) for each fusion partner was calculated based on the relative amounts of strain-specific reads. Genes in a given fusion partner were defined by pre- and post-fusion expression levels as activatable (pre-fusion TPM < 1, prefusion partner TPM ≥ 2, post-fusion TPM ≥ 30% of total post-fusion TPM of both genomes, average post-fusion TPM per genome ≥ 2), occluded (pre-fusion TPM < 1, prefusion partner TPM ≥ 2, post-fusion TPM < 10% of total post-fusion TPM of both genomes, average post-fusion TPM per genome ≥ 2), or extinguished (pre-fusion TPM < 1, prefusion partner TPM ≥ 2, average post-fusion TPM per genome < 2).

Metagene analysis of SEF2 occluded genes in hybrid cells during differentiation

Strain-specific gene expression was calculated for SEF2-E14 or SEF2-G14 fusion cells before and after differentiation. Total TPM, calculated from all reads from both partners of hybrid cells, were used to select for activated genes (total TPM post-differentiation >= 4, fold change > 4) during NSC differentiation. For each gene, expression from the E14 or G14 genome post differentiation was scaled to unit. Expression from the E14 or G14 genome pre-differentiation, as well as expression from the SEF2 genome pre- and post-differentiation, was scaled to this unit.

Chromatin assembly

Chromatinization was performed by Chromatin Assembly Kit from Active Motif (Cat #: 53500). The HeLa core histones in the kit were replaced with equal amounts of recombinant histones from NEB (Cat#: M2508S and M2509S) for recombinant histone samples. As a modification of the manual that improved chromatin assembly efficiency, high salt buffer and low salt buffer were mixed together during the incubation step of h-NAP-1 and core histones. Following chromatin reconstitution, 5 mM $MgCl_2$ was added to the solution and centrifuged for 15 minutes to precipitate the assembled chromatin. This removed unchromatinized or partially chromatinized DNA. The pellet was resuspended with Gene Pulser electroporation buffer with 1 mM EDTA. Undissolved chromatin pellet was removed by centrifugation before electroporation.

Partial digestion assay

Following purification of assembled chromatin, 3 ul of 0.1 M $CaCl_2$ was added to 100 ul 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 adding 34 ul 4X Enzymatic Stop Solution in the Chromatin Assembly Kit. Following purification of digested DNA by QIAquick PCR Purification Kit from Qiagen, gel electrophoresis was used to visualize the pattern of nucleosome array.

Cell transfection with naked or chromatinized DNA and measurement of gene expression

293T cells were transfected by electroporation. Cells were trypsinized and counted. One million cells were resuspended in Gene Pulser Electroporation Buffer containing 1 mM EDTA and 1 ug naked or chromatinized plasmid DNA (see below for plasmid ID used). 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 visualized by fluorescence microscopy. To assay for promoter strength, electroporated cells were separated into two batches, one to extract mRNA for RT-qPCR quantification of reporter expression, and the other to isolate nuclear plasmid DNA for qPCR quantification. The mRNA quantity was normalized to the plasmid DNA quantity to obtain the final, normalized reporter expression levels. Chemical transfection (i.e., by lipofectamine) was not used due to the possibility that the chemicals used could disrupt DNA-histone association (*57*).

CRISPR-mediated gene knockout

Guide RNAs were designed at 5' and 3' UTRs of Sox2, Olig2 and Hmgb2 genes to delete their full coding sequences. CRISPR plasmids were then constructed, each expressing an sgRNA pair corresponding the 5' and 3' UTR targets of a gene, plus Cas9 and blasticidin drug resistance (see below for plasmid ID used). Following transfection of plasmid DNA, B6NSC was selected by 30 ug/ml blasticidin for 2 days to enrich for cells that took in functional copies of the plasmid, and then sorted into CELLstart coated 96-well plates to derive clonal lines. The resulting clones were genotyped with primers across the two target sites to screen for successful deletion, and primers within the deleted fragment to screen for homogenous knockout. Knockout clones were further validated by RNA-Seq data, wherein the coding sequences of target genes were devoid of reads.

Rescue of Sox2 knockout

A lentiviral vector containing the PGK promoter driving Sox2 as well as a neomycin resistance gene were created and packaged into virus by VectorBuilder (see below for plasmid ID used). Sox2 knockout B6NSC clones were transduced at MOI of 5 and selected by G418 for 7 days. The rescue of Sox2 expression was confirmed by RNA-seq data, wherein the read coverage of Sox2 coding sequence was recovered.

Plasmids and lentiviral vectors

Plasmids/vectors for chromatinization assay: VB171220-1258gxn, VB200911-1183yfq, VB210716-1111vfu and VB210719-1129pad; half-chromatinization assay: VB220428-1064cuh, VB221020-1032pzd, and VB220428-1062bsm; CRISPR-mediated knockout of Sox2, Olig2 and Hmgb2, respectively: VB220823-1355ufr, VB220823-1356hth and VB220823-1358fhs; lentiviral

labelling of cells: VB150915-10026 (EGFP and puromycin resistance); VB150925-10020 (mCherry and hygromycin resistance); lentiviral expression of SV40-T, Esrrb, Klf4 and Sox2, respectively: VB171106-1316rqy, VB180510-1202zrv, VB181219-1169xkg, VB230911-1130rye. All plasmids were constructed and lentiviruses packaged by VectorBuilder, and their maps and sequences can be retrieved by the above vector IDs at https://vectorbuilder.com by following menu link “Design Vector”, then “Retrieve Vector Information”. For the half-chromatinization assay, VB220428-1064cuh or VB221020-1032pzd (with piggyBac ITRs) and VB220428-1062bsm were digested by BbsI to create the mCherry fragment and the Turbo-GFP fragment with compatible sticky ends, which was subsequently gel-purified for the chromatin assembly assay.

CUT&Tag

CUT&Tag was performed with NovoNGS CUT&Tag 3.0 High-Sensitivity Kit for Illumina from Novoprotein following vendor’s instructions. Both monoclonal antibody from Cell Signaling (Cat #: 23064) and polyclonal antibody from Abcam (Cat #: ab97959) against Sox2 were used.

ACKNOWLEDGMENTS

We thank Marcelo Nobrega, Heng-Chi Lee, Alex Ruthenburg and Guohong Li for scientific advice, Marcelo Nobrega for administrative support, and Austin Smith and Graziano Martello for providing the Esrrb knockout ES cells. Data on Esrrb were also posted in a previous bioRxiv preprint (DOI: 10.1101/2021.05.04.442547) not yet published in a peer-reviewed journal. This work was funded by VectorBuilder, Frontier Explorer Foundation, Department of Human Genetics at the University of Chicago, and Chicago Biomedical Consortium with support from Searle Funds at The Chicago Community Trust.

AUTHOR CONTRIBUTIONS

Supervision, formulation of direction and acquisition of funding: BTL; Conceptualization and methodology: BW, JHL, KMF, BTL; Investigation, data analysis and visualization: BW, JHL, KMF, LZ, CJF, BG, XD, BXX, CZZ, GF, BTL.

COMPETING INTERESTS

The authors declare no competing interest.

DATA AND MATERIALS AVAILABILITY

All data and materials are deposited or available upon request.

REFERENCES

1. C. H. Waddington, *The Strategy of the Genes*. (George Allen & Unwin, London, 1957).
2. C. H. Waddington, *Principles of Development and Differentiation*. (Macmillan, New York, 1966).
3. A. G. Fisher, Cellular identity and lineage choice. *Nat Rev Immunol* **2**, 977-982 (2002).
4. J. H. Lee *et al.*, Systematic identification of cis-silenced genes by trans complementation. *Human molecular genetics* **18**, 835-846 (2009).

5. B. T. Lahn, The "occlusis" model of cell fate restriction. *Bioessays* **33**, 13-20 (2011).
6. J. Gaetz *et al.*, Evidence for a critical role of gene occlusion in cell fate restriction. *Cell research* **22**, 848-858 (2012).
7. K. M. Foshay *et al.*, Embryonic stem cells induce pluripotency in somatic cell fusion through biphasic reprogramming. *Mol Cell* **46**, 159-170 (2012).
8. T. J. Looney *et al.*, Systematic mapping of occluded genes by cell fusion reveals prevalence and stability of cis-mediated silencing in somatic cells. *Genome research* **24**, 267-280 (2014).
9. M. Boshart, D. Nitsch, G. Schutz, Extinction of gene expression in somatic cell hybrids--a reflection of important regulatory mechanisms? *Trends Genet* **9**, 240-245 (1993).
10. J. Nichols, A. Smith, Naive and primed pluripotent states. *Cell stem cell* **4**, 487-492 (2009).
11. C. S. Lee, J. R. Friedman, J. T. Fulmer, K. H. Kaestner, The initiation of liver development is dependent on Foxa transcription factors. *Nature* **435**, 944-947 (2005).
12. A. Wang *et al.*, Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell stem cell* **16**, 386-399 (2015).
13. A. Mayran *et al.*, Pioneer and nonpioneer factor cooperation drives lineage specific chromatin opening. *Nat Commun* **10**, 3807 (2019).
14. M. B. Gerstein *et al.*, Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91-100 (2012).
15. T. Graf, T. Enver, Forcing cells to change lineages. *Nature* **462**, 587-594 (2009).
16. H. M. Blau, How fixed is the differentiated state? Lessons from heterokaryons. *Trends Genet* **5**, 268-272 (1989).
17. U. Lakshmipathy, C. Verfaillie, Stem cell plasticity. *Blood Rev* **19**, 29-38 (2005).
18. T. Braun, G. Buschhausen-Denker, E. Bober, E. Tannich, H. H. Arnold, A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. *Embo J* **8**, 701-709 (1989).
19. A. Smith, Formative pluripotency: the executive phase in a developmental continuum. *Development (Cambridge, England)* **144**, 365-373 (2017).
20. I. G. Brons *et al.*, Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191-195 (2007).
21. P. J. Tesar *et al.*, New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196-199 (2007).
22. B. Feng *et al.*, Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat Cell Biol* **11**, 197-203 (2009).
23. G. Guo *et al.*, Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development (Cambridge, England)* **136**, 1063-1069 (2009).
24. Y. Buganim *et al.*, The developmental potential of iPSCs is greatly influenced by reprogramming factor selection. *Cell stem cell* **15**, 295-309 (2014).
25. G. Martello *et al.*, Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell stem cell* **11**, 491-504 (2012).
26. I. Flores, R. Benetti, M. A. Blasco, Telomerase regulation and stem cell behaviour. *Current opinion in cell biology* **18**, 254-260 (2006).
27. L. Bintu *et al.*, Dynamics of epigenetic regulation at the single-cell level. *Science (New York, N.Y* **351**, 720-724 (2016).
28. D. Reinberg, L. D. Vales, Chromatin domains rich in inheritance. *Science (New York, N.Y* **361**, 33-34 (2018).
29. J. H. Lee *et al.*, Chromatin analysis of occluded genes. *Human molecular genetics* **18**, 2567-2574 (2009).
30. P. J. Laybourn, J. T. Kadonaga, Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. *Science (New York, N.Y* **254**, 238-245 (1991).

31. P. B. Becker, C. Wu, Cell-free system for assembly of transcriptionally repressed chromatin from Drosophila embryos. *Mol Cell Biol* **12**, 2241-2249 (1992).
32. C. R. Clapier, J. Iwasa, B. R. Cairns, C. L. Peterson, Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* **18**, 407-422 (2017).
33. S. Cereghini, M. Yaniv, Assembly of transfected DNA into chromatin: structural changes in the origin-promoter-enhancer region upon replication. *EMBO J* **3**, 1243-1253 (1984).
34. S. C. Biddie *et al.*, Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol Cell* **43**, 145-155 (2011).
35. R. C. Adam *et al.*, NFI transcription factors provide chromatin access to maintain stem cell identity while preventing unintended lineage fate choices. *Nat Cell Biol* **22**, 640-650 (2020).
36. M. Lattke *et al.*, Extensive transcriptional and chromatin changes underlie astrocyte maturation in vivo and in culture. *Nat Commun* **12**, 4335 (2021).
37. L. H. Pevny, S. K. Nicolis, Sox2 roles in neural stem cells. *Int J Biochem Cell Biol* **42**, 421-424 (2010).
38. A. Amador-Arjona *et al.*, SOX2 primes the epigenetic landscape in neural precursors enabling proper gene activation during hippocampal neurogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E1936-1945 (2015).
39. J. L. Mateo *et al.*, Characterization of the neural stem cell gene regulatory network identifies OLIG2 as a multifunctional regulator of self-renewal. *Genome research* **25**, 41-56 (2015).
40. A. B. Abraham *et al.*, Aberrant neural stem cell proliferation and increased adult neurogenesis in mice lacking chromatin protein HMGB2. *PLoS One* **8**, e84838 (2013).
41. R. Bronstein, J. Kyle, A. B. Abraham, S. E. Tsirka, Neurogenic to Gliogenic Fate Transition Perturbed by Loss of HMGB2. *Front Mol Neurosci* **10**, 153 (2017).
42. H. S. Kaya-Okur, D. H. Janssens, J. G. Henikoff, K. Ahmad, S. Henikoff, Efficient low-cost chromatin profiling with CUT&Tag. *Nat Protoc* **15**, 3264-3283 (2020).
43. B. E. Bernstein *et al.*, A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326 (2006).
44. M. A. Lodato *et al.*, SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. *PLoS genetics* **9**, e1003288 (2013).
45. M. Bergsland *et al.*, Sequentially acting Sox transcription factors in neural lineage development. *Genes & development* **25**, 2453-2464 (2011).
46. M. Ramalho-Santos, S. Yoon, Y. Matsuzaki, R. C. Mulligan, D. A. Melton, "Stemness": transcriptional profiling of embryonic and adult stem cells. *Science (New York, N.Y* **298**, 597-600 (2002).
47. N. B. Ivanova *et al.*, A stem cell molecular signature. *Science (New York, N.Y* **298**, 601-604 (2002).
48. J. Cai, M. L. Weiss, M. S. Rao, In search of "stemness". *Exp Hematol* **32**, 585-598 (2004).
49. A. Avgustinova, S. A. Benitah, Epigenetic control of adult stem cell function. *Nat Rev Mol Cell Biol* **17**, 643-658 (2016).
50. S. H. Stricker, A. Koferle, S. Beck, From profiles to function in epigenomics. *Nature reviews* **18**, 51-66 (2017).
51. Y. Zeng *et al.*, Lin28A Binds Active Promoters and Recruits Tet1 to Regulate Gene Expression. *Mol Cell* **61**, 153-160 (2016).
52. K. M. Jozwik, I. Chernukhin, A. A. Serandour, S. Nagarajan, J. S. Carroll, FOXA1 Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3. *Cell Rep* **17**, 2715-2723 (2016).
53. E. M. Riising *et al.*, Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol Cell* **55**, 347-360 (2014).

54. M. Hosogane, R. Funayama, M. Shirota, K. Nakayama, Lack of Transcription Triggers H3K27me3 Accumulation in the Gene Body. *Cell Rep* **16**, 696-706 (2016).
55. M. Khan, S. Gasser, Generating Primary Fibroblast Cultures from Mouse Ear and Tail Tissues. *J Vis Exp*, (2016).
56. R. C. McCarthy *et al.*, Characterization of a novel adult murine immortalized microglial cell line and its activation by amyloid-beta. *J Neuroinflammation* **13**, 21 (2016).
57. D. Lundberg *et al.*, Interactions between cationic lipid bilayers and model chromatin. *Langmuir* **26**, 12488-12492 (2010).

Figure 1

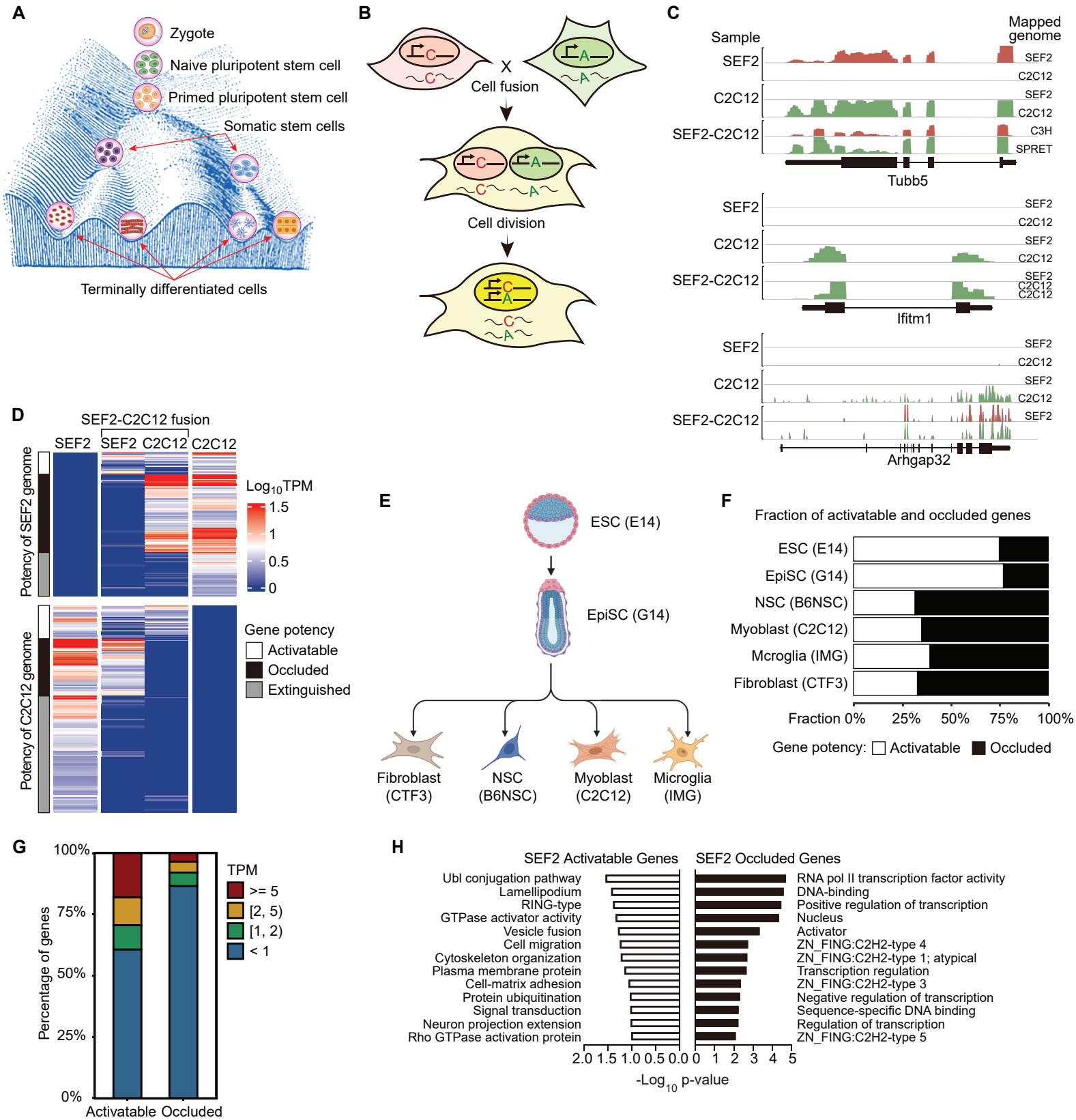


Figure 1. Restriction of cell lineage potency is associated with reduction in genome transcriptional potency. (**A**) Adaptation of Waddington's classic illustration metaphorizing lineage restriction (ref. 1), with balls representing cells in different stages of development. It shows that differentiation is irreversible (the ball can only roll downhill), is accompanied by a progressive reduction of fate potency (possible paths remaining for the ball become progressively limited), and eventually reaches a stable differentiated state (the ball comes to a rest in one of the valleys at the bottom of the hill). (**B**) Schematic of the cell fusion assay. (**C**) Strain-specific RNA-Seq profiles of representative constitutively expressed gene Tubb5, activatable gene Arhgap32, and occluded gene Ifitm1 in SEF2, before and after cell fusion. Only reads that can be mapped specifically to either the SEF2 or C2C12 genome based on inter-strain polymorphisms were used to build the profiles. (**D**) Gene expression heatmaps of unfused SEF2 (n = 303 genes) and C2C12 (n = 435), as well as fused cells. Shown are SEF2 or C2C12 genes classified as activatable, occluded or extinguished, based on their expression patterns pre- and post-fusion. (**E**) Cell lines of varying developmental potency used in the cell fusion assay. (**F**) Genome potency as reflected by the fraction of activatable and occluded genes in various cell lines. Genome transcriptional potency of G14 (n = 427), NSC (n = 426), C2C12 (n = 190), and IMG (n = 652) was measured by fusion with SEF2. Genome potency of E14 (n = 362) and CTF3 (n = 110) was measured by fusion with R1A (day 8 post-fusion) and C2C12, respectively. (**G**) Percentage of B6NSC occluded (n = 198) or activatable (n = 86) genes falling into high (TPM ≥ 5), medium (2 ≤ TPM < 5), low (1 ≤ TPM < 2) or no (TPM < 1) expression range in astrocytes differentiated from B6NSCs. (**H**) GO terms enriched in SEF2 occluded and activatable genes identified in SEF2-NSC fusion.

Figure 2

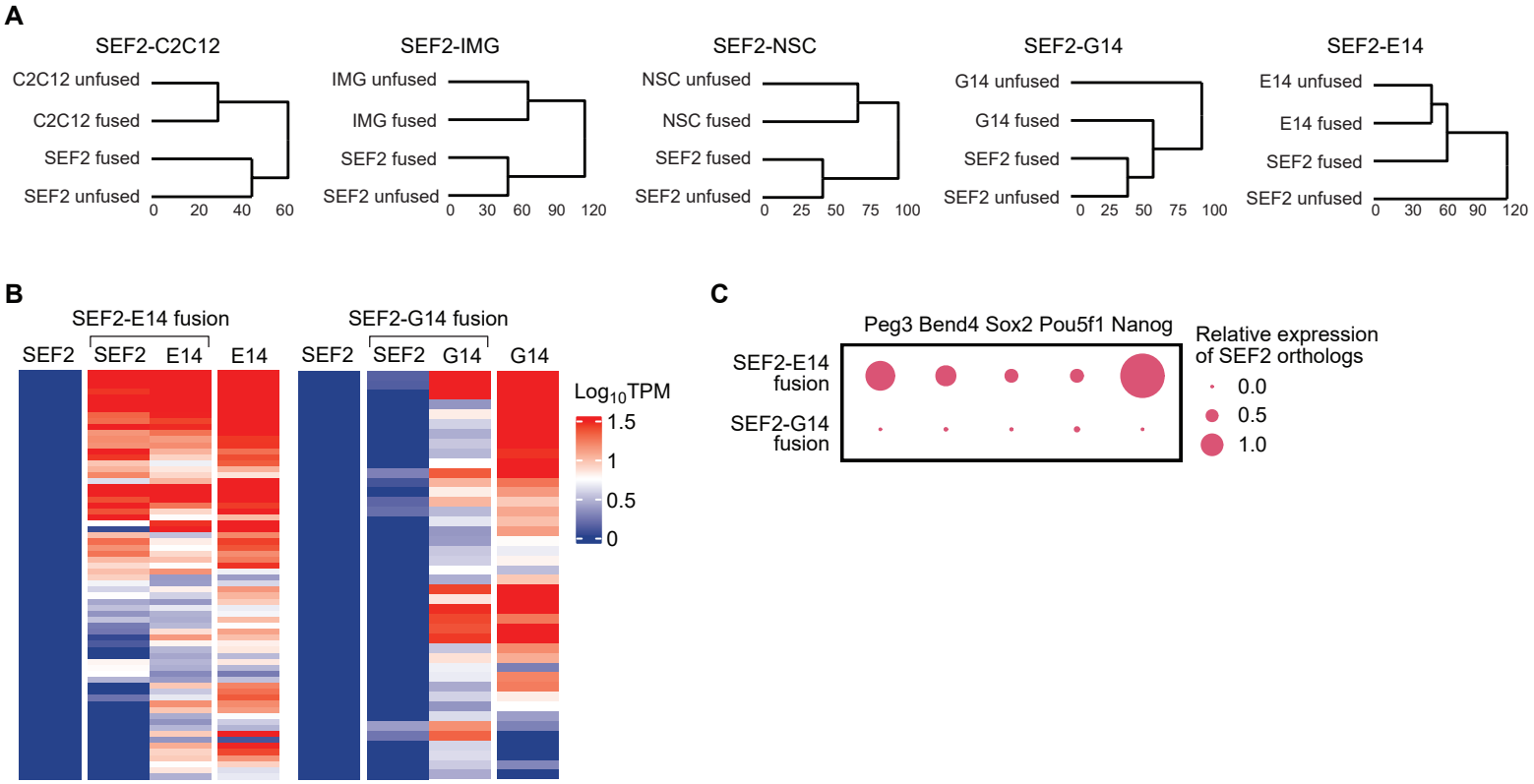


Figure 2. Deocclusion ability is present in naive pluripotency but lost in primed pluripotency. (**A**) Hierarchical clustering of pre- and post-fusion samples based on the expression patterns of genes differentially expressed between pre-fusion partner cells (TPM >= 2; $\log_2$ fold change > 2; SEF2-C2C12, n = 1308 genes; SEF2-IMG, n = 3164; SEF2-NSC, n = 2704; SEF2-G14, n = 3055; SEF2-E14, n = 3434). (**B**) Expression heatmap of SEF2 occluded genes with active orthologs in the pre-fusion partner cells. Depicted are SEF2-E14 (n = 68) and SEF2-G14 (n = 42) fusions (**C**) Relative expression of representative SEF2 occluded genes and pluripotency genes in SEF2-E14 and SEF2-G14 fusions. Peg3, Bend4 and Sox2 are identified as SEF2 occluded genes by SEF2-C2C12, SEF2-IMG and SEF2-B6NSC fusion, respectively. Expression levels from the SEF2 orthologs are scaled to that of their E14 or G14 fusion partner.

Figure 3

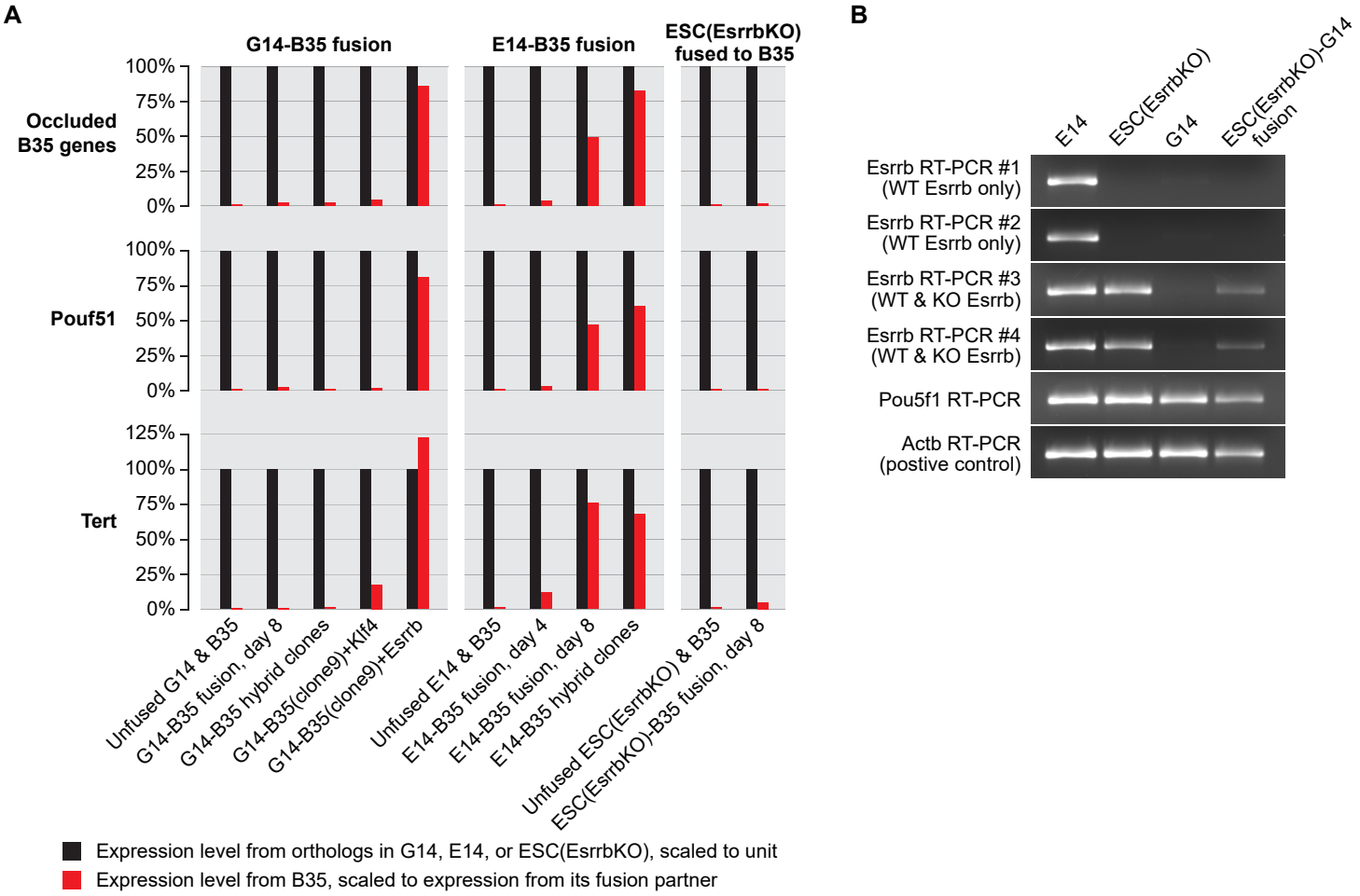


Fig 3. Deocclusion capacity is present in naive pluripotency, but abolished in primed pluripotency due to silencing of Esrrb by occlusion. (**A**) Expression levels of occluded B35 genes and their orthologs in fusions between B35 and various pluripotent cell types. (**B**) Cell fusion assay showing occluded status of Esrrb in G14. Esrrb expression was interrogated by four RT-PCR amplicons, two of which can only amplify the wildtype (WT) Esrrb transcripts, while the other two can amplify both WT and knockout (KO) Esrrb transcripts.

Figure 4

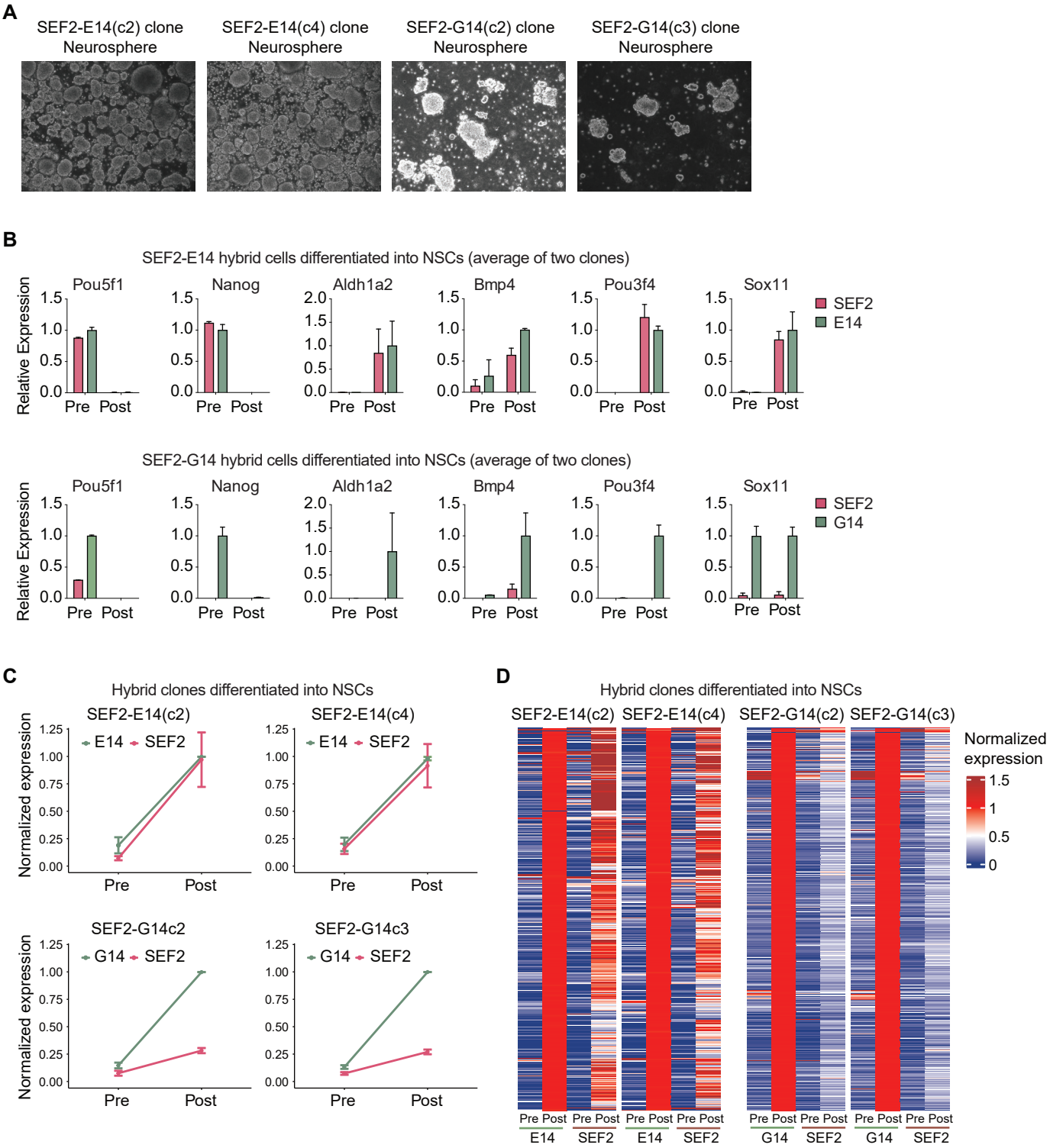


Figure 4. Occluded SEF2 genes in SEF2-E14 but not SEF2-G14 hybrid cells possess the potency for proper activation in response to NSC differentiation. (**A**) Neurospheres successfully generated from the differentiation of four clonal hybrid lines, SEF2-E14(c2), SEF2-E14(c4), SEF2-G14(c2), and SEF2-G14(c3). (**B**) Expression of pluripotency genes and NSC markers before and after the differentiation of SEF2-E14 and SEF2-G14 hybrid clones into NSCs (2 hybrid clones). (**C**) Average expression of occluded SEF2 genes whose orthologs were upregulated in the E14 or G14 genome of SEF2-E14 (n = 63 genes) and SEF2-G14 (n = 41) hybrid clones upon differentiation into NSCs. Expression level of each gene is scaled to the post-differentiation expression level of E14 or G14. (**D**) Expression heatmaps of upregulated genes (TPM >= 2, $\log_2$ fold change > 2, n = 889) in SEF2-E14 and SEF2-G14 hybrid clones following differentiation into NSCs. Expression level of each gene is scaled to the post-differentiation expression level of E14 or G14.

Figure 5

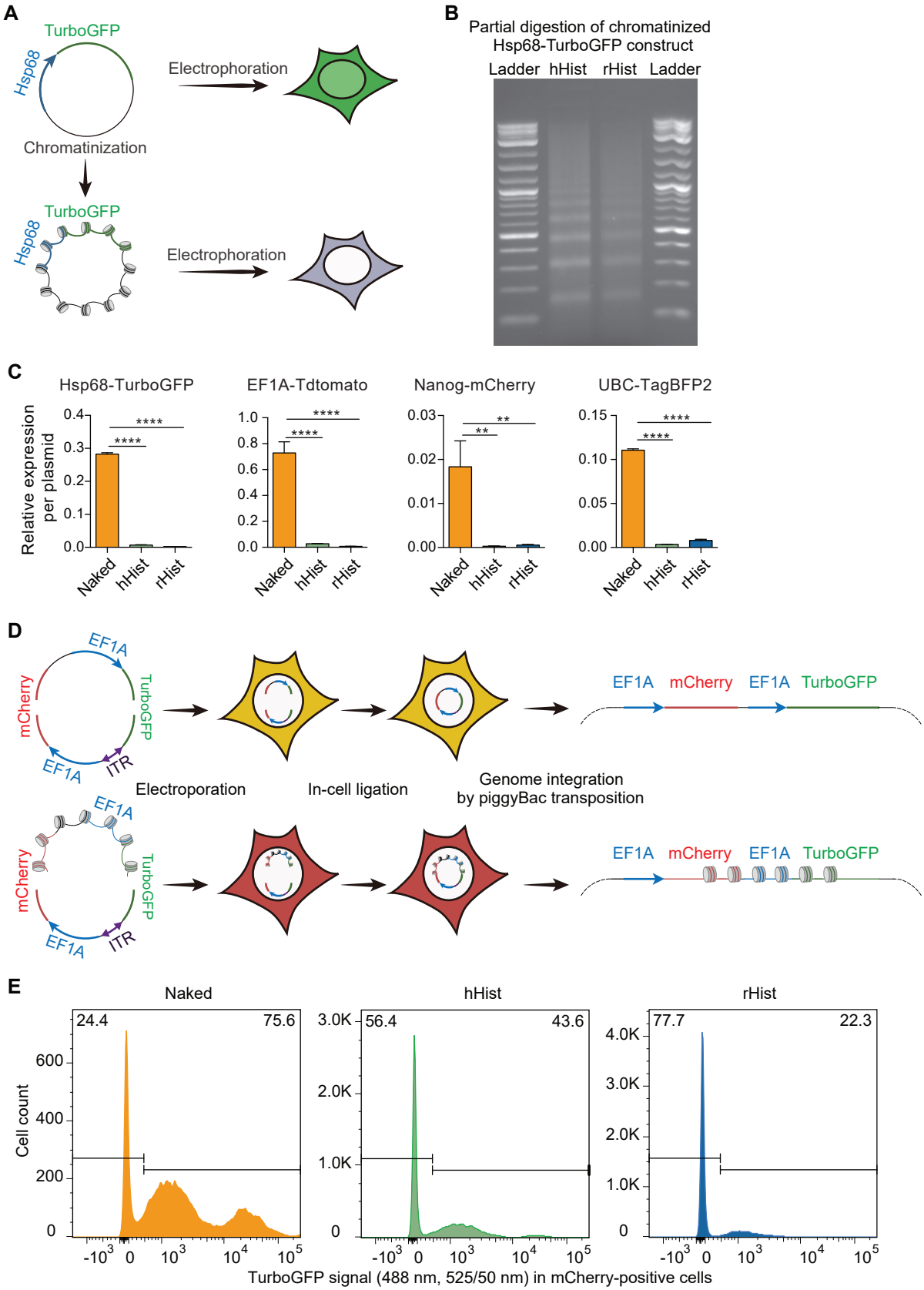


Figure 5. Chromatinized DNA undergoes occlusion by default and is heritable through DNA replication. (**A**) Schematic of the chromatinization assay to measure transcriptional potency of naked or in vitro chromatinized DNA construct, using the Hsp68-TurboGFP construct as an example (PolyA downstream of reporter not shown). (**B**) Representative partial digestion by MNase of plasmid DNA construct chromatinized with HeLa core histones (hHist) or recombinant core histones (rHist). (**C**) Expression of naked and chromatinized constructs driven by different promoters in 293T cells. Reporter expression was measured by RT-qPCR and normalized to the amount of plasmid DNA extracted from cell nuclei as measured by qPCR, then scaled to endogenous Actb expression (3 replicates, ordinary one-way ANOVA test). (**D**) Schematic of the assay involving half-chromatinization followed by genome integration (PolyA downstream of reporter not shown). Expression cassettes of TurboGFP and mCherry were split within their coding sequences. The TurboGFP fragment in either naked or chromatinized form was co-electroporated into 293T cells with the naked mCherry fragment. The two fragments ligated spontaneously after entering cells to reconstitute a circular construct with complete TurboGFP and mCherry expression cassettes. The piggyBac ITRs in the naked mCherry fragment enabled the reconstituted construct to integrate into the host genome by piggyBac transposition. (**E**) Histograms of TurboGFP fluorescence in mCherry-positive cells.

Figure 6

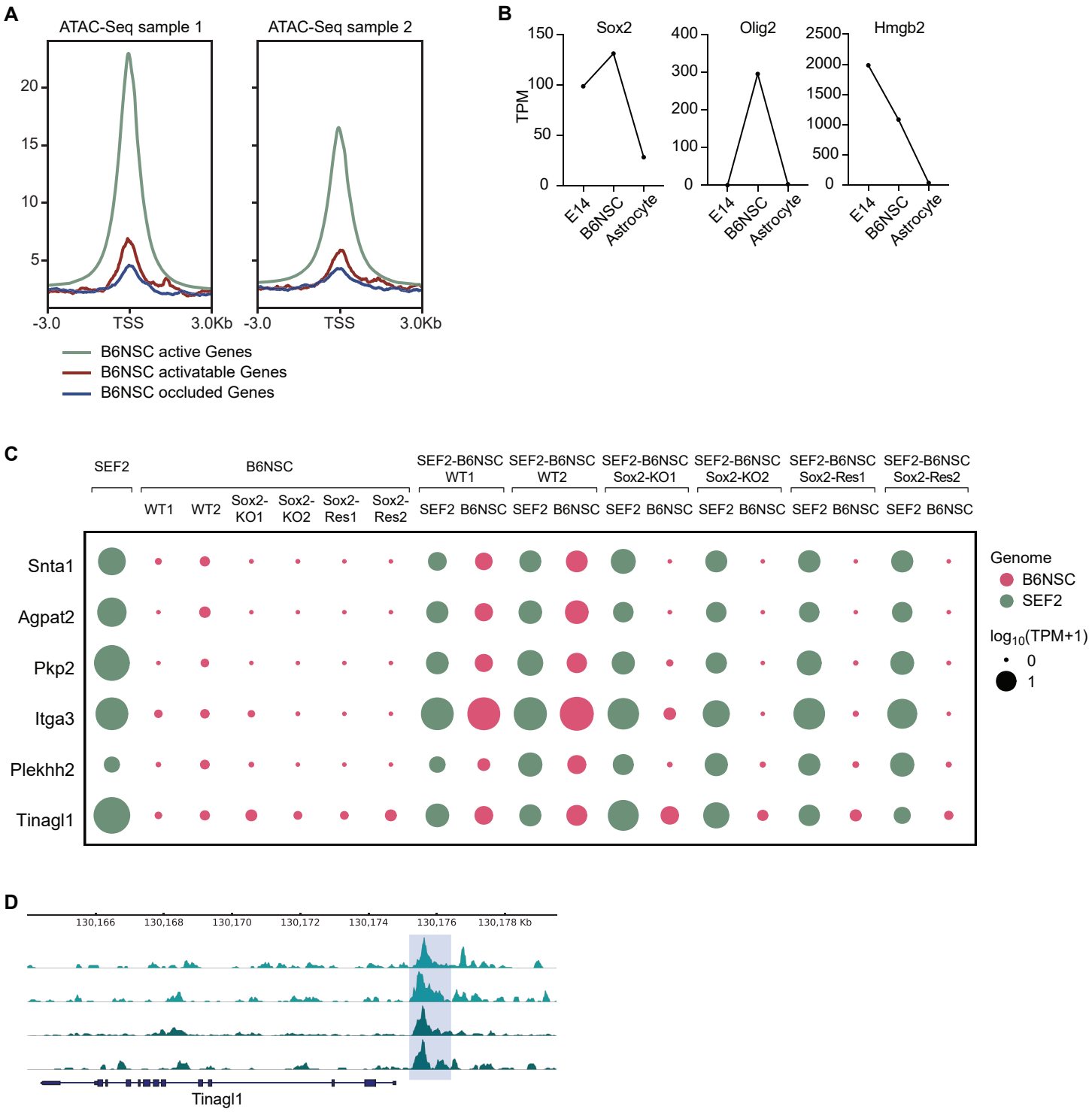


Figure 6. Sox2 acts as a PF to protect target genes from occlusion in NSCs. (**A**) ATAC-Seq profiles corresponding to activatable (n = 85) and occluded (n = 186) genes in B6NSC. (**B**) Expression of Sox2, Olig2 and Hmgb2 in E14, B6NSC, and B6NSC-differentiated astrocytes. (**C**) Bubble plot depicting expression levels in pre- and post-fusion SEF2 and B6NSC for activatable genes in wildtype B6NSC that became occluded following Sox2 knockout. Two B6NSC clones homozygous for Sox2 deletion were fused to SEF2 either directly (Sox2-KO1 and Sox2-KO2), or rescued for Sox2 expression by a lentivirus vector expression Sox2 before cell fusion (Sox2-Res1 and Sox2-Res2). Two wildtype B6NSC clones (WT1 and WT2) were used as control. (D) Sox2 binding profiles of a representative B6NSC activatable gene that became occluded following Sox2 knockout. Sox2 binding was measured by CUT&Tag with a monoclonal antibody (top two tracks for two replicates) and a polyclonal antibody (bottom two tracks for two replicates).

Figure 7

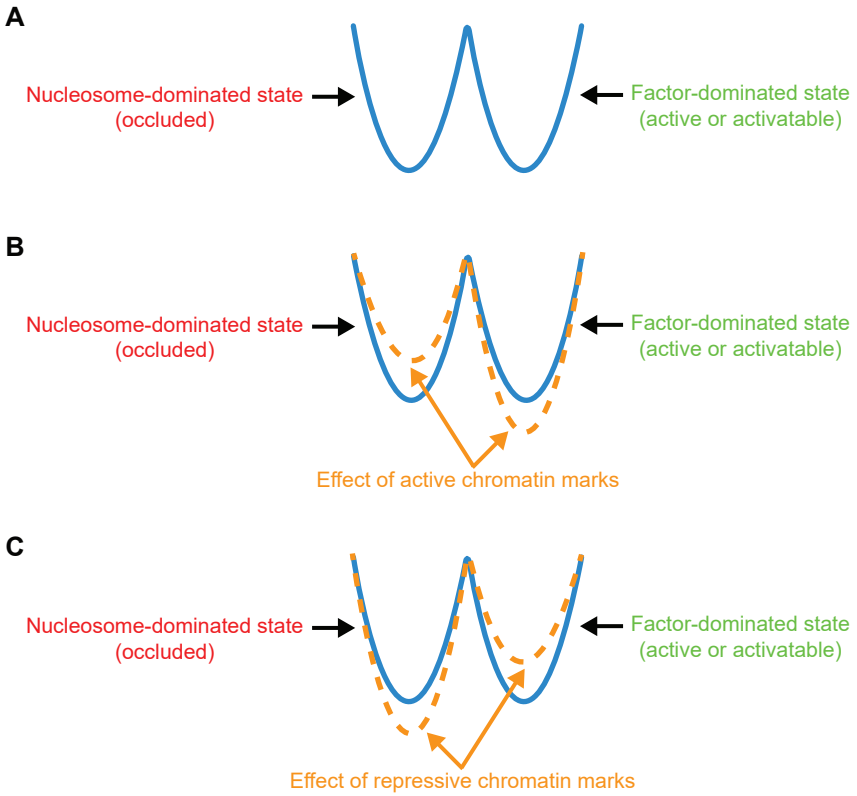


Figure 7. Bistability of genes and the modulating effect of chromatin marks on the bistability. (**A**) Genes can reside in one of two energetically stable states (represented by two energy wells) with a high energy barrier that prevents easy transition between them. In the nucleosome-dominate state, genes are occluded. In the factor-dominated state, genes are either active (if the factors are TFs) or activatable (if the factors are PFs). (**B**) Active chromatin marks increase the stability of factor-dominated state while decrease the stability of nucleosome-dominated state. (**C**) Repressive chromatin marks decrease the stability of factor-dominated state while increase the stability of nucleosome-dominated state.

Figure S1

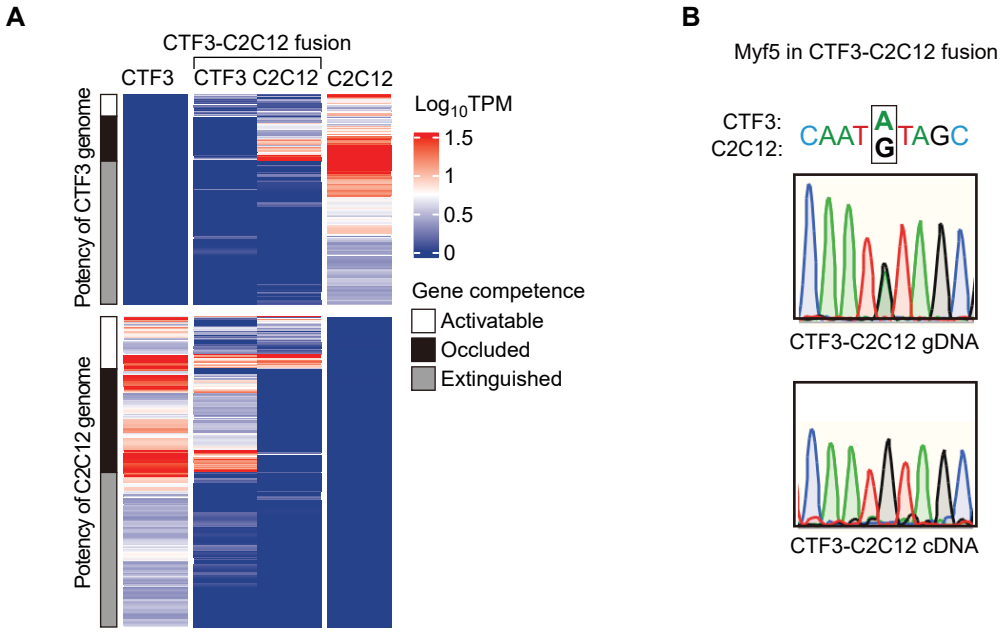
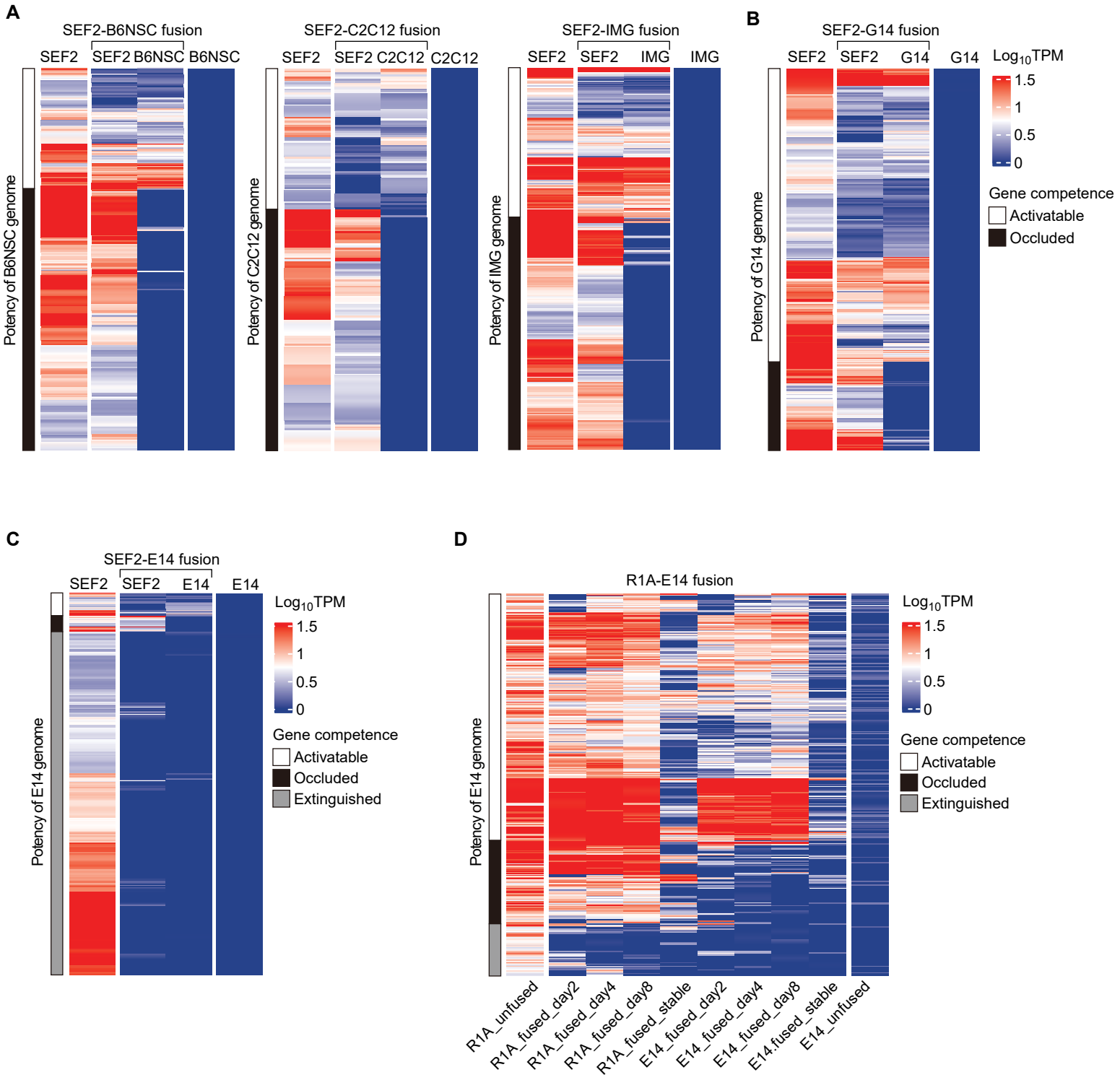


Figure S1. Strain-specific RNA-Seq of fusion cells provides a measurement of gene potency. (**A**) Heatmap of informative silent gene (n = 858) expression in unfused CTF3 and C2C12 cells, as well as strain-specific expression in CTF3-C2C12 fusion cells. Activatable, occluded and extinguished genes are annotated based on their expression before and after cell fusion. (**B**) Sanger sequencing of RT-PCR (CTF3-C2C12 cDNA) and genomic DNA PCR (CTF3-C2C12 gDNA) products of Myf5, an occluded gene in CTF3, in CTF3-C2C12 fusion cells. Relative abundance of Myf5 genomic DNA and mRNA in hybrid cells are reflected by the signal at the SNP site.

Figure S2



Figrue S2. Gene potency of ESCs can be measured during early but not late stages of cell fusion. (**A**)(**B**) Heatmap of informative silent gene expression in NSC (n = 426), C2C12 (n = 190), IMG (n = 652) and G14 (n = 427) unfused cells and their fusion cells with SEF2. (**C**) Heatmap of E14 informative silent gene (n = 788) expression in unfused SEF2 and E14 cells, as well as strain-specific expression in SEF2-E14 fusion cells. (**D**) Time-course expression of E14 informative silent genes (n = 417) in R1A, E14 and R1A-E14 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.

Figure S3

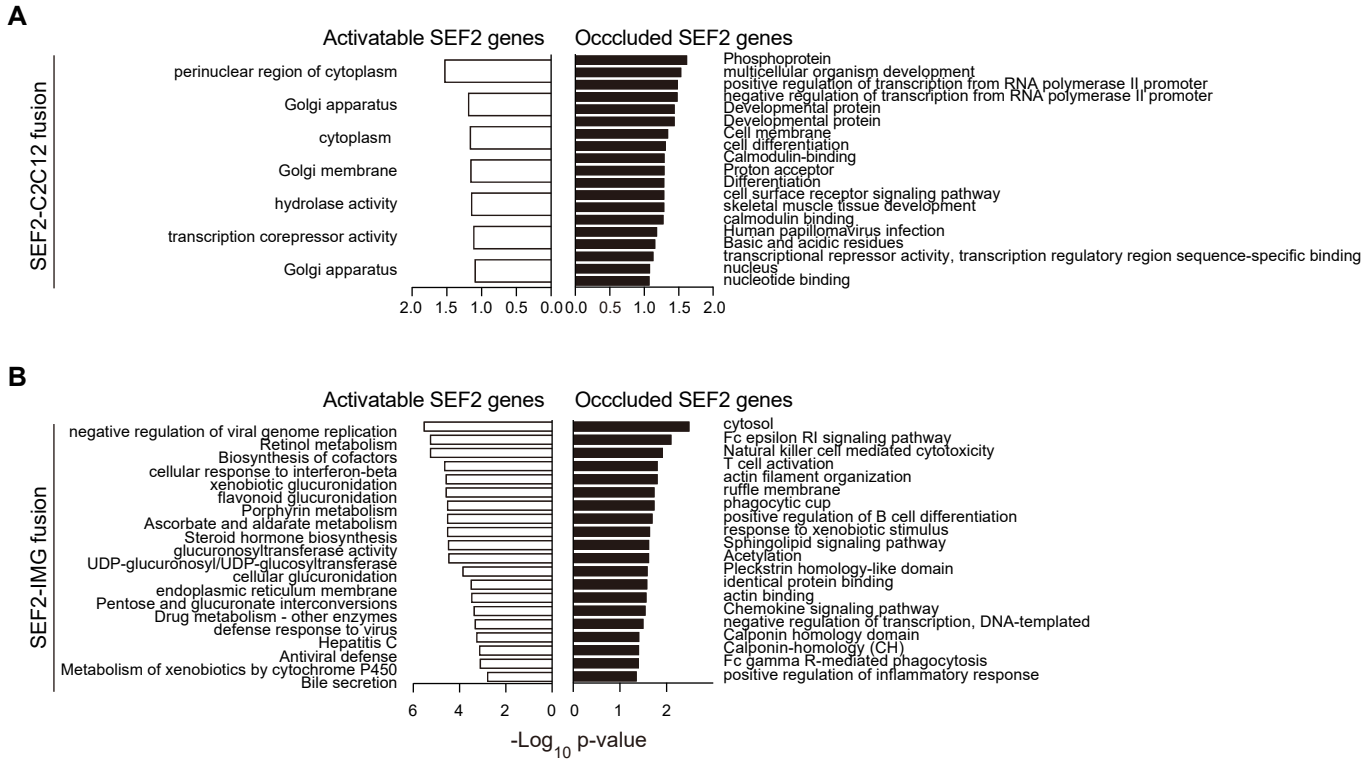


Figure S3. SEF2 occluded genes are enriched for regulatory functions. GO terms enriched in SEF2 activatable and occluded genes identified in SEF2-C2C12 (A) and SEF2-IMG (B) fusions.

Figure S4

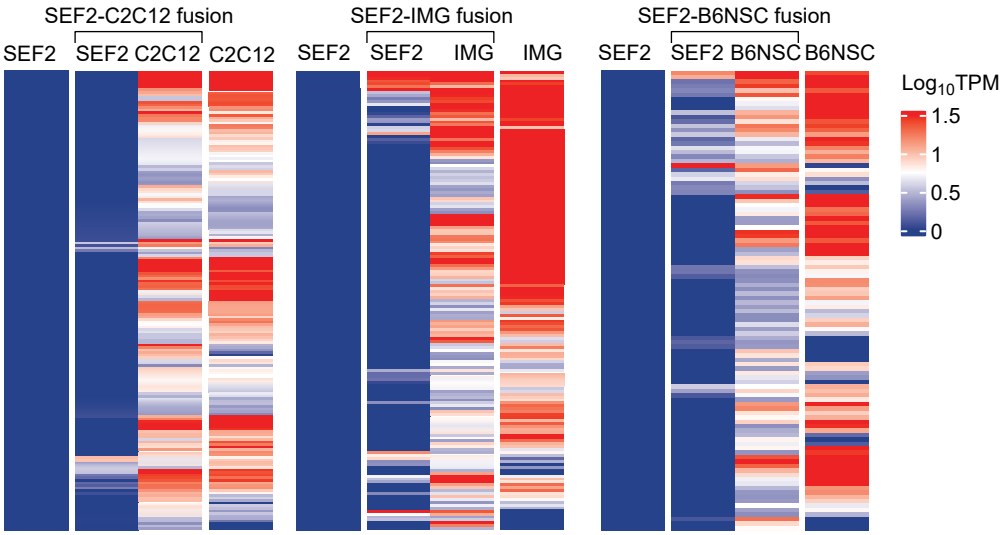


Figure S4. Expression heatmaps of SEF2 occluded genes in various somatic-somatic fusions. SEF2 genes shown to be occluded in SEF2-C2C12 (n = 180), SEF2-IMG (= 157) and SEF2-B6NSC (n = 104) fusions are combined, with non-extinguished genes in each fusion sample displayed.

Figure S5

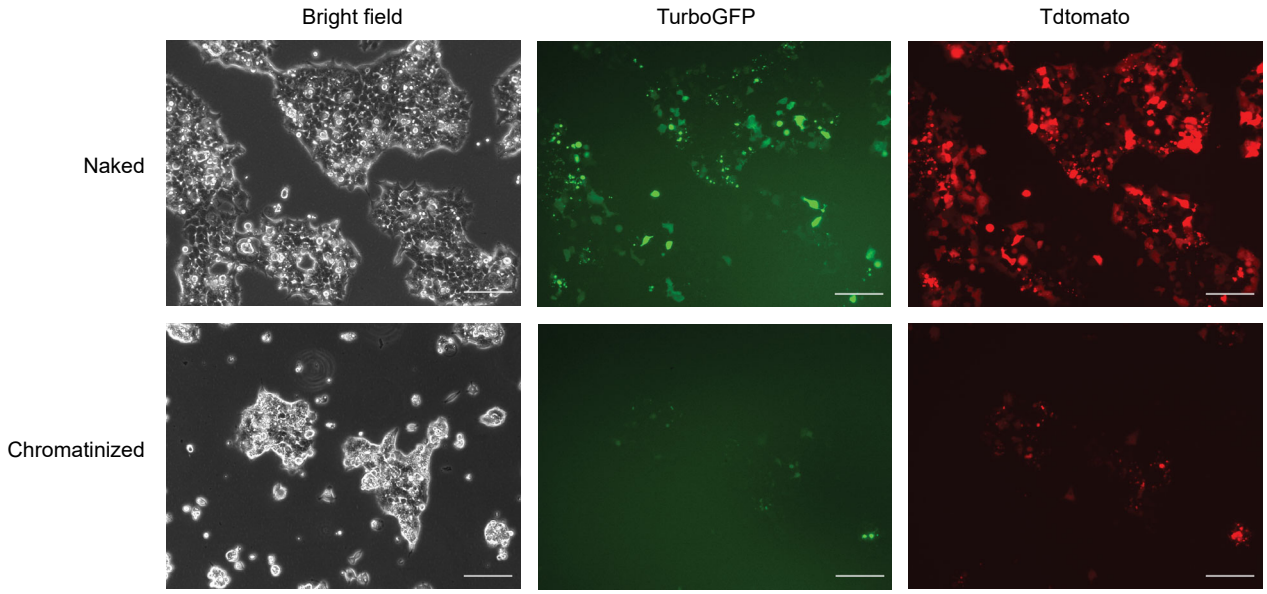


Figure S5. Chromatinized DNA undergoes occlusion. Fluorescence images of 293T cells electorporated with naked or chromatinized plasmids containing Hsp68 driving TurboGFP or EF1A driving Tdtomato. Scale bar: 200 um.

Figure S6

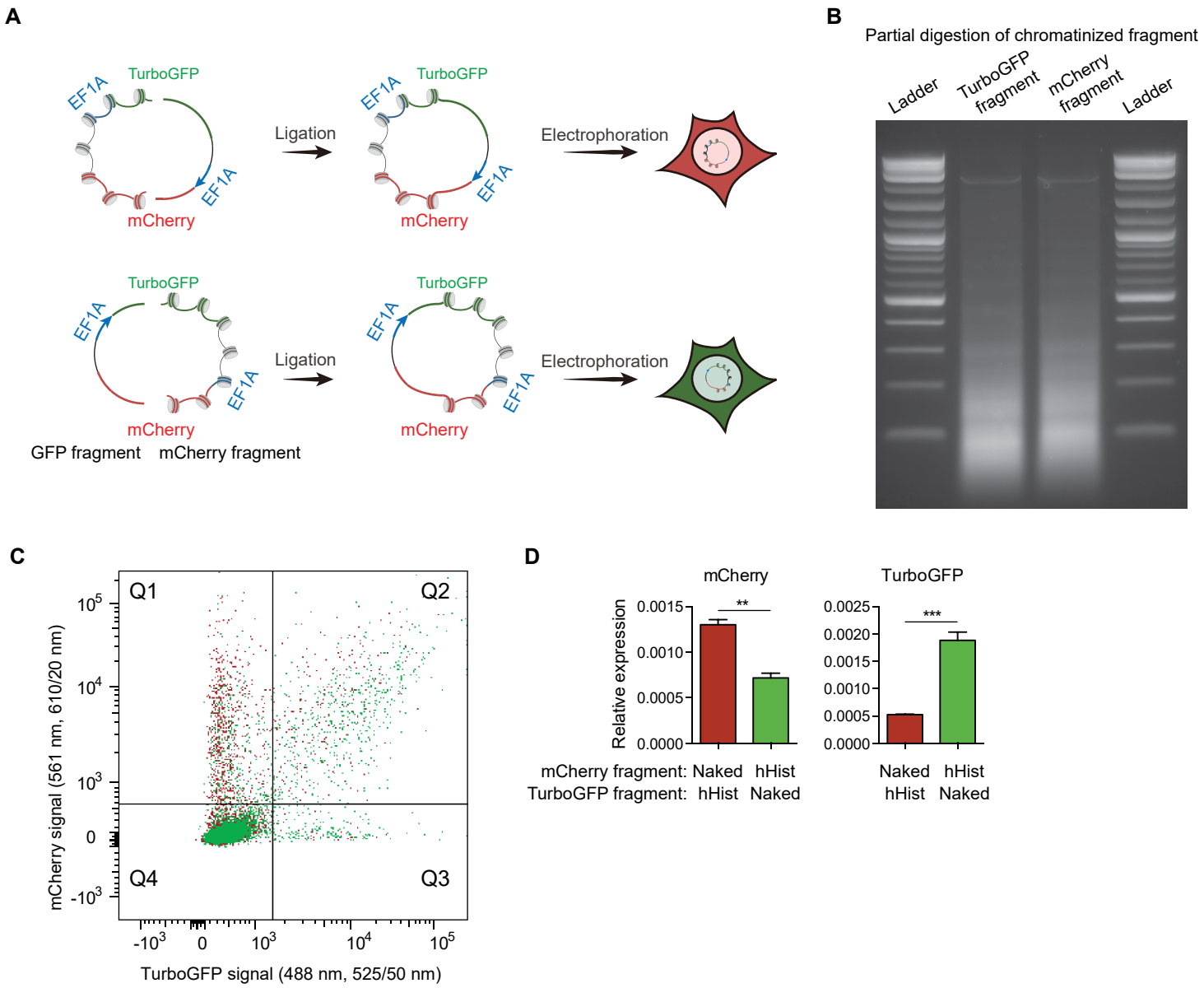


Figure S6. Half-chromatinized episomal DNA are occluded in the chromatinized half. (**A**) Schematic of the half-chromatinization assay. Expression cassettes of TurboGFP and mCherry are split in the middle of the coding sequence. Either the TurboGFP or mCherry fragment is chromatinized while the other one is kept naked. The two fragments are ligated to reconstitute complete TurboGFP and mCherry reporters, followed by electroporation into 293T cells. (**B**) Partial digestion by MNase of half-chromatinized DNA following ligation. (**C**) Flow cytometric analysis of 293T cells electroporated with half-chromatinized DNA. Green dots represent cells containing ligation product of chromatinized mCherry fragment with naked TurboGFP. Red dots represent cells containing ligation product of chromatinized TurboGFP fragment with naked mCherry. (**D**) Relative expression of mCherry and TurboGFP from ligated half-chromatinized DNA. Either the mCherry or TurboGFP fragment was chromatinized with Hela core histones (hHist), and ligated to the naked form of the TurboGFP or mCherry fragment, respectively. RNA levels were measured by RT-qPCR and normalized to Actb (n = 3, unpaired t-test). Primers were designed across ligation junctions.

Figure S7

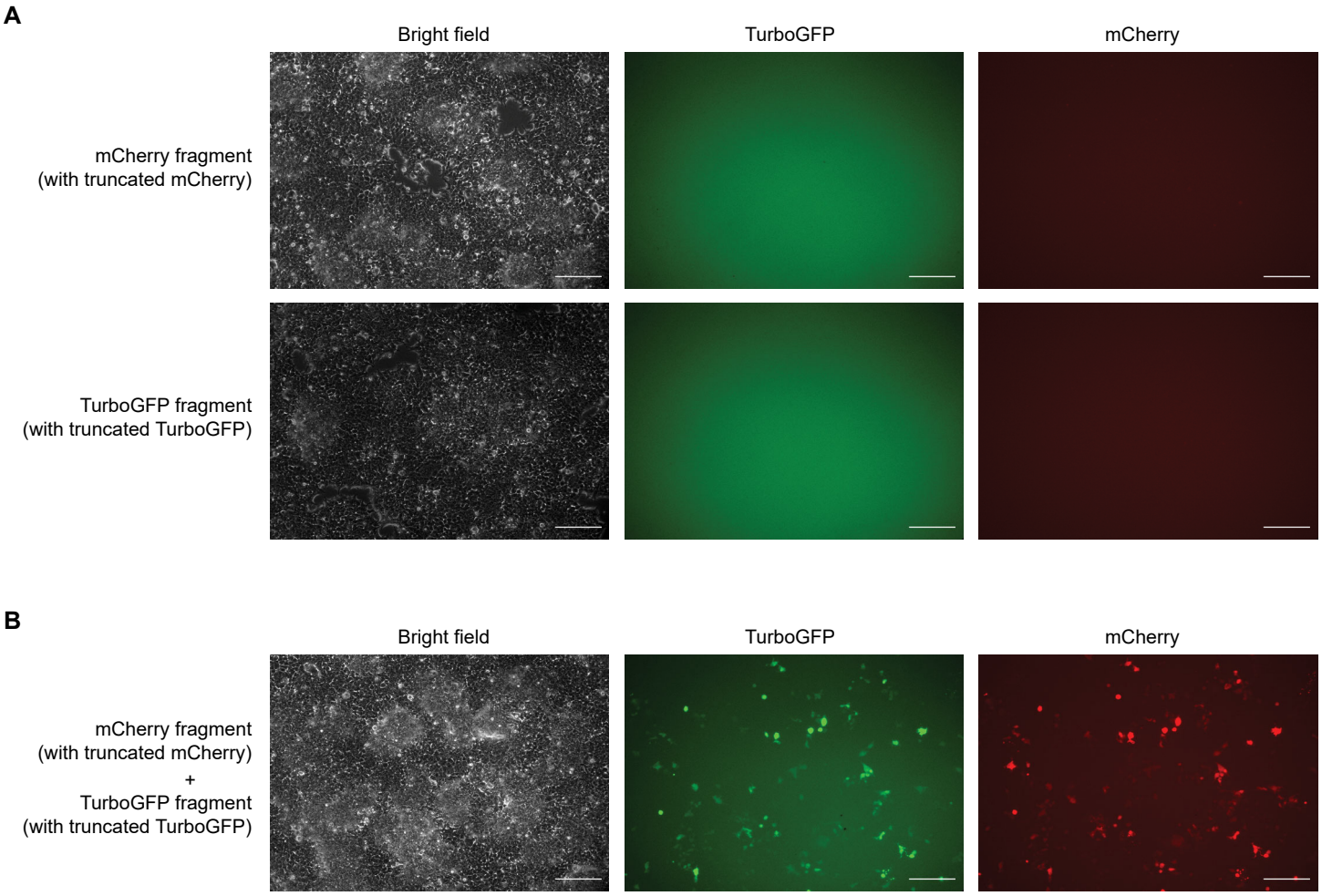


Figure S7. Spontaneous ligation between mCherry and TurboGFP fragments in 293T cells. Expression cassettes of TurboGFP and mCherry were split in the middle of the coding sequence. The two fragments, which had matching sticky ends, were electroporated into 293T cells either individually that failed to produce any fluorescence (A), or simultaneously that produced both TurboGFP and mCherry fluorescence due to spontaneous ligation between the two fragments in cells (B). Scale bar: 200 um.

Figure S8

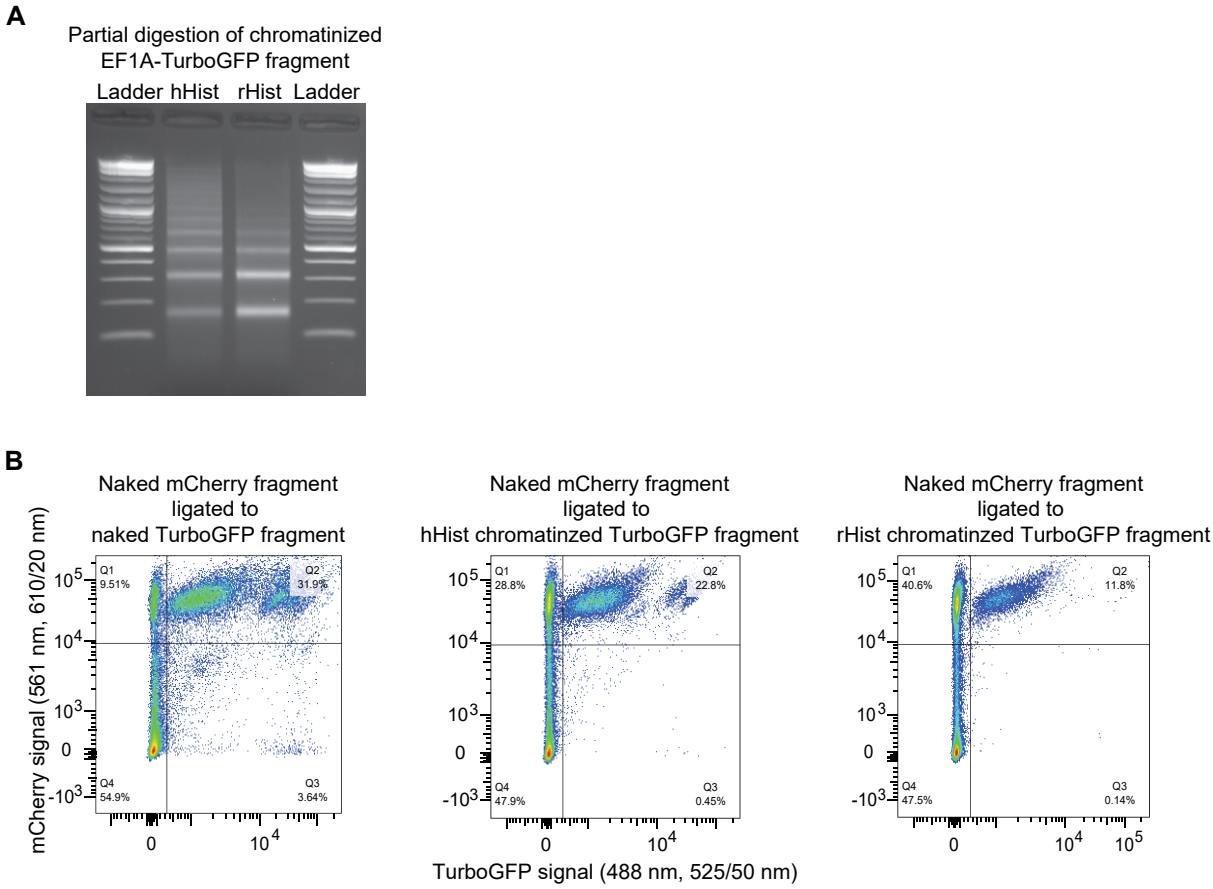


Figure S8. Heritability of chromatinization-mediated gene occlusion through DNA replication. (**A**) Partial digestion by MNase of chromatinized TurboGFP fragment. (**B**) Flow cytometric analysis of 293T cells carrying integrated copies of the naked mCherry fragment ligated with naked, Hela histone (hHist) chromatinzed, or recombinant histone (rHist) chromatinized TurboGFP fragments.

Figure S9

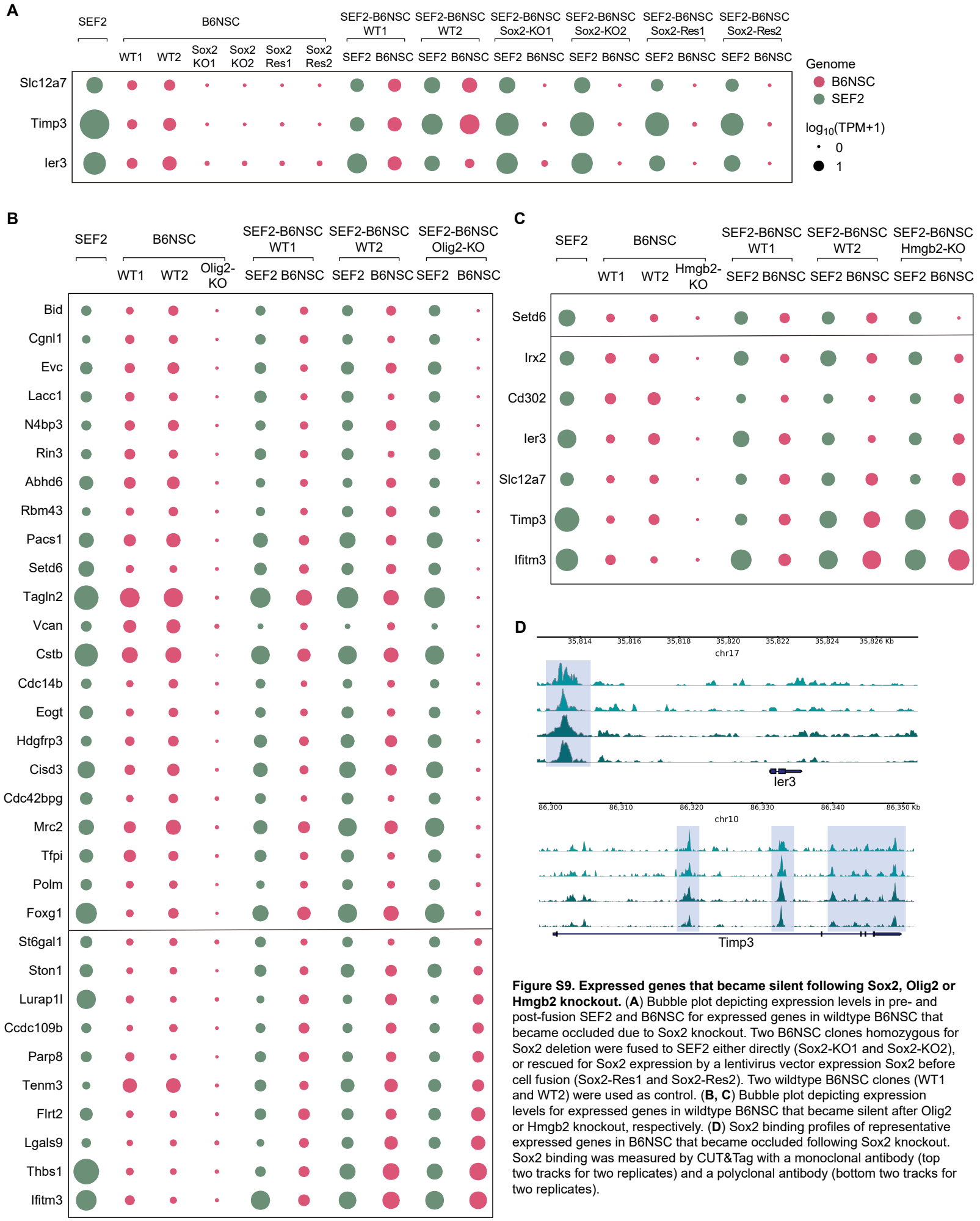


Figure S9. Expressed genes that became silent following Sox2, Olig2 or Hmgb2 knockout. (**A**) Bubble plot depicting expression levels in pre- and post-fusion SEF2 and B6NSC for expressed genes in wildtype B6NSC that became occluded due to Sox2 knockout. Two B6NSC clones homozygous for Sox2 deletion were fused to SEF2 either directly (Sox2-KO1 and Sox2-KO2), or rescued for Sox2 expression by a lentivirus vector expression Sox2 before cell fusion (Sox2-Res1 and Sox2-Res2). Two wildtype B6NSC clones (WT1 and WT2) were used as control. (**B, C**) Bubble plot depicting expression levels for expressed genes in wildtype B6NSC that became silent after Olig2 or Hmgb2 knockout, respectively. (**D**) Sox2 binding profiles of representative expressed genes in B6NSC that became occluded following Sox2 knockout. Sox2 binding was measured by CUT&Tag with a monoclonal antibody (top two tracks for two replicates) and a polyclonal antibody (bottom two tracks for two replicates).

Figure S10

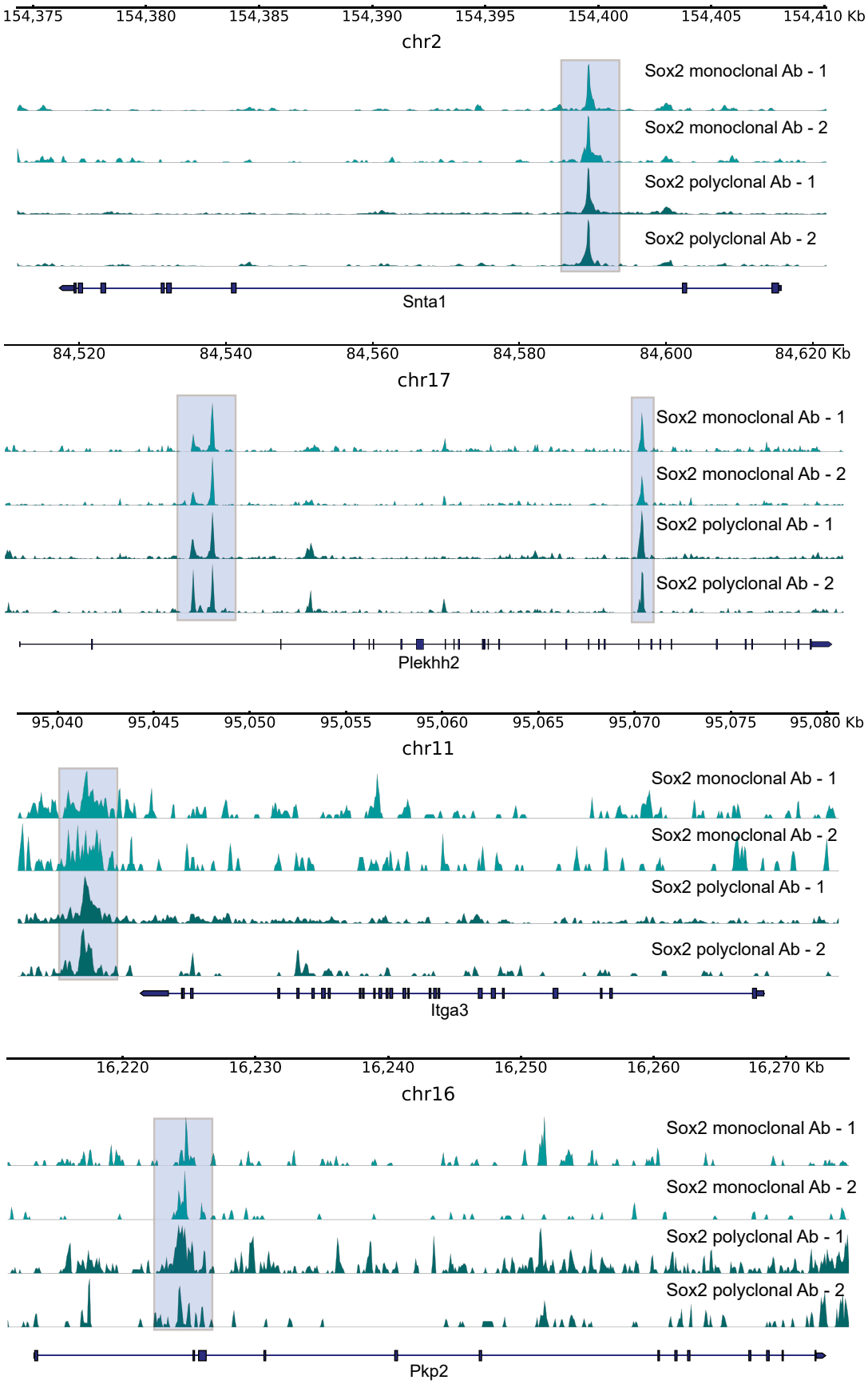


Figure S10. Sox2 binding profiles of activatable genes in B6NSC that became occluded following Sox2 knockout. Sox2 binding was measured by CUT&Tag with a monoclonal antibody (top two tracks for two replicates) and a polyclonal antibody (bottom two tracks for two replicates).

SUPPLEMENTARY TEXT

Several sources of uncertainty exist in the assessment of the transcriptional potency of genes by cell fusion. First, while occlusion is measured at the gene level, it might actually occur at the level of individual cis-regulatory elements (e.g., enhancers). Specifically, a gene could possess multiple enhancers, with only a subset being occluded in a given cell line while the others remaining activatable. Such a gene would appear occluded in fusions that only brought in TFs cognate to its occluded enhancers, but the same gene would appear activatable in fusions that brought in TFs cognate to its activatable enhancers. It is even possible that the same enhancer could be occluded in regards to one set of TFs but activatable in regards to another set of TFs. Second, the statistical power in measuring expression levels of genes could be limited if they are lowly expressed or have few polymorphic sites to facilitate strain-specific mapping of RNA-Seq reads from hybrid cells. This, together with the inaccuracy of RNA-Seq, could lead to stochasticity in determining the occluded or activatable status of genes, which could occasionally result in the same gene being annotated as occluded in one fusion but activable in another. Third, while we attempted to mitigate interspecies incompatibility by fusing cells of the same species, we did deliberately choose to fuse cells from relatively divergent mouse strains in order to maximally exploit inter-strain sequence polymorphisms to specifically map gene expression in hybrid cells to the two fusion partners. Some level of inter-strain incompatibility might still exist, similar to that observed for cis-acting expression quantitative trait loci (cis-eQTL) (*1*). As a result, some activatable genes could be misclassified as occluded if their TFs in hybrid cells failed to activate them to sufficient levels due to inter-strain incompatibly. Fourth, while our analysis excluded fusion samples showing signs of chromosome loss, it remains possible, though perhaps not common, that local mutations such as circumscribed deletions could knockout the expression of a gene, resulting in it appearing occluded. For fusions between mouse and rat cells, the greater interspecies incompatibility would further exaggerate the overestimation of occluded genes and underestimation of activatable genes. For early-stage R1A-E14 fusion samples, additional overestimation of occluded E14 genes was possible if there were remnant mRNAs from extinguished R1A-specific genes that made the corresponding E14 orthologs seemingly occluded. These caveats notwithstanding, the internal consistency of our data across different fusion experiments involving multiple cell types indicate a relatively high degree of reliability of the fusion assay in ascertaining occluded and activatable genes.

SUPPLEMENTARY REFERENCES

1. U. Vosa *et al.*, Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nature genetics* **53**, 1300-1310 (2021).