Method

# Systematic mapping of TF-mediated cell fate changes by a pooled induction coupled with scRNA-seq and multi-omics approaches

Muyoung Lee,[1,3] Qingqing Guo,[1,3] Mijeong Kim,[1,3] Joonhyuk Choi,[1] Alia Segura,[1] Alper Genceroglu,[1] Lucy LeBlanc,[1] Nereida Ramirez,[1] Yu Jin Jang,[1] Yeejin Jang,[1] Bum-Kyu Lee,[2] Edward M. Marcotte,[1] and Jonghwan Kim[1]

[1]Department of Molecular Biosciences, The University of Texas at Austin, Austin, Texas 78712, USA; [2]Department of Biomedical Sciences, Cancer Research Center, University at Albany, State University of New York, Rensselaer, New York 12144, USA

Transcriptional regulation controls cellular functions through interactions between transcription factors (TFs) and their chromosomal targets. However, understanding the fate conversion potential of multiple TFs in an inducible manner remains limited. Here, we introduce iTF-seq as a method for identifying individual TFs that can alter cell fate toward specific lineages at a single-cell level. iTF-seq enables time course monitoring of transcriptome changes, and with biotinylated individual TFs, it provides a multi-omics approach to understanding the mechanisms behind TF-mediated cell fate changes. Our iTF-seq study in mouse embryonic stem cells identified multiple TFs that trigger rapid transcriptome changes indicative of differentiation within a day of induction. Moreover, cells expressing these potent TFs often show a slower cell cycle and increased cell death. Further analysis using bioChIP-seq revealed that GCM1 and OTX2 act as pioneer factors and activators by increasing gene accessibility and activating the expression of lineage specification genes during cell fate conversion. iTF-seq has utility in both mapping cell fate conversion and understanding cell fate conversion mechanisms.

[Supplemental material is available for this article.]

Master transcription factors (TFs), which control cellular identity and development, can often change cell fate upon induction in a process known as reprogramming, best exemplified by somatic cell reprogramming in which overexpression of four TFs generates induced pluripotent stem (iPS) cells from terminally differentiated cells (Takahashi and Yamanaka 2006; Takahashi et al. 2007; Yu et al. 2007; Park et al. 2008). Similarly, induction of a single master TF is often sufficient to reprogram cells in certain contexts; for example, ectopic expression of *Myod1* in fibroblasts generates myoblasts, and *Gata1* overexpression converts avian myelomonocytic cells into eosinophils, thromboblasts, and erythrocytes (Choi et al. 1990; Kulessa et al. 1995). Additionally, ectopic expression of individual trophoblast TFs, such as *Cdx2*, *Gata3*, or *Arid3a*, has been shown to convert embryonic stem (ES) cells to trophoblast-like cells (Niwa et al. 2005; Ralston et al. 2010; Rhee et al. 2014; Blij et al. 2015).

Reprogramming has enormous potential for regenerative medicine, as it enables the generation of desired cell types. To systematically test the potential of individual TFs in converting cell fates, recent studies have performed large-scale overexpression screens (Parekh et al. 2018; Nakatake et al. 2020; Ng et al. 2021; Joung et al. 2023) in human pluripotent stem cells, which are plastic and capable of differentiating into all three germ lineages and even extraembryonic lineages in vitro (Keller 1995; Smith 2001; Cho et al. 2012; Toyooka 2021). Although these studies successfully identified numerous TFs triggering differentiation of pluripo-

tent stem cells, one study relying on a reporter was unable to annotate lineage specification (Ng et al. 2021), and the other studies using a constitutive promoter for the overexpression were unable to control the duration of TF induction (Parekh et al. 2018; Joung et al. 2023). Furthermore, none of the studies investigated the dynamic changes in the transcriptome or the action mechanisms of potent TFs during cell fate conversion.

The advent of single-cell RNA sequencing (scRNA-seq) has revolutionized our understanding of tissue and organ behavior at the level of individual cells (Ofengeim et al. 2017; Hwang et al. 2018; Papalexi and Satija 2018; Svensson et al. 2018; Madissoon et al. 2020; Li et al. 2020; Paik et al. 2020; Shiau et al. 2021). With scRNA-seq, cellular heterogeneity within an organ can be profiled at an unprecedented resolution, enabling the unbiased investigation of the single-cell transcriptome in a massively parallel manner. This approach has been used to study cellular heterogeneity in normal and abnormal settings and has also been combined with other high-throughput techniques, such as RNA interference, CRISPR-Cas9-mediated perturbation, and overexpression screens, to address numerous questions in biology and medicine without relying on a limited number of reporter genes (Adamson et al. 2016; Dixit et al. 2016; Jaitin et al. 2016; Aarts et al. 2017; Datlinger et al. 2017).

Here, we introduce iTF-seq, a novel method for a pooled induction screen of individual TFs with cell fate conversion potential. In this method, ectopic expression of TFs is performed using a transposon-based doxycycline (Dox)-inducible system, which enables greater transfection and genomic integration efficiency

for the generation of individual cell lines for TF induction. Our analysis pipeline detects the induced TF (iTF) in each cell and profiles the resulting transcriptome that emerges from this induction. iTF-seq also allows subsequent multi-omics approaches through metabolic biotinylation of TFs for protein–DNA interaction (PDI) and protein–protein interaction (PPI) mappings centered on the potent TFs. In a pilot test of 80 TFs, iTF-seq identified multiple novel TFs with rapid cell fate conversion potential toward specific lineages, which underscores the importance of TFs in cell fate determination and provides valuable insights into the generation of desired cell types by TFs. The approach will serve as a framework for understanding common or unique TF-mediated cell fate conversion mechanisms.

## Results

### Construction of an SBFB vector

We developed a Sleeping Beauty (SB) transposon-based inducible vector in combination with a metabolic biotinylation (FB) system to enable the inducible expression of TFs in ES cells, which we refer to as the SBFB vector (Ivics et al. 1997; Kim et al. 2008; Hackett et al. 2010). Previous research has suggested that transposon-mediated approaches are superior to lentivirus methods for open reading frame (ORF) overexpression owing to the simplification of reagent preparation, reduced promoter silencing, and increased induction levels with higher genomic integration efficiency (Ng et al. 2021). In the SBFB system, the ectopically expressed proteins upon Dox treatment undergo biotinylation via a biotin ligase (BirA) that is constitutively expressed in J1 mouse ES cells (BirA-ES cells). This allows for streptavidin-mediated downstream multi-omics applica-

tions, including mapping of genomic targets and interacting partner proteins of the iTF of interest independently, in addition to the detection of protein induction and localization as previously described (Fig. 1A; Supplemental Fig. S1A,B; Kim et al. 2008, 2009, 2010).

### Generation of iTF lines

To investigate previously unreported TFs with cell fate conversion potential and to confirm the feasibility of the iTF-seq method, we generated cell lines for the induction of individual TFs in ES cells (iTF lines). A set of 80 TFs was selected as a proof of principle test. These TFs were chosen based on several factors, including cell or tissue type–specific expression patterns often displaying instructive roles in development (Zhang et al. 2019). Some of them are ubiquitously expressed, and we also included multiple previously reported TFs with reprogramming potential as a single factor, such as *Gata6* (Fujikura et al. 2002; Wamaitha et al. 2015), *Gata4* (Fujikura et al. 2002; Holtzinger et al. 2010), *Pdx1* (Bernardo et al. 2009), *Cited1* (Xu et al. 2018), *Cdx2* (Niwa et al. 2005; Blij et al. 2015), *Gli1* (Denham et al. 2010), *Arid3a* (Rhee et al. 2014), *Gata3* (Ralston et al. 2010), *Elf5* (Latos et al. 2015), *Nkx2-5* (Ruan et al. 2016), *Tfap2c* (Latos et al. 2015), *Fosl1* (Lee et al. 2018), and *Zfp36l1* (Tseng et al. 2017). Approximately 80% of the individual TFs tested did not have confirmed reprogramming potential from prior studies, as summarized in Supplemental Table S1. The cDNAs for the selected TFs were cloned into the SBFB vector, and 80 individual stable iTF lines were generated, as described in the Methods section, using transient expression of the SB transposase. This approach allowed us to generate stable iTF lines in a small-scale culture (96-well plate). We validated protein expression of
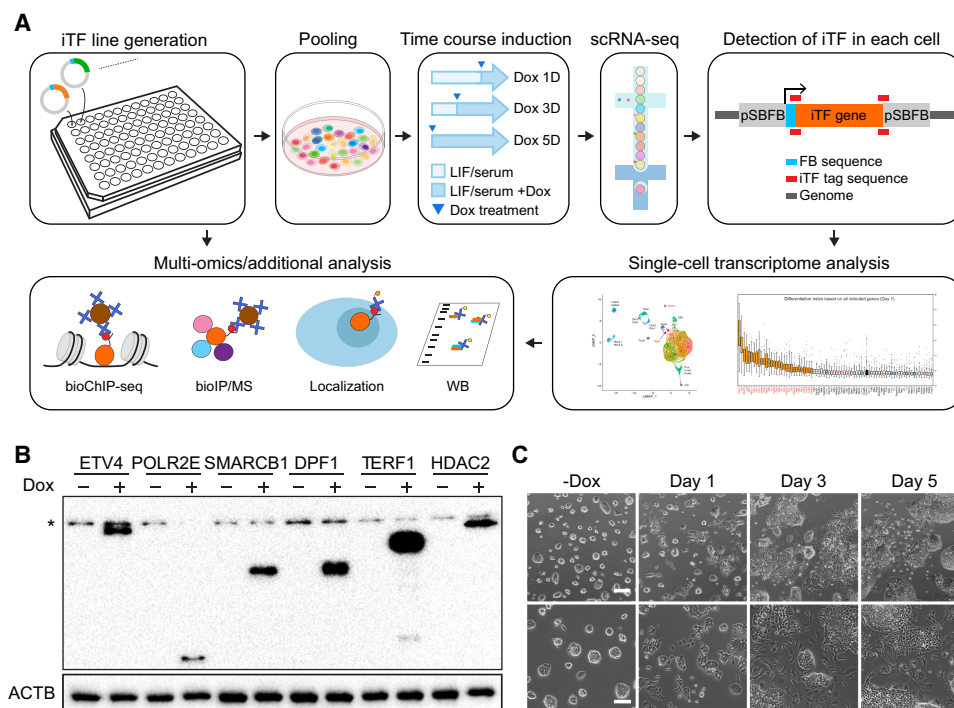


**Figure 1.** iTF-seq enables screening of numerous TFs with fate conversion potential. (*A*) Schematic representation of iTF-seq procedure and subsequent multi-omics approaches. (*B*) Representative western blot images showing inducible overexpression of biotinylated proteins in individual iTF lines detected by streptavidin-HRP. (*) Nonspecific bands, (Dox) doxycycline. (C) Morphology of 80 pooled iTF cell lines under uninduced condition (−Dox) and upon induction of TFs by treating Dox (0.5 μg/mL) for 1, 3, and 5 d. Scale bars, 100 μm or 200 μm (*upper*).

all individual biotinylated TFs with streptavidin-HRP after 1 d of Dox induction (0.5 μg/mL) (Fig. 1B). For some iTF lines, we observed significant changes in morphology upon induction (Supplemental Fig. S2A).

## Detection of each cell with a specific TF induction by iTF-seq

To investigate transcriptome and cell fate changes caused by the ectopic induction of individual TFs at single-cell resolution, we performed scRNA-seq of a pool of 80 stable iTF lines and wild-type ES cells upon treatment with Dox for 1, 3, and 5 d. Similar to the test of individual iTF lines, a substantial proportion of cells within the pool showed a flattened morphology indicating differentiation upon 1 d of induction (Fig. 1C).

To detect a specific TF expressed ectopically in each cell, we applied a barcode-independent approach. Parts of the SBFB vector are transcribed with the TF coding sequences upon Dox induction, and we determined the junctional sequences between the vector and TF coding region (iTF tags) (Fig. 1A). This approach also distinguishes endogenous versus ectopic TFs within the cell, as iTF tags are unique for ectopic TFs and cannot be found in the endogenous transcripts. From the iTF-seq results, we collected all reads matched with the iTF tags and successfully detected individual cells expressing specific iTFs. Although we expected that each cell expresses only one iTF based on our experimental design (Fig. 1A), we observed cells expressing iTF tags associated with multiple iTFs. The results may arise from the cumulative errors from single-cell dissociation, sequencing, and mapping processes. We optimized our analysis pipeline and decided to apply three unique molecular identifiers (UMIs) as a threshold for the number of UMIs for each iTF (see Methods) (Supplemental Fig. S2B). As a result, we successfully detected 77 iTFs (7938 cells), 76 iTFs (5714 cells), and 68 iTFs (7902 cells) from day 1, 3, and 5 samples, respectively (Supplemental Table S2). The proportion of single TF overexpressing cells identified is 62.7% (day 1), 60.9% (day 3), and 58.9% (day 5), calculated among all cells. These numbers are consistently ~60% for cells that passed quality control on each respective day.

## iTF-seq captures changes in transcriptome mediated by single TF induction

With the obtained iTF-seq data, we performed UMAP clustering and visualized the positions of cells to quickly monitor the global gene expression differences among the detected cells (Fig. 2A–C). We observed one big cluster of cells along with multiple small cell clusters distinctly separated from the big cluster, indicating that the cells within the small clusters show unique gene expression profiles and are potentially differentiated owing to the induction of specific TFs with cell fate conversion potential. Indeed, as shown in Figure 2A, we found that many cells expressing iTFs with already known reprogramming potential, such as *Pdx1*, *Gata3*, and *Fosl1* (Supplemental Table S1), belong to such separated cell clusters. Notably, in contrast to many reprogramming processes characterized by low efficiencies, we observed that TFs inducing cells to locate in separated cell clusters are potent, as most such TF-overexpressing cells tend to position in separated cell clusters rather than in the large undifferentiated cell cluster. For instance, on day 1, all 93 (100%) ectopic *Dlx3*-expressing cells are localized within the separated cluster. Similarly, among 170 *Elf5*-expressing cells and 75 *Fosl2*-expressing cells, 97.8% and 80%, respectively, are found in separated clusters (Supplemental Table S3). The high efficiency of cell fate changes by inducing

TFs was also observed in a few validation cases in the recent scRNA-seq-based single TF overexpression screen (>80% efficiency confirmed through testing reporter gene expression) (Ng et al. 2021). We found many additional cell clusters expressing iTFs without previously known cell fate conversion potential. This implies that iTF-seq is sensitive enough to identify previously known and unknown TFs capable of cell fate changes. As expected, single-cell transcriptome and pseudobulk RNA-seq analysis indicated that the cells belonging to the separated clusters express a relatively lower expression level of ES cell marker genes, such as *Nanog* and *Sox2*, compared with the control cells, owing to the loss of ES cell identity after iTF induction (Fig. 2D; Supplemental Table S4).

## TF induction can rapidly trigger differentiation of ES cells

None of the prior TF overexpression studies monitored time-dependent transcriptome changes (Parekh et al. 2018; Nakatake et al. 2020; Ng et al. 2021; Joung et al. 2023), and cell fate changes were monitored upon 4–7 d of TF overexpression. We presume that time course experiments would give us additional insights into the roles of TFs with cell fate conversion potential as we observed morphological changes upon 1 d of TF induction in multiple individual cases (Supplemental Fig. S2A). The UMAP visualization also found that numerous TFs could change cell fates after only 1 d of induction, suggesting that TF-mediated reprogramming can occur rapidly. Considering the typically slow reprogramming processes, such as somatic cell reprogramming, this finding shed light on the speed of TF-mediated reprogramming (Fig. 2A). It is worth noting that previous studies have suggested extremely rapid and dynamic control of TF-mediated transcriptional activity in response to a variety of stimuli, as seen in the case of the Jun–Fos AP1 complex (Devary et al. 1991; Patel et al. 1994). This further highlights the potential for speedy cell fate manipulation through TF induction.

## Determining cell fate changes by calculating the differentiation index

Although we observed that the induction of multiple TFs triggered the differentiation of ES cells, the interpretation based on the UMAP results was subjective (Fig. 2A–C). To determine changes in cell fate, we developed a quantitative measure called the "differentiation index." We calculated the distances between the centroid of control cells and each control cell, using gene expression values as elements of the cells' expression vectors. Next, we measured the distances between the centroid of control cells and each cell population expressing a specific iTF and normalized them based on the mean and standard deviation of the control cells. To account for the interdependence of gene expressions, we applied PCA to the expression values before measuring distances from the centroid of control cells and used coefficients of PCs for the calculation (Supplemental Fig. S3A).

To assess cell fate changes, we calculated the differentiation index using the expression data of all genes detected by iTF-seq. As expected, many cell populations with specific iTF induction showed higher differentiation index values than did the control cells (Fig. 3A), indicating significant differences in overall transcriptomes. TFs that caused such disturbances were deemed potent TFs with the ability to induce cell fate conversion. Consistent with the UMAP data (Fig. 2A), such factors include *Gcm1*, *Jun*, *Nkx2-5*, *Fos*, *Dlx5*, and *Cdx2* (Fig. 3A). We also calculated the differentiation index for the time course data and found that 29, 20, and 21 TFs corresponding to each day had cell fate conversion potential
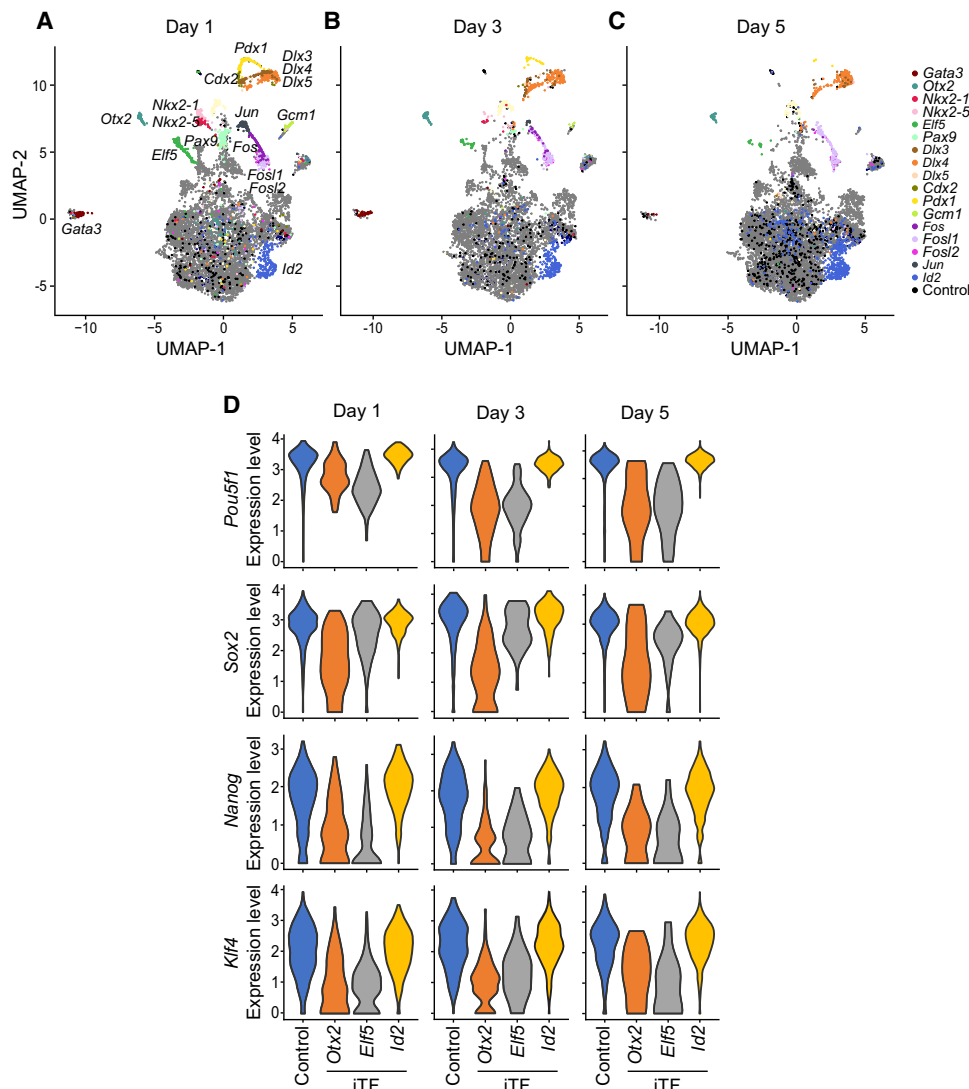
**Figure 2.** iTF-seq enables search for TFs with cell fate conversion potential. (*A–C*) UMAP plots of cells expressing individual iTFs and control cells after 1, 3, and 5 d of induction. Some iTF-expressing cells and control cells are marked with colors. (*D*) Relative expression levels of ESC markers (*Pou5f1*, *Sox2*, *Nanog*, and *Klf4*) in control and *Otx2-*, *Elf5-*, and *Id2*-expressing cells after 1, 3, and 5 d of induction.

(Mann–Whitney $U$ test, adjusted $P$-value < 0.01) (Supplemental Table S5). In contrast, cells with induction of many other individual TFs, such as *Cited1*, *Cited2*, *Erf*, *Hmgn2*, *Hopx*, *Id2*, *Rbpjl*, or *Snai1*, had similar or relatively lower differentiation index values compared with those of the control cells, indicating that these TFs did not significantly alter cell fates upon induction (Fig. 3A).

We further assessed the changes in expression of the extended core pluripotency-related genes, which are typically down-regulated upon differentiation of ES cells, and the polycomb repressive complex (PRC) target genes, which are up-regulated upon differentiation (Kim et al. 2010). We used these two gene sets to calculate differentiation indexes, and as shown in Figure 3, B and C, and Supplemental Figure S3B, the percentile ranks based on the index values using all genes displayed a strong positive correlation with the ranks calculated using the core-related genes (Fig. 3B) or PRC target genes (Fig. 3C). This provides independent confirmation that the potent TFs induced cell fate changes from ES cells to differentiated cells by down-regulating the ES

cell core factors and activating various lineage marker genes repressed by PRC in self-renewing, undifferentiated ES cells.

## Inducing TFs for a brief duration is sufficient for cell fate changes

A common feature observed during reprogramming or transdifferentiating factors is the activation of the endogenous expression of the induced factors. To investigate whether factors with a high differentiation index activate their endogenous expression, we examined the activation of endogenous TF upon the ectopic expression. We used specific primer pairs designed to amplify the junction between the 5′ or 3′ UTR and exons for the detection of endogenous transcripts. As shown in Supplemental Figure S4A, 13 out of 18 TFs tested with a high differentiation index showed the activation of endogenous TFs. In contrast, all tested TFs with a low differentiation index were confirmed to show no activation of endogenous TFs upon induction, despite showing comparable ectopic
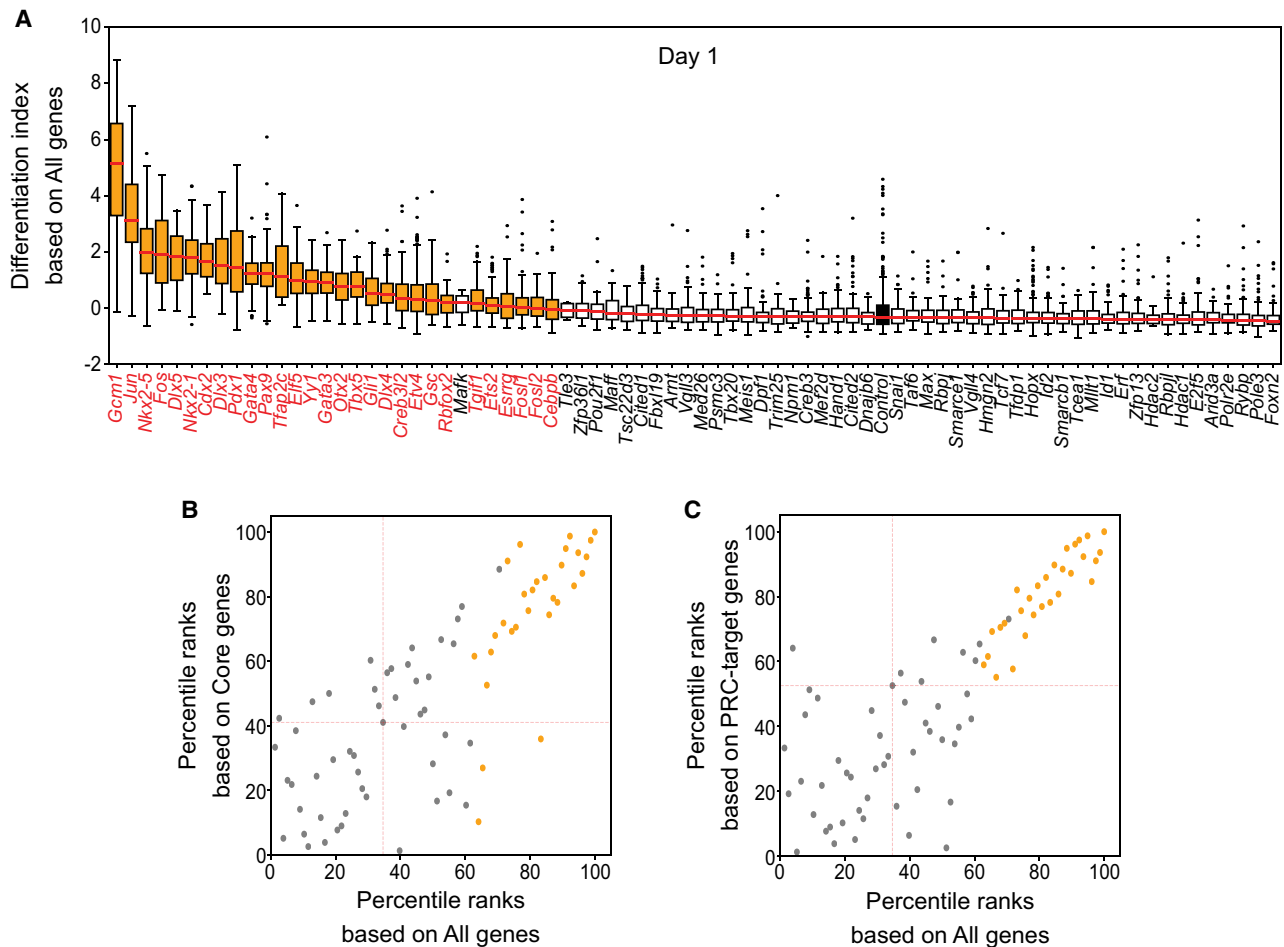
**Figure 3.** Determining cell fate changes by calculating differentiation index. (*A*) Differentiation index for cells expressing single iTF and control cells. Each box plot shows the distribution of the differentiation index for each cell population expressing one iTF (medians are marked with red lines). TFs with cell fate conversion potential are marked with orange, and the control is black. *Mafk* was not significant based on adjusted *P*-value. (*B*,*C*) Comparison between differentiation indexes based on all detected genes (all genes) and core pluripotency genes (core genes; *B*) or PRC target genes (*C*) after 1 d of induction (because the differentiation indexes for each gene group have different ranges, percentile ranks were used). TFs with cell fate conversion potential are marked with orange. Red dashed lines show the location of control cells. Some TFs (gray dots), like *Mafk*, are located near or even in the cluster of TFs with cell fate conversion potential (orange dots). These TFs were not selected based on the results of statistical testing.

induction levels to TFs with a high differentiation index (Supplemental Fig. S4A,B).

We investigated, aligned with examining reprogramming or cell fate–changing potential, whether differentiated cells induced by potent TFs can maintain their altered cell states without reverting to ES cell–like states or require continuous Dox treatment. To address this, we conducted a time course transcriptome profiling of Dox removal samples for potent TFs, including *Dlx5*, *Gata3*, *Gcm1*, *Otx2*, and *Pdx1*. In addition to maintaining continuous Dox treatment for 7 d as a control, we generated samples by discontinuing Dox after 1 day, 3 days, and 5 days of culture. As shown in Supplemental Figure S4C, visualizing 3000 genes showing higher expression variance, the overall gene expression patterns after Dox removal largely differ from those of control ES cells and align more closely with Dox-maintained cell lines for 7 d. In most instances, our findings indicate that a 3-d period of TF induction via Dox treatment is sufficient to commit cells to differentiated states, with no subsequent regaining of ES cell line expression patterns. We observed that even 1 d of induction followed by extended culture without Dox resulted in losing an ES cell–specific

transcriptome profile. To further assess the maintenance of differentiated status in potent TF-induced cells, we monitored the expression of pluripotency markers (*Pou5f1*, *Nanog*, and *Sox2*) in cells overexpressing *Yy1*, *Creb3l2*, *Cdx2*, *Ets2*, *Gsc*, or *Cebpb*, followed by treatment with Dox for 1 or 5 d. As shown in Supplemental Figure S4D, the expression of pluripotency marker genes decreased over time compared with their levels in control ES cells, albeit with variable patterns. In summary, there was no substantial evidence of cells returning to ES cell–like states over time.

## ES cells experiencing fate changes show slower proliferation and increased cell death

We observed that the proportion of some iTF-induced cell populations forming separate UMAP clusters, including *Pdx1*, *Cdx2*, *Gata3*, and *Gcm1*, gradually decreased with longer induction durations (Fig. 2A–C; Supplemental Table S2). Because ES cells have a higher proliferation rate than differentiated cells (Zaveri and Dhawan 2018) and we maintained leukemia inhibitory factor

(LIF) in the culture even after iTF induction, the dilution of iTF-induced cells by rapidly proliferating ES cells over time is likely responsible for this observation. Therefore, we hypothesized that some TFs, barely detected in day 3 or 5 samples, such as *Pdx1*, *Jun*, *Nkx2-5*, and *Gata3*, might still have significant reprogramming potential (Supplemental Fig. S5A; Supplemental Table S2). Additionally, as pluripotent stem cells undergo increased cell death upon differentiation (Duval et al. 2000; Bashamboo et al. 2006), we thought that cell death might also contribute to the decreased proportion of some TF-induced cells. To address these possibilities, we monitored the proliferation rate and cell death of multiple individual iTF lines associated with a high differentiation index (i.e., *Gcm1*, *Nkx2-5*, *Dlx5*, *Pdx1*, and *Gata3*) and a low differentiation index (i.e., *Meis1*, *Mef2d*, *Id2*, *Mllt1*, and *Erf*) in Figure 3A. Our results validated that cells with a high differentiation index, indicating rapid differentiation, show decreased proliferation with increased cell death, whereas the cells with a low differentiation index show no significant changes in cell proliferation and cell death (Fig. 4A,B; Supplemental Fig. S5B,C). Hence, our conclusion is that the rates of proliferation and cell death are intricately connected to TF-mediated differentiation of ES cells. The findings further indicate that when assessing differentiated or differentiating ES cells, proliferation and cell death rates should be regarded as pivotal factors in high-throughput differentiation screens. Moreover, these results underscore the advantages of using controlled TF induction, an approach not attainable with constitutive promoter-based methods.

## iTF-mediated rapid global expression changes and lineage specifications

Using the Dox-inducible system in our approach allowed us to reveal that a mere 1-d induction of potent TFs is sufficient to alter the global gene expression profile (Fig. 2). To delve deeper into the speed of cell fate changes, we conducted global gene expression profiling at additional time points, 3 h, 6 h, 12 h, day 1, day 3, and day 5, following the induction of *Otx2*, *Gcm1*, or *Dlx5*. As shown in Supplemental Figure S6A, although the timing varies depending on the iTFs, a 12-h induction is sufficient to transform the transcriptome from that of uninduced ES cells, suggesting that iTF-mediated cell fate changes are exceptionally rapid and dynamic processes.

To further investigate the specific lineages induced by each TF induction, we calculated differentiation indexes using 643 additional gene sets from the Molecular Signature Database (MSigDB), including three germ layer–related gene sets (total, eight: two from C2 curated and six from C5 ontology gene set collection) and C8 cell-type signature gene set collection (Subramanian et al. 2005; Liberzon et al. 2011). Overall, TFs with high differentiation index values based on all genes detected from our samples (Fig. 3A) also show relatively high values from the test of MSigDB gene sets (Supplemental Table S6), and some cases indicated incipient trajectories of early differentiation. For instance, *Gata4*- and *Gata3*-induced cells showed the highest values from endoderm gene sets, whereas *Elf5*-induced cells showed high values from mesoderm gene sets. Moreover, we found that *Gcm1* overexpression induces trophoblast or placenta-related cell types, and *Jun*, *Fos*, or *Gcm1* induction generates heart-related cell types. Notably, some TFs, such as *Jun*, *Fos*, and *Gcm1*, activated gene sets associated with multiple lineages or specialized cell types (Fig. 5A; Supplemental Table S6). We also computed a similarity matrix with the median values of differentiation indexes for all available

pairs of gene set groups and iTFs (Fig. 5B) and found that among TFs with cell fate conversion potential, those belonging to the same family often showed similar lineage specification potential, indicating either redundancy or positive feedback loops among these family TFs (e.g., GATA3 and GATA4; DLX3, DLX4, and DLX5; NKX2-1 and NKX2-5; and AP-1 complex proteins, including JUN, FOS, FOSL1, and FOSL2). To validate the induced lineage markers associated with cells overexpressing TF within the same family, we individually overexpressed each TF using iTF lines. Subsequently, we monitored the expression of associated lineage marker genes through reverse transcription quantitative polymerase chain reaction (RT-qPCR). Marker genes for distinct cell types/lineages were selected from MSigDB, PanglaoDB (Franzén et al. 2019), and the Azimuth data set (Hao et al. 2021). As illustrated in Supplemental Figure S6B, the induction of *Dlx3*, *Dlx4*, or *Dlx5* led to increased expression levels of marker genes associated with fetal lung ciliated epithelial cells, as indicated by a high differentiation index calculated using the Descartes fetal lung ciliated epithelial cell gene set (Supplemental Table S6). We also confirmed the common activation of lineage marker genes for fetal spleen vascular endothelial in *Nkx2-1* and *Nkx2-5* overexpressed cells (Supplemental Table S6; Supplemental Fig. S6C). In addition to using the differentiation index, we used SingleCellNet (Tan and Cahan 2019) in conjunction with training data (Han et al. 2018) to monitor iTF-mediated lineage specification. However, we encountered challenges in predicting potential lineages, indicating that SingleCellNet might not be optimal for analyzing our data, possibly owing to the early differentiation status of the iTF cells.

## GCMI and OTX2 function as activators during TF-mediated cell fate changes

Among genes that encode TFs that rapidly trigger cell fate changes, *Gcm1* and *Otx2* are of particular interest, as their cell fate conversion potential has not been reported in mouse ES cells. *Gcm1* had the highest differentiation index among all potent TF genes (Fig. 3A) and is expressed explicitly in a subset of mouse placental trophoblast cells, in which it plays a critical role in placental cell fusion (Stecca et al. 2002). Our analysis of pseudobulk RNA-seq data and independent RT-qPCR confirmed that *Gcm1*-induced cells express many trophoblast marker genes, such as *Gata3*, *Arid3a*, *Hand1*, *Tfap2c*, and *Cdx2* (Supplemental Fig. S7A; Supplemental Table S4), suggesting that *Gcm1* trans-differentiates ES cells to trophoblast-like cells. *Otx2*, on the other hand, encodes a TF previously implicated in brain, cerebellum, and eye development (Ruiz-Reig et al. 2019), and a high level of *Otx2* induces rod cell fate in retinal progenitors (Yamamoto et al. 2020). Recent screens also revealed that overexpression of human *OTX2* triggers differentiation of human pluripotent stem cells (Ng et al. 2021; Joung et al. 2023). We observed that *Otx2* induction activates the expression of multiple neuronal lineage markers in both pseudobulk RNA-seq data and RT-qPCR analysis, consistent with its in vivo function (Supplemental Fig. S7B; Supplemental Table S4).

To further investigate the cell fate–changing capacity of *Gcm1* and *Otx2*, we analyzed the pseudobulk RNA-seq data and identified differentially expressed genes upon the induction of each TF. As shown in Figure 6, A and B, Gene Ontology (GO) analysis revealed that the down-regulated genes upon the induction of *Gcm1* (491 genes) or *Otx2* (176 genes) are associated with ES cell maintenance and proliferation-related terms, indicating that both *Gcm1* and *Otx2* triggered the differentiation of ES cells upon induction. The up-regulated genes by *Gcm1* (1806 genes) were enriched in
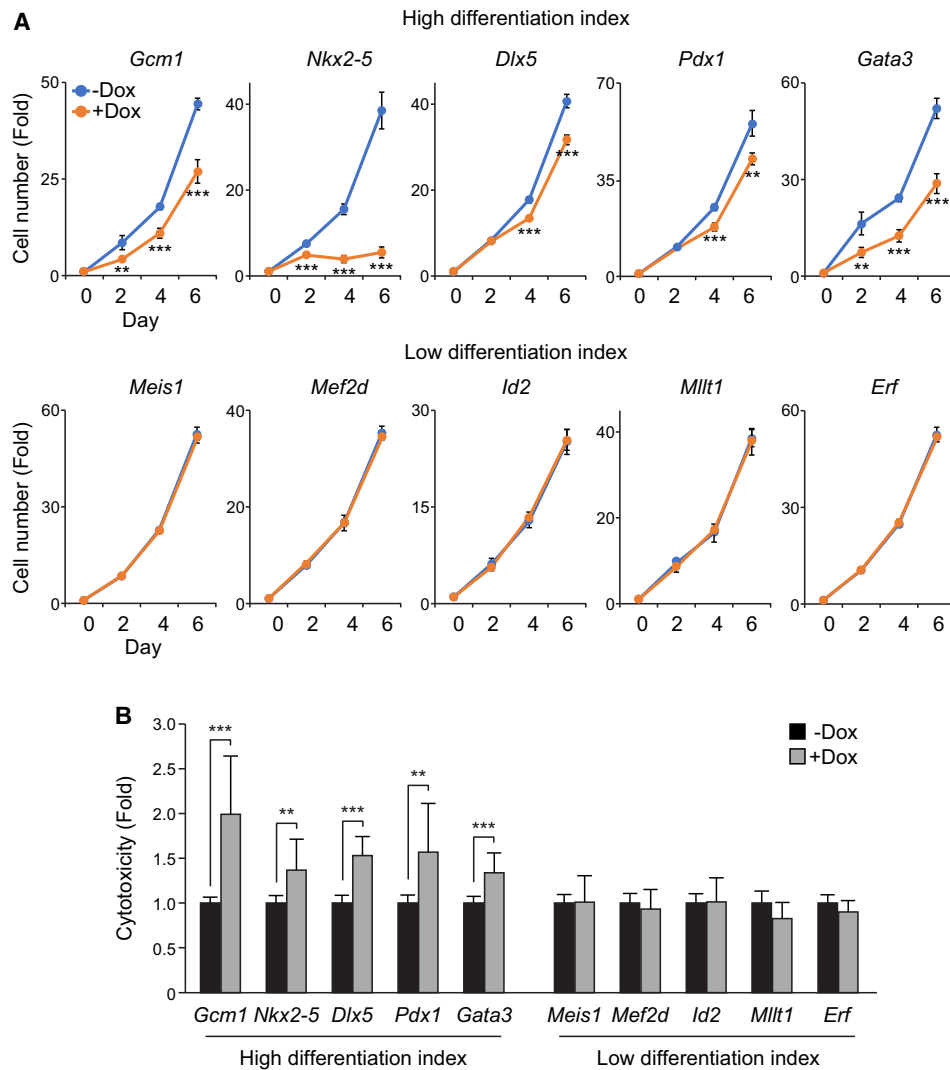
**Figure 4.** ES cells experiencing fate changes show slower proliferation and increased cell death. (*A*) Line plots depicting the time course measurement of cell number upon overexpression of iTFs via Dox induction (0.5 μg/mL; +Dox). The untreated cells (−Dox) were used as a control. The *top* row represents iTFs with high differentiation indexes, and the *bottom* row represents iTFs with low differentiation indexes. Cell numbers were counted up to 6 d after Dox induction, and the fold changes in cell numbers from day 0 were calculated and plotted (n = 2). Data were represented as mean ± SD, and a Student's *t*-test was performed to determine significance between −Dox and +Dox: (**) $P < 0.01$, (***) $P < 0.001$. (*B*) A bar plot shows the fold change in cell death in 10 iTF lines upon overexpression of the iTF by treating Dox (0.5 μg/mL) for 2 d. The cytotoxicity was calculated by measuring lactate dehydrogenase (LDH) activities in the culture media. The percentage of cytotoxicity was calculated for each value by dividing it against the value from the wells treated with lysis buffer, which provided the maximum LDH activity. The fold change was calculated against the control samples. Finally, the average percentage of cytotoxicity was calculated across replicates (n = 3–6), and a Student's *t*-test was performed to determine significance: (**) $P < 0.01$, (***) $P < 0.001$.

differentiation-related terms, including cell differentiation, tube development, and placenta development in accordance with the up-regulation of trophoblast marker genes. The 257 genes up-regulated in *Otx2*-induced cells were related to nervous system development, which is in accordance with previous reports (Ruiz-Reig et al. 2019). Overall, induction of *Gcm1* and *Otx2* resulted in cell fate conversion by down-regulation of ES cell marker genes and promoting the activation of lineage-specific genes.

As the SBFB vector used in iTF-seq enables biotinylated iTF-mediated mapping of PDI and PPI (Kim et al. 2009), we aimed to gain insight into the molecular mechanisms underlying the cell fate changes induced by *Gcm1* and *Otx2*. Therefore, we performed biotin-mediated chromatin immunoprecipitation followed by high-throughput sequencing (bioChIP-seq) to identify direct tar-

gets of these TFs (Supplemental Fig. S7C). As both TFs rapidly induced cell fate changes, we performed bioChIP-seq after 1 d of induction. bioChIP-seq analysis revealed 130,265 and 88,585 target peaks of GCM1 and OTX2, respectively, and motif analysis using HOMER (Heinz et al. 2010) showed that the top-ranking motifs of GCM1 and OTX2 were similar to their previously known motifs (Supplemental Fig. S7D). Both TFs showed a strong preference for binding to distal enhancers, which are known to be strongly associated with cell fate determination (Supplemental Fig. S7E; Ong and Corces 2012). Furthermore, a combined analysis of bioChIP-seq and pseudobulk RNA-seq data showed that both TFs had a strong occupancy on the genes activated by the induction of each TF (Fig. 6C,E). These results suggest that GCM1 and OTX2 function as transcriptional activators during cell fate conversion.
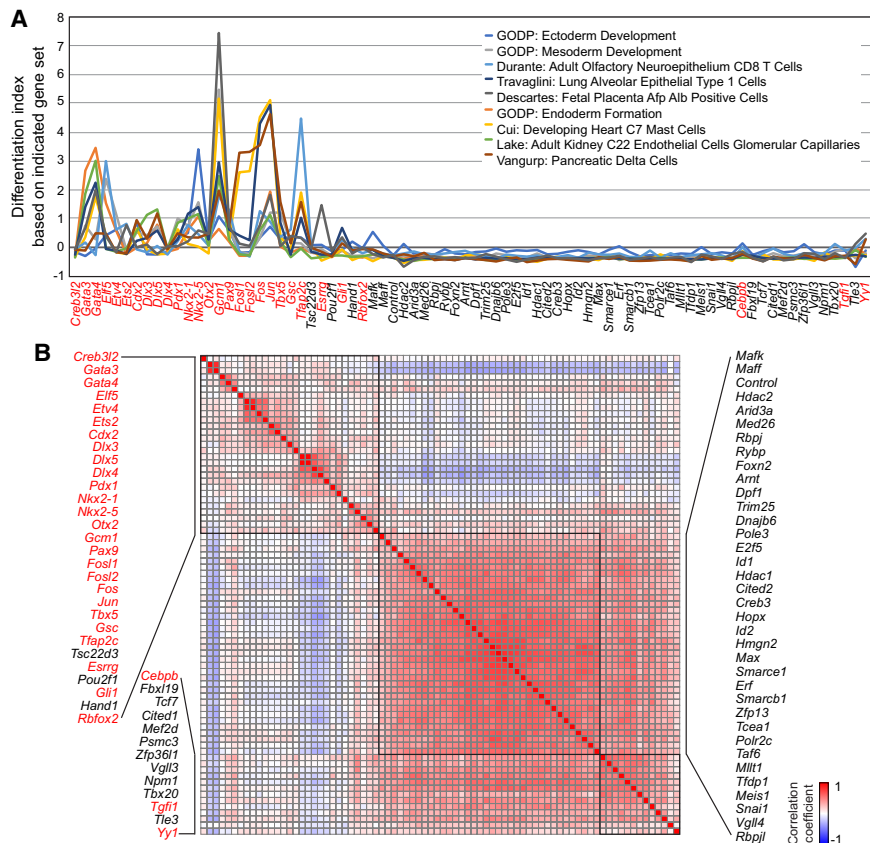
**Figure 5.** iTF-mediated lineage specification. (A) Differentiation index calculated with gene expression responses to the induction of 80 individual TFs (day 1) and indicated gene sets obtained from MSigDB signature gene set collection. Red text indicates TFs harboring cell fate conversion potential defined by differentiation indexing (Fig. 3A). (B) Similarity matrix of iTFs calculated from the differentiation indexes, which were based on all pairs of iTFs and 643 gene sets from MSigDB. Some groups of TFs show high similarity and form clusters of red cells and black boxes, meaning they have similar differentiation indexes across various cell type– or lineage-specific gene sets. Red text indicates TFs with cell fate conversion potential defined by differentiation index.

## GCM1 and OTX2 function as pioneer factors

Because the genes directly targeted by GCM1 and OTX2 during TF-mediated cell fate conversion were not generally active in self-renewing ES cells (Fig. 6C,E), we hypothesized that GCM1 and OTX2 might have pioneer factor activity (Iwafuchi-Doi and Zaret 2016; Morris 2016), which would allow them to bind to and open up previously inaccessible chromatin regions. To test this, we analyzed the chromatin accessibility of ES cells before and after induction of each TF using assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq). As shown in Figure 6, D and F, we found that the chromatin accessibility of control NANOG binding sites was similar to those in ES cells and the cells without TF induction. However, subsets of GCM1 and OTX2 target loci displayed extremely low ATAC-seq signals in control ES cells or the cells without induction, but their ATAC-seq signals increased significantly upon induction of GCM1 and OTX2, suggesting that these TFs likely act as pioneer factors and increase the chromatin accessibility of silenced lineage markers in ES cells, thereby facilitating cell fate conversion.

## Discussion

In summary, this study highlights the potential of iTF-seq to identify previously unknown potent TFs with cell fate conversion po-

tential. In addition, the iTF-seq approach provides a powerful tool for investigating the molecular mechanisms of cell fate conversion induced by TFs as we found that GCM1 and OTX2 act as pioneer factors to induce cell fate conversion by directly activating lineage-specific genes residing in regions of closed chromatin. Future studies using iTF-seq could focus on identifying additional TFs involved in cell fate conversion and investigating their specific mechanisms of action, such as potential roles as repressors or dual-function TFs. Additionally, mapping the interaction partner proteins of the TFs with the iTF-seq pipeline can provide further insights into their unique features and modes of action. Overall, iTF-seq has the potential to significantly advance our understanding of the complex molecular events underlying cell fate conversion.

Many previous studies have used multiple TFs for cell fate reprogramming. However, emerging research revealed that even a single TF can induce cell fate conversion toward a specific lineage or cell type, including recent large-scale screens (Parekh et al. 2018; Joung et al. 2023). By using appropriate culture conditions, like the somatic cell reprogramming case, a single TF-mediated cell fate change can be further polished for generating functional cell types. These findings highlight the merits of a single-factor system, particularly when integrated with downstream approaches aimed at understanding the action mechanisms of potent individual TFs, as elucidated in this study. Testing multiple TFs based on
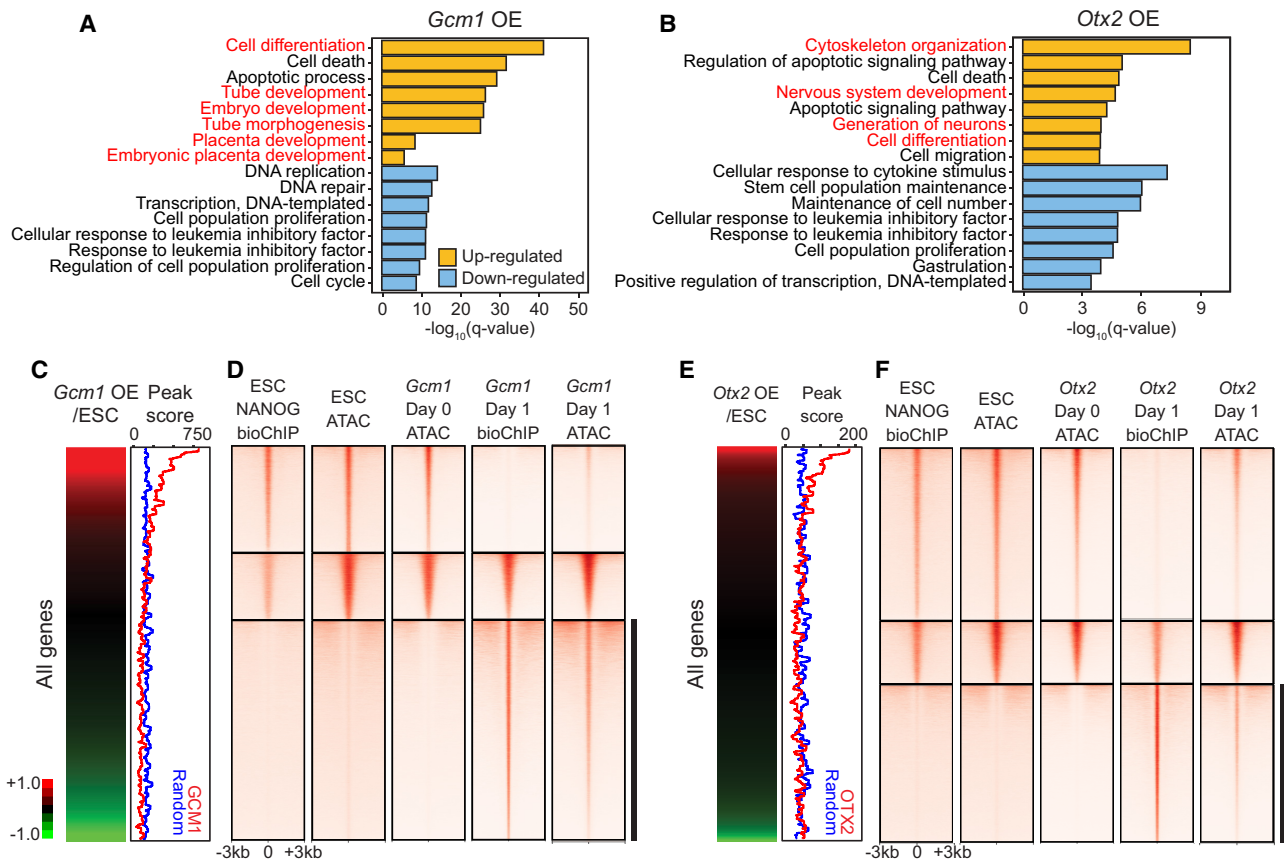
**Figure 6.** GCM1 and OTX2 function as activators and pioneer factors. (*A,B*) Selected GO biological process terms significantly enriched among differentially expressed genes up-regulated or down-regulated in *Gcm1*-overexpressing cells (*A*) or *Otx2*-overexpressing cells (*B*). (*C,E, left*) Expression profiles of *Gcm1*-overexpressing cells (*C*) or *Otx2*-overexpressing cells (*E*) with genes ordered by their expression levels in *Gcm1/Otx2*-overexpressing cells relative to the control ES cells. (*Right*) Corresponding peak scores to the expression profiles (red) and randomized peak scores (blue) are shown by applying moving window average (window size, 100; bin size, one). (*D,F*) bioChIP-seq of indicated TFs and ATAC-seq signals detected in control ES cells, −Dox cells (day 0), and *Gcm1*-overexpressing cells (*D*) or *Otx2*-overexpressing cells (*F*; day 1). Black bars indicate the regions closed in ES cells or day 0 cells but occupied by GCM1 (*D*) or OTX2 (*F*) and become open.

the results obtained from a single TF screening could enhance the efficiency of cell fate conversion.

Our observations on the decreased proportion of cells expressing TFs capable of triggering cell fate changes during prolonged induction are consistent with previous reports on the differences in cell cycle phase duration and cell death between pluripotent stem cells and differentiated cells. It is important to note that 20%–30% of mouse ES cells undergo cell death upon exit from self-renewal (Duval et al. 2000; Bashamboo et al. 2006), which may also contribute to cell fate conversion upon TF induction. In addition, we observed that the decreased expression levels of pluripotency markers persisted in cells overexpressing the iTFs, such as *Yy1*, *Creb3l2*, *Cdx2*, *Ets2*, *Gsc*, and *Cebpb*, which showed cell fate–changing effects only evident in day 1 results. This observation suggests that the loss of these cells on days 3 and 5 did not result from a reversion to ES cell–like states over time but may stem from other reasons. We confirmed that multiple iTF-triggered differentiated cells show increased cell death. Therefore, we conclude that both the slower cell cycle and enhanced cell death upon differentiation are responsible for the decreased proportion of cells observed during more prolonged induction of potent TFs. Our discoveries underscore the significance of considering cell cycle and cell death in perturbation screens related to the differentiation of

ES cells. This aspect has not been addressed in previous screens that lacked monitoring dynamic changes in cellular status through an inducible system. This understanding will be essential in designing more effective screening strategies for identifying potent TFs capable of inducing cell fate conversion.

The SB approach induces a spectrum of ectopic TF expression levels and a heterogeneous cell population, which may affect cell fate–changing efficiency. We observed a range of TF expressions among iTF overexpression cells (Supplemental Fig. S8A). The variability in ectopic TF levels may be attributed to the number of inserted iTFs and potential positional effects. In our observations, cells showing higher ectopic TF expressions (top 33%) tended to display relatively lower expression of ES cell markers (*Pou5f1*, *Sox2*, and *Nanog*), whereas cells with lower TF expression (bottom 33%) showed somewhat higher levels of ES cell marker genes compared with cells expressing high levels of iTF. It is noteworthy that in both cases, the levels of ES cell markers were lower than those in undifferentiated control ES cells, indicating the ES cell differentiation. Although we acknowledge that this variability in ectopic TF levels introduces noise into the system, we recognize its potential to offer valuable insights into TF dosage effects. We unexpectedly observed that the induction levels of *Gata3* influence lineage specification decisions toward trophoblast or primitive endoderm

lineages. Cells with high *Gata3* expression displayed trophoblast-like gene expression programs, whereas those with low *Gata3* expression expressed primitive endoderm-like gene expression patterns (Supplemental Fig. S8B). Consequently, although our system may generate complex expression patterns contingent on the levels of ectopic TFs, we posit that these resultant patterns may offer an additional layer of information for a deeper understanding of TF dosage effects in cell fate conversion, a facet not easily gleaned from other methodologies.

Although some tissue/cell type–specific TFs activated the expected tissue/cell type–specific genes upon induction, we also observed that several TFs activated various lineages or a broad spectrum of gene sets. Notably, we found that the induction of multiple family TFs resulted in similar transcriptome changes (Fig. 5; Supplemental Fig. S6B,C). Especially, ectopic induction of all tested *Jun/Fos* family TFs increased the expression levels of marker genes associated with multiple lineages, although each factor showed distinct kinetics. Although FOS and JUN showed rapid activation of lineage marker genes, with expression evident in day 1 samples, *Fosl1*- and *Fosl2*-overexpressing cells displayed a noticeable activation of lineage marker genes after 3 d of induction. Therefore, although their potency and target activation timing may vary, the immediate early genes, *Jun* and *Fos*, along with their family proteins, *Fosl1* and *Fosl2*, activated more than two lineages, suggesting that these TFs can activate multiple lineages (Supplemental Fig. S6D). Although not tested, our current experimental setting suggests that multiple family TFs generating similar transcriptome changes might have similar chromosomal target genes. As each TF within the same family often shows distinct cell type–specific expression patterns, our approach may help to elucidate the context-dependent roles of family TFs by testing additional cell types. Investigating whether these TFs occupy similar genomic targets would be of great interest in understanding the potential and general behavior of individual family TFs.

Our study showed that both GCM1 and OTX2 possess pioneer factor activity by binding closed chromatin in ES cells, leading to the opening of the target loci. It remains to be tested whether other potent TFs we identified also have pioneer activity and can contribute to generating open chromatin. Understanding the requirement of pioneer factor activity for TF-mediated cell fate conversion will be essential to uncover the underlying mechanisms of this process. Although GCM1 and OTX2 induced distinct gene expression profiles upon induction, it is possible that they share common interaction partner proteins, such as the factors involved in enhancer regulation, enhancer-related histone signatures, and chromatin remodeling, in addition to their unique interaction partner proteins. Therefore, identifying the interaction partner proteins of potent TFs will be crucial to understand both general and unique mechanisms of TF-mediated cell fate conversion. Our iTF-seq approach, which is also compatible with the biotinylation-mediated immunoprecipitation followed by mass spectrometry (bioIP/MS) approach (Kim et al. 2009), will provide a powerful platform to identify TFs capable of cell fate changes and to gain insights into the mechanisms of TF-mediated cell fate conversion. This approach holds great potential for generating desired cell types and for advancing our understanding of cell fate conversion processes.

Although we successfully identified potent cell fate–changing TFs that showed rapid effects based on day 1 results, the current method also presents limitations concerning the availability of informative cells (differentiated) at later time points, necessitating significantly more sequencing depth. Therefore, incorporating a depletion step to exclude undifferentiated cells, such as by using SSEA-1 expression, could enhance the efficiency of identifying potent cell fate–changing TFs using iTF-seq. Additionally, the SB transposon system integrates varying numbers of iTFs into each cell, resulting in a spectrum of TF expression levels among iTF lines. Although this variability in ectopic TF levels introduces noise into the system, it simultaneously provides valuable insights into TF dosage effects.

## Methods

### Cell culture

All the ES cell lines, including BirA-expressing mouse J1 ES cells (BirA-ES cells) and stable inducible lines, were cultured on 0.1% gelatin-coated plates in Dulbecco's Modified Eagle Medium (DMEM) (high glucose, Gibco 11965092) supplemented with 18% fetal bovine serum (Gemini Bio-Products 100106), 100 µM MEM nonessential amino acids (100X stock, Gibco 11140050), EmbryoMax nucleosides (MilliporeSigma ES-008-D), 2 mM L-glutamine, 50 U/L penicillin–streptomycin (100X stock, Gibco 10378016), 100 µM β-mercaptoethanol, and 1000 U/mL mouse LIF (Gemini Bio-Products 400–495) in a 37°C incubator with 5% $CO_2$. Cells were passaged every 2 d by dissociating into single cells with 0.25% trypsin-EDTA.

### Construction of pSBFB plasmid and generation of stable inducible overexpression cell lines

The luciferase ORF of pSBtet-GP (Addgene 60495) was replaced by a FLAG sequence, biotinylation site, and multiple cloning sites with NotI and NheI sequences (pSBFB). For cloning, total RNA was extracted from mouse J1 ES cells using the RNeasy plus mini kit (Qiagen 74136). Then, the RNA was reverse-transcribed with the ProtoScript II first strand cDNA synthesis kit (New England Biolabs [NEB] E6560). Each gene of interest was amplified from the cDNA and cut with NotI and/or NheI depending on the presence of respective cut sites within the ORF. Cloning was performed using T4 DNA ligase (NEB M0202) using pSBFB digested with NotI and/or NheI enzymes. The cloned vectors, verified with Sanger sequencing, were transfected (lipofectamine 3000, Thermo Fisher Scientific L3000008) into BirA-ES cells (Kim et al. 2008) along with pCMV(CAT)T7-SB100 (Addgene 34879), conferring transient expression of the SB transposase for inserting the target sequence into the genome of BirA-ES cells. To obtain stable cell lines, the transduced cells were selected using puromycin (1 µg/mL, Gibco A1113803) and G418 (250 µg/mL, Gibco 10131027) at 24 h after transfection. Protein expression of the transfected gene of each cell line was confirmed with western blotting after treatment with Dox (Fisher BioReagents BP26535) for 24 h at a concentration of 0.5 µg/mL.

### Western blotting

Dox-treated or untreated cells were lysed using Laemmli sample buffer (Bio-Rad 1610747) supplemented with 5% β-mercaptoethanol (MilliporeSigma M3148) and heated for 10 min at 95°C. The cell lysate was run on sodium dodecyl sulfate (SDS) polyacrylamide gel and transferred onto polyvinylidene fluoride (PVDF) membranes (MilliporeSigma IPVH00010). Then, membranes were blocked with either 5% nonfat milk or BSA in TBS-T (Tris-buffered saline with 0.1% Tween-20) for 1 h. Subsequently, the membranes were incubated with streptavidin–horseradish peroxidase conjugates (streptavidin-HRP, Cytiva RPN1231, 1:2000) or ACTB antibody (Abgent AM1829B, 1:20,000). For ACTB detection, the

membrane was incubated with a secondary antibody (Cell Signaling Technology 7076, 1:10,000) for 1 h at room temperature. The membrane was then exposed to ECL substrate (Cytiva RPN2232), and visualized on a ChemiDoc XRS+ imaging system (Bio-Rad).

## RT-qPCR

Total RNA was extracted using RNeasy plus mini kit (250, Qiagen 74136). For each sample, 500 ng RNA was reverse-transcribed with qScript cDNA supermix (Quantabio 95048-100). cDNA was diluted into 200 μL, and qPCR was performed using PerfeCTa SYBR Green FastMix (QuantaBio 95072-012). The primers used for qPCR are listed in Supplemental Table S7. Results were calculated from three independent samples (n = 3), and statistical significance was determined using two-tailed Student's *t*-test analysis.

## Immunofluorescence

Immunofluorescence was performed on *Gcm1* and *Otx2* SBFB cells after 24 h of Dox induction with uninduced control cells. Cells were washed twice with PBS and fixed with freshly cracked 4% paraformaldehyde in PBS for 20 min at room temperature. After washing three times with wash buffer (0.1% BSA in PBS), the cells were blocked in blocking buffer (PBS containing 0.1% BSA, 10% horse serum, and 0.3% Triton X-100) for 45 min. The cells were then incubated with streptavidin, Alexa Fluor 555 conjugate (Thermo Fisher Scientific S21381; diluted at 1:500 in PBS with 0.1% BSA and 10% horse serum) overnight at 4°C, followed by three washes with wash buffer. After incubation with DAPI for 5–10 min, the cells were rewashed and visualized under a fluorescence microscope.

## Pooled induction of target genes and scRNA-seq

In total, 80 iTF lines that were confirmed to express the protein of interest were pooled and cultured in ES cell media. Pooled cell lines were then treated with Dox (0.5 μg/mL) for 1, 3, and 5 d. At each time point, Dox-treated and control ES cells were dissociated into single cells with 0.25% trypsin-EDTA. The dissociated cells were washed twice with 0.04% BSA in PBS and resuspended to a concentration of 1000 cells per microliter. Then the cells were subjected to scRNA-seq. Single-cell libraries were generated using the chromium next GEM single-cell 3′ library kit v3.1 (10x Genomics), and the libraries were sequenced on a NovaSeq 6000 (Illumina).

## Sequencing read alignment and cell quality control

Sequencing reads were aligned with Cell Ranger v6.0.1 (Zheng et al. 2017). The mouse 2020-A reference built by 10x Genomics, which is based on Ensembl release 98 (Yates et al. 2020) GRCm38 genome assembly and GENCODE M23 annotation (Frankish et al. 2019), was used as a reference sequence. Cell quality control was performed with Seurat v4.0.3 (Hao et al. 2021). Based on plots illustrating the distribution of quality-control metrics, the criteria for filtering of cells were determined: A cell should have more than 1000 unique detected genes, with <10% of reads mapped to the mitochondrial genome, and more than 10,000 detected molecules. After filtering, 65%–77% of cells were selected (R Core Team 2021).

## Identification of cells overexpressing TFs

We defined a forward tag as 21 bp from a TF's start codon and defined its upstream tag as 7 bp from the vector. For the reverse tag, we concatenated a TF's last 21 bp (including the stop codon) with the downstream 7 bp from the vector and took its reverse complement. These tags were unique for each TF and allowed us to distinguish between the wild type and induced version of each gene. The only exception was the Dlx4 forward tag, which was 30 bp long because the 28-bp tag completely aligned to the reference genome.

BAM files were converted to FASTA files using SAMtools v1.7 (Li et al. 2009). Reads without either cell barcodes or UMIs were removed. Then, BLAST was performed to detect reads aligned with the tags. Collected reads with the tags were organized to the transcript level. The alignment results based on our tags were also compared with the alignment by Cell Ranger. When more than half of a transcript's reads aligned to two different genes by the two methods, that transcript was excluded. We did not exclude the cases in which a gene was annotated as "NA" by Cell Ranger but aligned with our tags and used the alignment results for further analysis.

## Adjustment of minimum UMI thresholds and cell selection

Because the iTF lines were generated individually, the number of unique iTFs expressed in a cell is theoretically one. However, errors from sequencing and library preparation steps such as polymerase chain reaction (PCR) can introduce errors that may lead to incorrect iTF prediction (Stoler and Nekrutenko 2021). To reduce false-positive events in detecting iTFs, we exploited UMIs. We applied different thresholds for the minimum number of UMI (one to five) and counted cells with single iTF detection. As the number of cells with a single iTF peaked in the minimum UMI 3 threshold, this threshold was adopted to designate iTF cells. The control cells were identified by excluding cells expressing any iTFs with a minimum UMI threshold of one.

## UMAP visualization of iTF-seq data

Uniform manifold approximation and projection (UMAP) plots were drawn with cells overexpressing a single iTF and control cells using Seurat v4.0.3 (Hao et al. 2021). Expression values were normalized and variance-stabilized by SCTransform (Hafemeister and Satija 2019). The location of whole cells was visualized, and TF-overexpressing cells for each TF and control cells were marked.

## Pseudobulk RNA-seq processing

Pseudobulk RNA-seq expression values were calculated by taking the mean of gene expression values for every gene from iTF-overexpressing cells and control cells, respectively. Log-normalized versions of corrected counts by SCTransform (Hafemeister and Satija 2019) were used (Supplemental Table S4).

## Calculation of differentiation index

First, because gene expression values are not independent of each other, we performed principal component analysis (PCA) after normalizing, variance-stabilizing, and estimating missing values with SCTransform. Fifty principal components (PCs) were used because the variance explained by each PC becomes very small around the 50th PC. The centroid of control cells based on the PCs was calculated; distances between the centroid and each control cell were measured; and these distances were converted into *z*-values. Then, distances between TF-overexpressing cells and the centroid of control cells were measured and normalized using the mean and standard deviation from control cells. For these calculations, gene expression measurements for the 80 TFs were excluded.

In addition to all genes, for some analyses, differentiation indexes were calculated with other specific gene sets: core module genes, PRC module genes (Kim et al. 2010), and 643 gene sets from the MSigDB (Subramanian et al. 2005; Liberzon et al.

2011). The 80 TFs of interest were excluded from all gene sets, and ribosomal genes (Nakao et al. 2004) were additionally excluded from the two module gene sets. For gene sets from MSigDB, we chose three germ layer–related gene sets (total, eight: GO:0007398, GO:0007492, GO:0001706, GO:0035987, GO:0007498, GO:0048332, GO:0048333, and GO:0031016) from C2 curated and C5 ontology gene set collection and all the gene sets (total, 704) from the C8 cell type signature gene set collection. Mouse orthologs of human genes were listed using Ensembl BioMart (Kinsella et al. 2011), and gene sets whose size was fewer than 15 genes were filtered out. The median values of differentiation indexes were used for the hierarchical clustering of all available pairs of gene sets and iTFs (Supplemental Table S6).

### Analysis of differentially expressed genes

Global gene expression patterns of TF-overexpressing cells at each time point to their respective control cells were compared by using Seurat v4.0.3. We computed differential expression with DESeq2 (Love et al. 2014) based on raw expression values, excluding cases when the number of TF-overexpressing cells was fewer than three. Differentially expressed genes were filtered with these criteria: average $|\log_2(\text{Fold Change})| \geq 0.5$, and Bonferroni-corrected $P$-value < 0.05. iTFs themselves were not counted as up-regulated genes. The GO term enrichment tests were performed with the gprofiler2 R package (Kolberg et al. 2020).

### RNA sequencing and data analysis

Total RNA extraction was performed using RNeasy plus mini kit (250) (Qiagen 74136), and mRNA isolation was accomplished using magnetic mRNA isolation kit (oligo(dT) beads; NEB E7490). The sequencing libraries were prepared using NEBNext ultra DNA library prep kit (NEB E7645S) and were sequenced on an Illumina NovaSeq 6000 using 150-bp paired-end reads. Adapter/quality trimming was performed with Trim Galore! v0.6.10 (Babraham Bioinformatics; https://github.com/FelixKrueger/TrimGalore). Quantification of transcript expressions was performed with Salmon v1.10.2 (Patro et al. 2017), and aggregation to the gene level was performed with the R package tximport v1.22.0 (Soneson et al. 2015).

### Cell proliferation assay

iTF lines were seeded at a density of $6 \times 10^4$ cells/mL in the indicated ES cell media in 0.1% gelatin-coated 24-well plates. Dox was treated 24 h after seeding at a concentration of 0.5 µg/mL. Cells were counted at four time points: on the day of Dox induction and on days 2, 4, and 6 of Dox induction, using the Invitrogen countess automated cell counter (Invitrogen). Briefly, cells were trypsinized with 0.25% trypsin-EDTA, and then, 10 µL of the cell suspension was mixed with 10 µL of 0.4% trypan blue and counted by the cell counter.

### Cell death assay

iTF lines were seeded at a density of $5 \times 10^5$ cells/mL in the indicated ES cell media in 0.1% gelatin-coated 96-well plates. Dox was treated 24 h after seeding at a concentration of 0.5 µg/mL. After 2 d of Dox induction, lactate dehydrogenase (LDH) activity was quantified using the CyQUANT LDH cytotoxicity assay kit (Thermo Fisher Scientific C20301) according to the manufacturer's instruction. The 10× lysis buffer and nuclease-free water were added directly to the media and incubated at 37°C. After 45 min, 50 µL of the media was transferred to a new 96-well plate containing an equal volume of reaction mix. The plates were incubated for

30 min at room temperature in the dark. Stop buffer was added to each well, and the final absorbance was measured at 490 nm with the Tecan M1000 plate reader. The results were normalized to wells that had been treated with lysis buffer, providing the maximum LDH activity.

### bioChIP-seq and ATAC-seq

bioChIP-seq was performed on *Gcm1* and *Otx2* iTF lines with or without Dox induction, as well as ES cells expressing biotinylated NANOG. For each bioChIP reaction, $5 \times 10^6$ cells were fixed using 1% formaldehyde for 7 min at room temperature, and then the fixation was stopped by the addition of glycine to a final concentration of 0.125 M. The bioChIP was performed as described previously (Kim et al. 2009). The genomic DNA was sonicated into fragments of ~200 bp using a Bioruptor (Diagenode). The cross-linked DNA fragments were precleared using Protein A agarose (Sigma-Aldrich 11134515001) and immunoprecipitated using Dynabeads MyOne streptavidin T1 (Invitrogen 65602). The sequencing libraries were prepared using the NEBNext ultra DNA library prep kit (NEB E7645S) and were sequenced on an Illumina NovaSeq 6000 using 50-bp paired-end reads. An assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) was performed on iTF lines for *Gcm1* and *Otx2* under uninduced or 1-d-induced conditions. ATAC-seq libraries were generated using Diagenode's ATAC-seq kit (C01080001) followed by sequencing on Illumina NovaSeq 6000 using 50-bp paired-end reads.

### bioChIP-seq and ATAC-seq data analysis

Published ES cell ATAC-seq data were downloaded from the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSM2412020. The FASTQ files of bioChIP-seq and ATAC-seq were trimmed with Trim Galore! v0.6.7 and mapped to the mm10 mouse reference genome using Bowtie 2 v2.4.4 with default parameters (Langmead and Salzberg 2012; https://github.com/FelixKrueger/TrimGalore). All alignments were further sorted by Picard Tools v2.27.2 and filtered by SAMtools v1.14 with MAPQ ≥ 10 (Broad Institute) (Li et al. 2009; http://broadinstitute.github.io/picard). The PCR duplicates were removed using Picard Tools v2.27.2. Peaks were identified using MACS2 v2.2.7.1 with a cutoff of FDR < 0.05 and annotated with HOMER v4.11 (Zhang et al. 2008; Heinz et al. 2010). For visualization, RPGC normalized bigWig files were generated using deepTools bamcoverage v3.5.0 (Ramírez et al. 2014).

## Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE218628 and GSE220724. The codes used in this study are available at GitHub (https://github.com/marcottelab/iTF-seq) and as Supplemental Code.

## Competing interest statement

The authors declare no competing interests.

## Acknowledgments

## References

Aarts M, Georgilis A, Beniazza M, Beolchi P, Banito A, Carroll T, Kulisic M, Kaemena DF, Dharmalingam G, Martin N, et al. 2017. Coupling shRNA screens with single-cell RNA-seq identifies a dual role for mTOR in reprogramming-induced senescence. *Genes Dev* **31:** 2085–2098. doi:10.1101/gad.297796.117

Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, Chen Y, Villalta JE, Gilbert LA, Horlbeck MA, Hein MY, et al. 2016. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167:** 1867–1882.e21. doi:10.1016/j.cell.2016.11.048

Bashamboo A, Taylor AH, Samuel K, Panthier JJ, Whetton AD, Forrester LM. 2006. The survival of differentiating embryonic stem cells is dependent on the SCF-KIT pathway. *J Cell Sci* **119:** 3039–3046. doi:10.1242/jcs.03038

Bernardo AS, Cho CH-H, Mason S, Docherty HM, Pedersen RA, Vallier L, Docherty K. 2009. Biphasic induction of Pdx1 in mouse and human embryonic stem cells can mimic development of pancreatic β-cells. *Stem Cells* **27:** 341–351. doi:10.1634/stemcells.2008-0310

Blij S, Parenti A, Tabatabai-Yazdi N, Ralston A. 2015. *Cdx2* efficiently induces trophoblast stem-like cells in naïve, but not primed, pluripotent stem cells. *Stem Cells Dev* **24:** 1352–1365. doi:10.1089/scd.2014.0395

Cho LT, Wamaitha SE, Tsai IJ, Artus J, Sherwood RI, Pedersen RA, Hadjantonakis AK, Niakan KK. 2012. Conversion from mouse embryonic to extra-embryonic endoderm stem cells reveals distinct differentiation capacities of pluripotent stem cell states. *Development* **139:** 2866–2877. doi:10.1242/dev.078519

Choi J, Costa ML, Mermelstein CS, Chagas C, Holtzer S, Holtzer H. 1990. MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. *Proc Natl Acad Sci USA* **87:** 7988–7992. doi:10.1073/pnas.87.20.7988

Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Klughammer J, Schuster LC, Kuchler A, Alpar D, Bock C. 2017. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* **14:** 297–301. doi:10.1038/nmeth.4177

Denham M, Thompson LH, Leung J, Pébay A, Björklund A, Dottori M. 2010. Gli1 is an inducing factor in generating floor plate progenitor cells from human embryonic stem cells. *Stem Cells* **28:** 1805–1815. doi:10.1002/stem.510

Devary Y, Gottlieb RA, Lau LF, Karin M. 1991. Rapid and preferential activation of the c-*jun* gene during the mammalian UV response. *Mol Cell Biol* **11:** 2804–2811. doi:10.1128/mcb.11.5.2804-2811.1991

Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, et al. 2016. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167:** 1853–1866.e17. doi:10.1016/j.cell.2016.11.038

Duval D, Reinhardt B, Kedinger C, Boeuf H. 2000. Role of suppressors of cytokine signaling (Socs) in leukemia inhibitory factor (LIF)-dependent embryonic stem cell survival. *FASEB J* **14:** 1577–1584. doi:10.1096/fj.99-0810com

Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, et al. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47:** D766–D773. doi:10.1093/nar/gky955

Franzén O, Gan LM, Björkegren JLM. 2019. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)* **2019:** baz046. doi:10.1093/database/baz046

Fujikura J, Yamato E, Yonemura S, Hosoda K, Masui S, Nakao K, Miyazaki Ji J, Niwa H. 2002. Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev* **16:** 784–789. doi:10.1101/gad.968802

Hackett PB, Largaespada DA, Cooper LJ. 2010. A transposon and transposase system for human application. *Mol Ther* **18:** 674–683. doi:10.1038/mt.2010.2

Hafemeister C, Satija R. 2019. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20:** 296. doi:10.1186/s13059-019-1874-1

Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F, et al. 2018. Mapping the mouse cell atlas by Microwell-seq. *Cell* **172:** 1091–1107.e17. doi:10.1016/j.cell.2018.02.001

Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. 2021. Integrated analysis of multimodal single-cell data. *Cell* **184:** 3573–3587.e29. doi:10.1016/j.cell.2021.04.048

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38:** 576–589. doi:10.1016/j.molcel.2010.05.004

Holtzinger A, Rosenfeld GE, Evans T. 2010. *Gata4* directs development of cardiac-inducing endoderm from ES cells. *Dev Biol* **337:** 63–73. doi:10.1016/j.ydbio.2009.10.003

Hwang B, Lee JH, Bang D. 2018. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* **50:** 1–14. doi:10.1038/s12276-018-0071-8

Ivics Z, Hackett PB, Plasterk RH, Izsvák Z. 1997. Molecular reconstruction of *Sleeping Beauty*, a *Tc1*-like transposon from fish, and its transposition in human cells. *Cell* **91:** 501–510. doi:10.1016/S0092-8674(00)80436-5

Iwafuchi-Doi M, Zaret KS. 2016. Cell fate control by pioneer transcription factors. *Development* **143:** 1833–1837. doi:10.1242/dev.133900

Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, Salame TM, Tanay A, van Oudenaarden A, Amit I. 2016. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* **167:** 1883–1896.e15. doi:10.1016/j.cell.2016.11.039

Joung J, Ma S, Tay T, Geiger-Schuller KR, Kirchgatterer PC, Verdine VK, Guo B, Arias-Garcia MA, Allen WE, Singh A, et al. 2023. A transcription factor atlas of directed differentiation. *Cell* **186:** 209–229.e26. doi:10.1016/j.cell.2022.11.026

Keller GM. 1995. In vitro differentiation of embryonic stem cells. *Curr Opin Cell Biol* **7:** 862–869. doi:10.1016/0955-0674(95)80071-9

Kim J, Chu J, Shen X, Wang J, Orkin SH. 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132:** 1049–1061. doi:10.1016/j.cell.2008.02.039

Kim J, Cantor AB, Orkin SH, Wang J. 2009. Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. *Nat Protoc* **4:** 506–517. doi:10.1038/nprot.2009.23

Kim J, Woo AJ, Chu J, Snow JW, Fujiwara Y, Kim CG, Cantor AB, Orkin SH. 2010. A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* **143:** 313–324. doi:10.1016/j.cell.2010.09.010

Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, et al. 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* **2011:** bar030. doi:10.1093/database/bar030

Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. 2020. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res* **9:** 709. doi:10.12688/f1000research.24956.2

Kulessa H, Frampton J, Graf T. 1995. GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. *Genes Dev* **9:** 1250–1262. doi:10.1101/gad.9.10.1250

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359. doi:10.1038/nmeth.1923

Latos PA, Sienerth AR, Murray A, Senner CE, Muto M, Ikawa M, Oxley D, Burge S, Cox BJ, Hemberger M. 2015. Elf5-centered transcription factor hub controls trophoblast stem cell self-renewal and differentiation through stoichiometry-sensitive shifts in target gene networks. *Genes Dev* **29:** 2435–2448. doi:10.1101/gad.268821.115

Lee BK, Uprety N, Jang YJ, Tucker SK, Rhee C, LeBlanc L, Beck S, Kim J. 2018. Fosl1 overexpression directly activates trophoblast-specific gene expression programs in embryonic stem cells. *Stem Cell Res* **26:** 95–102. doi:10.1016/j.scr.2017.12.004

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079. doi:10.1093/bioinformatics/btp352

Li H, Huang Q, Liu Y, Garmire LX. 2020. Single cell transcriptome research in human placenta. *Reproduction* **160:** R155–R167. doi:10.1530/REP-20-0231

Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. 2011. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27:** 1739–1740. doi:10.1093/bioinformatics/btr260

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550. doi:10.1186/s13059-014-0550-8

Madissoon E, Wilbrey-Clark A, Miragaia RJ, Saeb-Parsy K, Mahbubani KT, Georgakopoulos N, Harding P, Polanski K, Huang N, Nowicki-Osuch K, et al. 2020. scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biol* **21:** 1. doi:10.1186/s13059-019-1906-x

Morris SA. 2016. Direct lineage reprogramming via pioneer factors; a detour through developmental gene regulatory networks. *Development* **143:** 2696–2705. doi:10.1242/dev.138263

Nakao A, Yoshihama M, Kenmochi N. 2004. RPG: the ribosomal protein gene database. *Nucleic Acids Res* **32:** D168–D170. doi:10.1093/nar/gkh004

Nakatake Y, Ko SBH, Sharov AA, Wakabayashi S, Murakami M, Sakota M, Chikazawa N, Ookura C, Sato S, Ito N, et al. 2020. Generation and profiling of 2135 human ESC lines for the systematic analyses of cell states perturbed by inducing single transcription factors. *Cell Rep* **31:** 107655. doi:10.1016/j.celrep.2020.107655

Ng AHM, Khoshakhlagh P, Rojo Arias JE, Pasquini G, Wang K, Swiersy A, Shipman SL, Appleton E, Kiaee K, Kohman RE, et al. 2021. A comprehensive library of human transcription factors for cell fate engineering. *Nat Biotechnol* **39:** 510–519. doi:10.1038/s41587-020-0742-6

Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, Yagi R, Rossant J. 2005. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123:** 917–929. doi:10.1016/j.cell.2005.08.040

Ofengeim D, Giagtzoglou N, Huh D, Zou C, Yuan J. 2017. Single-cell RNA sequencing: unraveling the brain one cell at a time. *Trends Mol Med* **23:** 563–576. doi:10.1016/j.molmed.2017.04.006

Ong CT, Corces VG. 2012. Enhancers: emerging roles in cell fate specification. *EMBO Rep* **13:** 423–430. doi:10.1038/embor.2012.52

Paik DT, Cho S, Tian L, Chang HY, Wu JC. 2020. Single-cell RNA sequencing in cardiovascular development, disease and medicine. *Nat Rev Cardiol* **17:** 457–473. doi:10.1038/s41569-020-0359-y

Papalexi E, Satija R. 2018. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol* **18:** 35–45. doi:10.1038/nri.2017.76

Parekh U, Wu Y, Zhao D, Worlikar A, Shah N, Zhang K, Mali P. 2018. Mapping cellular reprogramming via pooled overexpression screens with paired fitness and single-cell RNA-sequencing readout. *Cell Syst* **7:** 548–555.e8. doi:10.1016/j.cels.2018.10.008

Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, Lerou PH, Lensch MW, Daley GQ. 2008. Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451:** 141–146. doi:10.1038/nature06534

Patel LR, Curran T, Kerppola TK. 1994. Energy transfer analysis of Fos-Jun dimerization and DNA binding. *Proc Natl Acad Sci USA* **91:** 7360–7364. doi:10.1073/pnas.91.15.7360

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14:** 417–419. doi:10.1038/nmeth.4197

Ralston A, Cox BJ, Nishioka N, Sasaki H, Chea E, Rugg-Gunn P, Guo G, Robson P, Draper JS, Rossant J. 2010. Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development* **137:** 395–403. doi:10.1242/dev.038828

Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42:** W187–W191. doi:10.1093/nar/gku365

R Core Team. 2021. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

Rhee C, Lee BK, Beck S, Anjum A, Cook KR, Popowski M, Tucker HO, Kim J. 2014. Arid3a is essential to execution of the first cell fate decision via direct embryonic and extraembryonic transcriptional regulation. *Genes Dev* **28:** 2219–2232. doi:10.1101/gad.247163.114

Ruan Z, Zhu L, Yin Y, Chen G. 2016. Overexpressing NKx2.5 increases the differentiation of human umbilical cord drived mesenchymal stem cells into cardiomyocyte-like cells. *Biomed Pharmacother* **78:** 110–115. doi:10.1016/j.biopha.2016.01.020

Ruiz-Reig N, Rakotobe M, Bethus I, Le Menn G, Huditz HI, Marie H, Lamonerie T, D'Autréaux F. 2019. Developmental requirement of homeoprotein Otx2 for specific habenulo-interpeduncular subcircuits. *J Neurosci* **39:** 1005–1019. doi:10.1523/JNEUROSCI.1818-18.2018

Shiau F, Ruzycki PA, Clark BS. 2021. A single-cell guide to retinal development: cell fate decisions of multipotent retinal progenitors in scRNA-seq. *Dev Biol* **478:** 41–58. doi:10.1016/j.ydbio.2021.06.005

Smith AG. 2001. Embryo-derived stem cells: of mice and men. *Annu Rev Cell Dev Biol* **17:** 435–462. doi:10.1146/annurev.cellbio.17.1.435

Soneson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4:** 1521. doi:10.12688/f1000research.7563.1

Stecca B, Nait-Oumesmar B, Kelley KA, Voss AK, Thomas T, Lazzarini RA. 2002. Gcm1 expression defines three stages of chorio-allantoic interaction during placental development. *Mech Dev* **115:** 27–34. doi:10.1016/S0925-4773(02)00095-3

Stoler N, Nekrutenko A. 2021. Sequencing error profiles of Illumina sequencing instruments. *NAR Genom Bioinform* **3:** lqab019. doi:10.1093/nargab/lqab019

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102:** 15545–15550. doi:10.1073/pnas.0506580102

Svensson V, Vento-Tormo R, Teichmann SA. 2018. Exponential scaling of single-cell RNA-seq in the past decade. *Nat Protoc* **13:** 599–604. doi:10.1038/nprot.2017.149

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126:** 663–676. doi:10.1016/j.cell.2006.07.024

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131:** 861–872. doi:10.1016/j.cell.2007.11.019

Tan Y, Cahan P. 2019. SingleCellNet: a computational tool to classify single cell RNA-seq data across platforms and across species. *Cell Syst* **9:** 207–213.e2. doi:10.1016/j.cels.2019.06.004

Toyooka Y. 2021. Pluripotent stem cells in the research for extraembryonic cell differentiation. *Dev Growth Differ* **63:** 127–139. doi:10.1111/dgd.12716

Tseng KY, Chen YH, Lin S. 2017. Zinc finger protein ZFP36L1 promotes osteoblastic differentiation but represses adipogenic differentiation of mouse multipotent cells. *Oncotarget* **8:** 20588–20601. doi:10.18632/oncotarget.15246

Wamaitha SE, del Valle I, Cho LT, Wei Y, Fogarty NM, Blakeley P, Sherwood RI, Ji H, Niakan KK. 2015. Gata6 potently initiates reprograming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. *Genes Dev* **29:** 1239–1255. doi:10.1101/gad.257071.114

Xu Y, Luo X, Fang Z, Zheng X, Zeng Y, Zhu C, Gu J, Tang F, Hu Y, Hu G, et al. 2018. Transcription coactivator Cited1 acts as an inducer of trophoblast-like state from mouse embryonic stem cells through the activation of BMP signaling. *Cell Death Dis* **9:** 924. doi:10.1038/s41419-018-0991-1

Yamamoto H, Kon T, Omori Y, Furukawa T. 2020. Functional and evolutionary diversification of *Otx2* and *Crx* in vertebrate retinal photoreceptor and bipolar cell development. *Cell Rep* **30:** 658–671.e5. doi:10.1016/j.celrep.2019.12.072

Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, et al. 2020. Ensembl 2020. *Nucleic Acids Res* **48:** D682–D688. doi:10.1093/nar/gkz1138

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, et al. 2007. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318:** 1917–1920. doi:10.1126/science.1151526

Zaveri L, Dhawan J. 2018. Cycling to meet fate: connecting pluripotency to the cell cycle. *Front Cell Dev Biol* **6:** 57. doi:10.3389/fcell.2018.00057

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9:** R137. doi:10.1186/gb-2008-9-9-r137

Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, Luo T, Xu L, Liao G, Yan M, et al. 2019. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res* **47:** D721–D728. doi:10.1093/nar/gky900

Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8:** 14049. doi:10.1038/ncomms14049

# Systematic mapping of TF-mediated cell fate changes by a pooled induction coupled with scRNA-seq and multi-omics approaches

Muyoung Lee, Qingqing Guo, Mijeong Kim, et al.

| | |
|---|---|
| **Supplemental Material** | http://genome.cshlp.org/content/suppl/2024/04/05/gr.277926.123.DC1 |
| **P<P** | Published online April 5, 2024 in advance of the print journal. |
| **Open Access** | Freely available online through the *Genome Research* Open Access option. |
| **Creative Commons License** | This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at http://creativecommons.org/licenses/by/4.0/. |
| **Email Alerting Service** | Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.** |

To subscribe to *Genome Research* go to:
**https://genome.cshlp.org/subscriptions**