Direct Induction of the Three Pre-implantation Blastocyst Cell Types from Fibroblasts

Authors
Hana Benchetrit, Mohammad Jaber, Valery Zayat, ..., Oren Ram, Tommy Kaplan, Yosef Buganim

Correspondence
yossibug@ekmd.huji.ac.il

In Brief
Benchetrit and colleagues identified a combination of factors that can produce the three in vitro equivalent cell types of the blastocyst, iPSCs, iTSCs, and iXENs, from fibroblasts. They found that Esrrb was most potent in inducing pluripotency by the activation of a XEN-like state and Eomes most potent in promoting trophoblast fate.

Highlights
- Gata3, Eomes, Tfp2c, Myc, and Esrrb convert fibroblasts into iPSCs, iTSCs, and iXENs
- Esrrb, but not other pluripotency genes, can shift the TSC fate into pluripotency
- Esrrb induces pluripotency by the activation of a unique XEN-like state
- The interplay between Eomes and Esrrb determines cell fate decision

Benchetrit et al., 2019, Cell Stem Cell 25, 983–994
June 6, 2019 © 2019 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.stem.2019.03.018
Direct Induction of the Three Pre-implantation Blastocyst Cell Types from Fibroblasts

Hana Benchetrit,1,4 Mohammad Jaber,1,4 Valery Zayat,1,4 Shulamit Sebban,1 Avital Pushett,1 Kirill Makedonski,1 Zvi Zakheim,1 Ahmed Radwan,1 Noam Maoz,1 Rachel Lasry,1 Noa Renous,1 Michal Inbar,1 Oren Ram,2 Tommy Kaplan,3 and Yosef Buganim1,5,*
1Department of Developmental Biology and Cancer Research, Institute for Medical Research Israel-Canada, The Hebrew University-Hadassah Medical School, Jerusalem 91120, Israel
2The Silberman Institute of Life Sciences and the Edmond and Lily Safra Center for Brain Science, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Jerusalem 9190401, Israel
3School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel
4These authors contributed equally
5Lead Contact
*Correspondence: yossibug@ekmd.huji.ac.il
https://doi.org/10.1016/j.stem.2019.03.018

SUMMARY

Following fertilization, totipotent cells undergo asymmetric cell divisions, resulting in three distinct cell types in the late pre-implantation blastocyst: epiblast (Epi), primitive endoderm (PrE), and trophectoderm (TE). Here, we aim to understand whether these three cell types can be induced from fibroblasts by one combination of transcription factors. By utilizing a sophisticated fluorescent knockin reporter system, we identified a combination of five transcription factors, Gata3, Eomes, Tfap2c, Myc, and Esrrb, that can reprogram fibroblasts into induced pluripotent stem cells (iPSCs), induced trophoblast stem cells (iTSCs), and induced extraembryonic endoderm stem cells (iXENs), concomitantly. In-depth transcriptomic, chromatin, and epigenetic analyses provide insights into the molecular mechanisms that underlie the reprogramming process toward the three cell types. Mechanistically, we show that the interplay between Esrrb and Eomes during the reprogramming process determines cell fate, where high levels of Esrrb induce a XEN-like state that drives pluripotency and high levels of Eomes drive trophectodermal fate.

INTRODUCTION

After fertilization, a rapid reprogramming of the chromosomal content of the newly formed cell occurs, resulting in a totipotent zygote. Upon division, totipotency is gradually lost and an early blastocyst is formed, containing two compartments that are more committed: the inner cell mass (ICM), which will generate the embryo proper (epiblast-Epi) and primitive endoderm (PrE), and an outer layer of trophectoderm (TE), which will give rise to extraembryonic tissues, such as the placenta (Chen et al., 2010). Several models explain the first cell fate decision in the embryo (Jaber et al., 2017; Wu and Scho¨ ler, 2016), but it remains to be elucidated which are the key factors that derive each of the two lineages (i.e., ICM and TE). In embryonic stem cells (ESCs), reciprocal repression between the pluripotency master regulator, Oct4, and the TE key gene, Cdx2, has been suggested to determine cell lineage specification (Niwa et al., 2005). Accordingly, knockout of Oct4 or overexpression of Cdx2 leads to transdifferentiation of ESCs into trophoblast stem-like cells (Niwa et al., 2000, 2005). Single-cell RNA sequencing throughout mouse pre-implantation development identified targets of the master pluripotency regulators Oct4 and Sox2 as being highly heterogeneously expressed within 4-cell stage embryos, with Sox21 showing one of the most heterogeneous expression profiles that drives cell fate commitment (Goolam et al., 2016).

The induction of pluripotency from somatic cells by a small number of defined factors (Takahashi and Yamanaka, 2006) opened a new avenue in basic research (Buganim and Jaenisch, 2012), in which cell-type-specific combinations of key master regulators are identified by demonstrating their capability to impose a stable alternative cell fate (Xu et al., 2015). Recently, we and others have shown that the introduction of Gata3, Eomes, Tfap2c, and Myc (GETM) (Benchetrit et al., 2015) or Ets2 (Kubaczka et al., 2015) in fibroblasts can initiate a reprogramming process that leads to the formation of stable and fully functional induced trophoblast stem cells (iTSCs). The success in inducing pluripotent stem cell (PSC) and TSC states by ectopic expression of transcription factors led us to search for a combination of factors that would hold the capacity to convert fibroblasts into both iPSCs and iTSCs. We hypothesized that identifying such a combination would help to elucidate the countervailing forces that drive each lineage.

RESULTS

Ectopic Expression of Esrrb Drives the TSC Reprogramming Combination toward Pluripotency

To distinguish between PSC and TSC fates, we established a fluorescent knockin reporter system harboring 4 unique reporters: (1) Nanog-2A-EGFP, a cytoplasmic reporter that specifically marks PSCs; (2) Elf5-2A-EYFP-NLS, a nuclear reporter that is specific for TSCs; and (3) Utf1-2A-tdTomato and (4) Esrrb-2A-TagBFP,
cytoplasmic reporters that mark both cell types. Although Utf1 levels were reduced and a small fraction of the Nanog-2A-EGFP protein was not cleaved (Figure S1A), the targeted engineered KH2-ESC line (termed BYKE hereinafter) gave rise to chimeras with germline transmission following blastocyst injection (Figure S1B), indicating full developmental potential.

To test whether the BYKE system is suitable for reprogramming studies, we injected BYKE ESCs into host blastocysts and selected for BYKE mouse embryonic fibroblasts (MEFs) at 13.5 days post-coitum (DPC).

BYKE-MEFs were reprogrammed either into iPSCs by Oct4, Sox2, Klf4, and Myc (OSKM) or into iTSCs by GETM (Figure 1A). Reprogramming efficiency ranged between 2% and 20% in OSKM-iPSC reprogramming, and between 0.3% and 3% in GETM-iTSC reprogramming, depending on infection efficiency and duration of reprogramming. BYKE-iPSC clones were triple positive for the 3 PSC reporters, and BYKE-iTSC clones were triple positive for the 3 TSC reporters (Figure S1C). Taken together, these data demonstrate that the BYKE system is adequate for reprogramming and all 4 reporters are active.

We assumed that GETM would be a good initial core reprogramming combination to produce both iPSCs and iTSCs, as these factors are expressed at early embryonic developmental stages (Auman et al., 2002; Home et al., 2009; McConnell et al., 2005).

We screened for factors that would shift the TSC fate, dictated by GETM, to pluripotency and generate iPSCs. We infected BYKE-MEFs with doxycycline (dox)-inducible lentiviruses encoding for the GETM proteins together with individual factors that are expressed at early developmental stages, such as Klf5, Tead4, and Sox15; factors that drive pluripotency, like Klf4, Oct4, and Nanog; or factors that are expressed in both compartments or shared between iPSCs and iTSCs, such as Sall4, Lin28, Utf1, Sox2, and Esrb (Figure 1B).

ePSC formation was scored by fluorescence-activated cell sorting (FACS) after 20 days of dox exposure followed by 10 days of dox removal. Intriguingly, the only factor that could shift the GETM fate and form iPSCs was Esrb (S). The resulting iPSCs were positive for the 3 PSC reporters (Figure 1B) and showed tight dome morphology similar to ESCs. Interestingly, all factor combinations resulted in a small fraction of Utf1-2A-tdTomato-only-positive cells (Figure 1B). qPCR analysis on sorted cells revealed that these cells are trophoblast differentiated cells (Figure 1C). These results suggest that Esrb harbors a unique property that allows it to shift TSC fate, driven by GETM, into PSC fate.

Ectopic Expression of GETMs Produces Both iPSCs and iTSCs

We examined whether the 5 factor combination (5F or GETMS) is capable of also producing iTSCs or whether the addition of Esrb to GETM abrogates this capability. MEFs were infected with GETMS (Figure S1D) and then reprogrammed in ESC or in TSC medium. Strikingly, infected BYKE-MEFs that were cultured in ESC medium produced iPSCs and trophoblast differentiated cells, and infected BYKE-MEFs grown in TSC medium formed mostly iTSCs, with few iPSC colonies (Figures 1D and 1E). Cultivating the infected cells for 2 weeks in 10% fetal bovine serum (FBS) in DMEM with dox, followed by 1 week in TSC reprogramming medium, and subsequently removing dox for 14 days in TSC culture medium produced both stable iPSC and iTSC colonies in the same dish (Figures 1F and S1E).

6/7 examined iPSC clones and 4/7 iTSC clones were positive for all 5 transgenes in their genome (Figure S1F), indicating that the GETMS combination holds the potential to produce both iPSCs and iTSCs.

The Interplay between Esrb and Eomes Determines Cell Identity during Reprogramming with GETMS

We asked whether a specific stoichiometry of reprogramming factors within the GETMS combination is favorable for the production of iPSCs versus iTSCs or whether GETMS produce transient bi-potential cells. Thus, we examined the expression levels of GETMS in various secondary (2') MEF systems. 2' MEF systems were obtained by infecting MEFs with dox-inducible lentiviruses encoding for GETMS, and then “primary” iPSC colonies were isolated and injected into host blastocysts to derive “secondary” MEFs. This allowed us to generate multiple 2' MEF systems in which each system captures only one stoichiometry of factors and can be immediately activated by the

Figure 1. Esrb Drives TSC Reprogramming with GETM to Pluripotency

(A) Schematic representation, bright field and fluorescence images of the 3 PSC reporters (left), 3 TSC reporters (right), and FACS analysis of the 4 reporters in BYKE-MEFs undergoing reprogramming for 20 days with OSKM (left) or GETM (right) followed by 10 days of dox removal. A typical experiment out of 3 independent experiments is shown.

(B) BYKE-MEFs were infected with GETM plus individual factors as indicated and tested for iPSC formation. The graph summarizes FACS analysis of the 3 PSC reporters in BYKE-MEFs undergoing reprogramming for 20 days with GETM plus the indicated single factor, followed by 10 days of dox removal. A typical experiment out of 3 independent experiments is shown.

(C) BYKE-MEFs were reprogrammed by GETM + Esrb (GETMS) for 20 days with dox followed by 2 days of dox removal. Subsequently, different populations of cells (Utf1-2A-tdTomato+/Esrb-2A-TagBFP+/Nanog-2A-EGFP- and Utf1-2A-tdTomato+/Esrb-2A-TagBFP+/Nanog-2A-EGFP+) were sorted and tested for expression of ESC-specific (Oct4 and Nanog), TSC-specific (Elf5 and Cdx2), and trophoblast differentiation-specific (Gcm1, Ctsq, and Tpbpa) genes. mRNA levels were normalized to Gapdh. Error bars presented as a mean ± SD of 2 duplicate runs from a typical experiment out of 3 independent experiments.

(D and E) FACS analyses for the 4 reporters on BYKE-MEFs undergoing reprogramming with GETM for 20 days followed by 10 days of dox removal under (D) ESC (GETMS-ESM) or (E) TSC (GETMS-TSM) culture conditions.

(F) Left) Reprogramming strategy for producing both iPSCs and iTSCs with GETMS factors. (Right) FACS analysis of the 4 reporters in BYKE-MEFs undergoing reprogramming with GETMS following the aforementioned reprogramming strategy.

(G) qPCR of the GETMS transgenes normalized to Gapdh in BYKE-GETMS#10 2' MEFS after 72 h of dox induction. Error bars presented as a mean ± SD of 2 duplicate runs from a typical experiment out of 3 independent experiments.

(H and I) Graphs summarizing FACS analyses for all 4 reporters (H) and representative bright field and fluorescence images (I) after 20 days of dox followed by 10 days of dox removal of BYKE-GETMS#10 2' MEFS infected with additional single factor as depicted in ESC or TSC medium. Error bars presented as a mean ± SD of 2 duplicate runs from a typical experiment out of 3 independent experiments.

Related to Figure S1.
Figure 2. GETMS-iPSCs and GETMS-iTSCs Are Fully Functional
(A) (Left) Oct4-GFP MEFs were infected with GETMS factors, with Esrrb marked by EGFP and Eomes marked by tdTomato. 6 days post-dox induction, 16,000 of Esrrb-EGFP-positive and Eomes-tdTomato-high cells or 16,000 of Esrrb-EGFP-positive and Eomes-tdTomato-low cells were sorted and plated on feeder cells. Reprogramming continued for 20 more days in ESC medium followed by 10 days of dox removal. Left: number of Oct4-GFP-positive iPSC colonies is shown. Right: FACS analysis of Oct4-GFP-positive cells is shown. (Right) Oct4-GFP MEFs were infected with GETMS factors, with Eomes marked by EGFP and Esrrb marked by tdTomato. 6 days post-dox induction, 80,000 of Eomes-EGFP-positive and Esrrb-tdTomato-high cells or 80,000 of Eomes-EGFP-positive and Esrrb-tdTomato-low cells were sorted and plated on feeder cells.
addition of dox. We then examined the correlation between stoichiometry and reprogramming outcome.

In a 2× MEF system (2× BYKE-GETMS#10), which expressed low levels of all 5 factors (Figure 1G) and which following dox addition resulted in proliferation and differentiated trophoblasts, additional injection of Esrrb, but not any of the other 4 factors, produced iPSCs under ESC culture conditions (Figures 1H, upper, and 1I, left). In contrast, under TSC culture conditions, the addition of Eomes, but not any of the other factors, produced iTSCs (Figures 1H, lower, and 1I, right).

When a 2× MEF system (2× KH2-NgUrEb-GETMS#1) with very high levels of Eomes was examined (Figure S1G), no iPSC formation was noted, even when ESC conditions were used and Esrrb was overexpressed in the cells (Figure S1H). Overexpression of Eomes in this 2× system under TSC culture conditions facilitated further formation of iTSCs, but excessive levels of Esrrb or Gata3 hindered iTSC production (Figure S1I). According to its main role in inducing a TSC fate, removing Eomes from the 5F combination (Gata3, Tfp2c, Myc, and Esrrb [GTMS]) still allowed the production of iPSCs under ESC culture conditions (Figure S1J).

Lastly, to determine whether the ratio between Esrrb and Eomes can predict the outcome of resulting colonies early in reprogramming, we infected Oct4-GFP MEFs with GTM together with Eomes-2A-tdTomato and Esrrb-2A-EGFP. At day 6 of reprogramming, an early time point before Oct4-GFP signal can be detected, two cell populations were sorted, Esrrb-2A-EGFP positive with high Eomes-2A-tdTomato expression (Esrrb+/EomesHigh) and Esrrb-2A-EGFP positive with low Eomes-2A-tdTomato expression (Esrrb+/EomesLow). Reprogramming toward iPSCs was continued for each population separately. Notably, Esrrb+/EomesLow cells gave rise to 4-fold more Oct4-GFP iPSC colonies compared to Esrrb+/EomesHigh cells (Figure 2A).

A reciprocal experiment resulted in enrichment for iTSC colonies in cells with Eomes-2A-EGFP-positive and low Esrrb-2A-tdTomato expression, as assessed by the TSC-specific cell surface marker Cd40 (Figure 2A). Taken together, these results indicate that a bi-potential state is not achieved during reprogramming with GETMS but rather that the interplay between Eomes and Esrrb, along with culture conditions, are the factors that determine cellular identity.

**GETMS-iPSCs and GETMS-iTSCs Are Fully Functional**

To examine whether GETMS-iPSCs and GETMS-iTSCs are fully functional and that their developmental potential resembles ESCs and blastocyst-derived TSCs (bdTSCs), respectively, we injected GETMS-tdTomaTo-marked iPSCs and GETMS-EGFP-marked iTSCs into blastocysts and monitored their contribution to 13.5-DPC embryos and placentas. All injected GETMS-iPSC clones produced high-grade chimeras with a contribution that was comparable to their ESC parental clone (Figures 2B and S2A). Of note, contribution to the extraembryonic mesoderm or endoderm, which originate from the ICM, was noted in the placenta as well (shown as tdTomato signal in the blood vessels of the placenta; Figure S2B).

In contrast, GETMS-iTSC clones contributed solely to the trophoblast compartment of the placenta, with no signal in the embryo (Figures 2B, 2C, and S2A). Moreover, GETMS-iTSCs formed transient hemorrhagic lesions with large lacunae when injected subcutaneously into nude mice and produced all trophoblastic cell types (i.e., giant multinucleated trophoblast cells, spangiotrophoblast cells, and syncytiotrophoblast cells) within the lesions (Figures 2D, S2C, and S2D). GETMS-iTSCs showed the same developmental potential in vitro, following removal of Fgf4 and heparin from culture medium, as depicted by qPCR for the relevant markers and cell cycle FACS analysis (Figures S2E and S2F).

Lastly, whole-transcriptome analysis of 2 clones from each GETMS-iPSC, OSKM-iPSC, ESC, GETMS-iTSC, GETM-iTSC, bdTSC, and MEF clustered all PSCs together and all TSCs together and far from MEFs (Figure 2E). These data indicate that GETMS can produce high-quality iPSCs and iTSCs, which resemble ESCs and bdTSCs, respectively, in their function and transcription.

**In-Depth Transcriptomic, Chromatin, and Epigenetic Analyses of GETMS Reprogramming**

Our next goal was uncovering the molecular mechanisms underlying GETMS reprogramming. We performed RNA sequencing (RNA-seq) and the assay for transposase-accessible chromatin (ATAC-seq) on cells in reprogramming with various combinations of factors: OSKM and GTMS for iPSC-only reprogramming, GETM for iTSC-only reprogramming, and GETMS in ESC medium (GETMS-ESM) or in TSC medium (GETMS-TSM). We compared the reprogramming cells to MEF, ESC, iPSC, and bdTSCs controls following 3 days of transgene expression, an early time point where changes are relatively homogeneous among all reprogramming cells (Buganim et al., 2013). Principal-component analysis (PCA) of the transcriptional profiles of the various conditions clustered ESCs and bdTSCs far from all reprogramming combinations and MEF controls, as expected (Figure 3A). Plotting only MEFs and MEFs undergoing reprogramming showed a clear difference between the transcriptome of OSKM-expressing MEFs, control MEFs, and all GTM-containing combinations (Figure 3B).
Figure 3. Unbiased Comparative Transcriptome and Chromatin Analysis of the Various Reprogramming Combinations

(A) PCA of global gene expression profiles from RNA-seq data of two biological replicates of MEFs, bdTSCs, ESCs, and the indicated reprogramming combinations after 72 h of dox induction.

(legend continued on next page)
Clustering of 13,309 expressed genes revealed 19 clusters with unique signatures (Figure 3C): MEF-specific genes (clusters nos. 1, 2, and 6; e.g., Thy1 and Col5a1), OSKM-specific genes (cluster no. 3; e.g., Cd34 and Aldh3a1), GETM, GETMS, and GTMS-specific genes (cluster no. 9; e.g., Pcsa and Pri32), GTMS, GETM, GETMS, and bdTSC-specific genes (cluster no. 11; e.g., Gatsl13 and Mga4t1b), and ESC and bdTSC or ESC- or TSC-specific genes (clusters nos. 10, 12, 15, 16, 17, 18, and 19; e.g., Gdf3 and Elf5).

Similar cluster analysis was performed for 188,432 intrinsic and distal ATAC-seq peaks in all 8 conditions, yielding 18 distinct clusters with a chromatin-reshaping pattern that closely resembles that of the transcriptome (Figure 3D).

Using GREAT, we allocated distal and intrinsic peaks to their adjacent genes and ran gene ontology (GO) annotation for the ATAC-seq and RNA-seq clusters (Tables S1 and S2). Although TSC- and iPSC-specific clusters were enriched for blastocyst formation (false discovery rate [FDR] = 2.7E–16), TSC-specific clusters for placenta development (FDR = 5.9E–23), GETM-, GETMS-, and bdTSC-specific cluster for focal cell adhesion (FDR = 7.8E–14), and labyrinthine layer development (FDR = 9.7E–14) as expected, OSKM-specific cluster was enriched also for genes that cause placental defects (Table S1; Figure S3A), suggesting a role of OSKM in repressing placental genes early on. The other reprogramming clusters were enriched for processes involved in MET, proliferation and metabolic shift, all known to take place at early time points of reprogramming with OSKM (Tables S1 and S2).

The strongest 1,000 ATAC-seq peaks were located at promoters (40%–50%); however, although ATAC-seq peaks at the TSS region (±2.5 Kb) of the 19 RNA clusters could not identify unique signatures (Figure S3B), distal and intrinsic peaks showed a clear overlap with the transcriptome (Figure 3D). These data suggest that intronic and distal regions are more specifically arranged in each cell type and are reshaped early in reprogramming. Indeed, unique intronic and distal peaks were observed in OSKM-specific markers, such as Gdf3 (Buginan et al., 2012), and for iTSC-specific markers, such as Elf5 (Benchetrit et al., 2015; Kubaczka et al., 2015), in the various combination of factors (Figures 3E and 3F).

We compared global ATAC-seq peaks located near active genes (FPKM > 1) to those located near inactive genes (FPKM < 1) in OSKM and GETM reprogramming. Interestingly, although peaks near active genes showed a significant enrichment for Theliger stage 1, 2, and 4 (FDR = 6.8E–10) for OSKM and TE (FDR = 2.3E–22) for GETM, no specific enrichment was found for ATAC-seq peaks located near inactive genes (Table S3).

In contrast, GETMS–specific peaks near active genes were enriched for genes regulating PrE (p ≤ 0.00003) and TE (p ≤ 0.0001), and peaks located near inactive genes were enriched for genes regulating gonad primordium (p ≤ 0.00003; Table S3). These results suggest that OSKM, GETM, and GETMS bind and activate genomic loci involved in early embryogenesis already at day 3 of reprogramming and imply that OSKM and GETMS combinations produce iPSCs by reshaping distinct genomic loci.

Esrrb Drives GETM Reprogramming toward Pluripotency by Inducing a XEN-like State

To understand how Esrrb in conjunction with GETM drives iPSC formation, we initially searched for RNA binding motifs that are enriched within ESrrb-specific ATAC-seq peaks (GETMS and GTMS) compared to controls (GETM, MEFs, bdTSCs). The most significant binding motifs within Esrrb-specific peaks were shown to be enriched in regulatory elements of germ cells (Eik1, Eik4, Nrfl, YY1, Sp1, and Zfh; Hammoud et al., 2014; Luoh et al., 1997; Thomas et al., 2007; Wu et al., 2009), XENs or PrE (Elf1, Nr5a2, and Klf5; Cho et al., 2012; Lin et al., 2010; McDonald et al., 2014), and PSCs (Nr5a2, Klf4, and Zfh; Table S4; Figures 3G and S3C).

Given that a XEN-like state is achieved during chemical reprogramming just before pluripotency (Li et al., 2017; Zhao et al., 2015) and the above binding motifs analysis, we sought to determine whether such a state is induced in reprogramming with GETMS. We performed independent per-study differential gene expression analyses of RNA-seq data from the present study and from 3 different studies describing a XEN-like signature (Li et al., 2017; Parenti et al., 2016; Zhao et al., 2015). We revealed a statistically significant XEN-like signature at day 3 of factor induction in combinations that include Esrrb with an over-all 225 genes that were upregulated by GETMS and GTMS combinations, but not in controls. 96 of them were shared with one or more of the studies describing a XEN-like signature (Figures 3H and 3I).

Chromatin immunoprecipitation (ChIP)-seq for H3K4me2 and H3K27ac was conducted on MEFs overexpressing GETMS, GTMS, or Esrrb alone (S) as control, following 3 days of dox. Differential binding analysis of H3K4me2 ChIP-seq data, contrasting GETMS and GTMS against the control samples, MEFs
and MEFs +S, found 116 genes that were intersected with the 225 XEN-signature genes (p ≤ 2.7E-08; Figure S3D). Indeed, activation signals were accumulated in XEN-specific loci, such as Myh6, as well as in Oct4 in GETMS and GTMS compared to MEFs + S control (Figure S3E).

Accordingly, we were able to isolate stable XEN-like stem cells (GETMS-iXENs) after 20 days of GETMS expression by applying XEN culture conditions. GETMS-iXENs exhibited a similar morphology to that of blastocyst-derived XENs (bdXENs) and expressed XEN-specific markers (Figures 4A and S4A). Importantly, higher levels of Oct4 and SalI4 were observed in stable GETMS-iXENs compared to bdXENs, explaining how a XEN-like state might lead to pluripotency (Figure S4A).

To examine whether XEN-specific genes are enriched in GETMS-reprogrammed cells just before the acquisition of pluripotency, we reprogrammed Oct4-GFP cells with either GETMS or with OSKM, and sorted Oct4-GFP-positive cells at day 12 of dox induction from both groups. qPCR analysis showed a clear induction of the XEN-specific gene, Myh6, in GETMS reprogrammable Oct4-GFP-positive cells compared to Oct4-GFP-negative or OSKM reprogrammable cells (Figure S4B). Myh6 was also higher in cells of GETMS reprogramming in a combination with high Esrrb expression sorted at day 6 (Figure S4C).

To understand whether a XEN-like state with high Oct4 expression can be boosted at the expense of the PSC state during GETMS reprogramming, we infected Oct4-GFP MEFs with GETMS together with either an empty vector or with key XEN genes, Sox17 and Gata6. A dramatic increase in Oct4-GFP-positive cells was noted in GETMS combinations harboring XEN-specific transgenes (Figure 4B). qPCR analysis for PSC and XEN-specific markers on sorted Oct4-GFP cells following 20 days of dox induction and 10 days of dox removal revealed increased percentage of Oct4-GFP cells in GETMS plus Sox17 and Gata6 due to increased number of iXENs at the expense of iPSCs (Figures 4C and 4D). The elevation in Oct4-GFP cells was mostly attributed to Esrrb, as GETMS combination together with Gata6 and Sox17 gave rise to very few Oct4-GFP cells (Figure S4D). These results suggest that Esrrb, in conjunction with GETM, is a potent inducer of a unique XEN-like state that leads to pluripotency.

**Esrrb Regulates XEN Genes in Other Systems**

To examine whether the activation of XEN-specific genes by Esrrb is a general phenomenon, we utilized ZHBTC4 ESCs, which are capable of transdifferentiating into TS-like cells by elimination of Oct4 expression using a Tet-OFF system (Niwa et al., 2000). We infected ZHBTC4 ESCs with Esrrb-2A-EGFP or Eomes-2A-EGFP and isolated EGFP-positive ESC clones that overexpressed either Esrrb or Eomes. As a control, we used the parental ZHBTC4 ESC line. All clones were grown under ESC culture conditions, and dox was added for 4 days to allow transdifferentiation. Each day, cells were collected and tested for expression of TSC-specific genes (Cdx2, Gata3, Tlap2c, Elf5, and Tead4), XEN-specific genes (Sox17, Gata6, Gata4 Sox7, and Stral), and PSC-specific genes (Zfp42 and Nanog).

As expected, Eomes-expressing cells showed a stronger induction of Elf5 and Tead4 compared to control and Esrrb-expressing cells (Figure S4E). On the other hand, a delay in the activation of TSC-specific genes was noted in Esrrb-expressing cells, which was accompanied by a strong upregulation of XEN-specific genes (Figures 4E and S4E). Accordingly, stable XEN-like colonies and a smaller number of stable TS-like colonies were observed in Esrrb-expressing cells compared to control cells that exhibited only stable TS-like colonies (Figures S4F–S4I).

In agreement with that, ChIP-qPCR for Esrrb in ZHBTC4 cells overexpressing Esrrb revealed enhanced binding of Esrrb to XEN-specific and TSC-specific genes compared to wild-type (WT) and CRISPR-Cas9-mediated knockdown (KD) control cells (Figures 4F and 4G). Taken together, these data imply that Esrrb drives pluripotency in conjunction with GETM via activation of a XEN-like state and by controlled repression of TSC-specific genes.

**DISCUSSION**

In this study, we asked whether one combination of transcription factors is capable of producing cells that are the in vitro equivalent of the ICM-Epi (i.e., iPSCs) and TE (i.e., iTSCs).

We identified a combination of 5 factors, GETMS, that can reprogram MEFs into both iPSCs and iTSCs. Although 3 or 4 factors out of these 5F combinations were sufficient to produce
either iPSCs (GTMS or ETMS) or iTSCs (GET or GETM), the only combination that could produce both cell types was GETMS. We show that, out of the 3 TSC key master regulators, Gata3, Tfap2c, and Eomes, the expression of Eomes is the most influential on the outcome of the induced cells, where high levels of Eomes drive cells toward the TSC state. In contrast, Esrrb was the only factor that could shift the TSC reprogramming process exerted by GETM toward pluripotency. This is in accordance with the observation that Esrrb was shown to act downstream to Sox21, a protein that was shown to initiate the pluripotency program at the 4-cell stage (Goolam et al., 2016). These results are extremely surprising, given that Esrrb is highly expressed in the ICM and TE compartments of the embryo (Deng et al., 2014) and is essential for TSC self-renewal (Latos et al., 2015).

Our molecular analyses show that GETMS triggered pluripotency via a distinct pathway and chromatin dynamics different than OSKM. Pluripotency markers and genomic loci, such as Gdf3, that were activated by OSKM in the reprogramming process were not expressed or marked by GETMS reprogramming and vice versa.

In contrast, Esrrb in conjunction with GETM, or when overexpressed in ESC that lost Oct4 expression, activated XEN cell-specific genes that could produce stable iXENs. However, in contrast to bdXENs, GETMS-iXENs expressed very high levels of Oct4. Although the activation of a XEN-like state with high levels of Oct4 by Esrrb can explain how pluripotency is initiated, we believe that the role of Esrrb in repressing the TSC state when cells start to acquire pluripotency is at least equally important for the stabilization process. These two parallel functions of Esrrb distinguish it from other pluripotency genes and might explain why Esrrb was the only factor that could shift the TSC fate into PSC fate.

Because all five factors have been shown to be expressed to some degree in the embryo before specification (Auman et al., 2002; Home et al., 2009; McConnell et al., 2005), it is tempting to speculate that the interplay between Esrrb and Eomes is also relevant for the first cell fate decision in the embryo.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

| KEY RESOURCES TABLE |
| CONTACT FOR REAGENT AND RESOURCE SHARING |
| EXPERIMENTAL MODEL AND SUBJECT DETAILS |
| METHOD DETAILS |
| BYKE Quadruple Reporter System |
| Molecular Cloning, Lentiviral Infection, and iTSC and iPSC reprogramming |
| FACS analysis |
| Chimeric embryo and placenta contribution |
| ZHBTc4 ESCs transdifferentiation into TS-like cells |
| Hemorrhagic lesion formation |
| Generation of ZHBTc4 ESCs Esrrb hypomorph using CRISPR/cas9 |
| Quantitative PCR |
| Immunostaining and western blot |
| Chromatin immunoprecipitation (ChIP) |
| RNA libraries and Sequencing |
| ATAC libraries and Sequencing |
| Mapping and analysis of RNA-Seq data |
| Principal component analysis |
| Cluster analysis |
| Mapping and analysis of ATAC-Seq data |
| Cluster analysis of ATAC-Seq peaks |
| Motif analyses and functional annotations |
| Mapping and analysis of ChiP-Seq |
| Venn diagram and heatmap of XEN signature |
| QUANTIFICATION AND STATISTICAL ANALYSIS |
| Identification of a XEN-like signature |
| DATA AND SOFTWARE AVAILABILITY |

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.stem.2019.03.018.

ACKNOWLEDGMENTS

Y.B. is supported by the European Research Council (ERC) (no. 676843), the Israeli Center of Research Excellence (i-CORE) program (center no. 41/11), the Israel Science Foundation (ISF) (no. 823/14), EMBO Young Investigator Programme (YIP), DKFZ-MOST (CA 177), and Howard Hughes Medical Institute (HHMI) (no. 55008727). A.R. is supported by BIRAX (030-5187). T.K. is supported by i-CORE program (centers no. 41/11 and no. 1796/12) and ISF, no. 913/15.

AUTHOR CONTRIBUTIONS

Y.B. conceived the study, wrote the manuscript, and prepared the figures. H.B., M.J., and Y.B. designed the experiments. H.B. and M.J. performed cloning, reprogramming, infection, iTSC and iPSC isolation, FACS, ChiP-qPCR, RNA preparation, and immunostaining. S.S. generated RNA and ATAC libraries. T.K. analyzed all RNA-seq, ATAC-seq, and ChiP-seq data. M.J. and O.R. performed ChiP-seq. Y.B. and V.Z. built the BYKE system. K.M. performed iTSC and iPSC injections. N.M. validated Esrrb KD. N.R. cloned factors. Z.Z. characterized GETMS-iXENs. A.R. analyzed the XEN-like signature. A.P. and H.B. performed the experiments with ZMBtC4. M.I. analyzed some of the ChiP-seq and RNA-seq data.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 25, 2018
Revised: January 3, 2019
Accepted: March 22, 2019
Published: April 25, 2019

SUPPORTING CITATIONS

The following reference appears in the Supplemental Information: Faddah et al. (2013).

REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169.
Auman, H.J., Nottoli, T., Lakiza, O., Winger, Q., Donaldson, S., and Williams, T. (2002). Transcription factor AP-2gamma is essential in the extra-embryonic lineages for early postimplantation development. Development 129, 2733–2747.
GATA3 is selectively expressed in the trophectoderm of peri-implantation development. Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-identity genes specify cell fates during direct reprogramming of mouse somatic cells. Cell Stem Cell 17, 543–556.

Brambrink, T., Foreman, R., Welstead, G.G., Lengner, C.J., Wernig, M., Suh, H., and Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell 2, 151–159.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218.

Buganim, Y., and Jaenisch, R. (2012). Transdifferentiation by defined factors as a powerful research tool to address basic biological questions. Cell Cycle 11, 4485–4486.

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

Buganim, Y., Faddah, D.A., and Jaenisch, R. (2013). Mechanisms and models of somatic cell reprogramming. Nat. Rev. Genet. 14, 427–439.

Chang, G., Gao, S., Hou, X., Xu, Z., Liu, Y., Kang, L., Tao, Y., Liu, W., Huang, B., Kou, X., et al. (2014). High-throughput sequencing reveals the disruption of methylation of imprinted genes in induced pluripotent stem cells. Cell Res. 24, 293–306.

Chen, L., Wang, D., Wu, Z., Ma, L., and Daley, G.Q. (2010). Molecular basis of the first cell fate determination in mouse embryogenesis. Cell Res. 20, 982–993.

Cho, L.T., Warnaitha, S.E., Tsai, I.J., Artus, J., Sherwood, R.I., Pedersen, R.A., Hadjantonakis, A.K., and Niakan, K.K. (2012). Conversion from mouse embryonic tropho-ecto-meric to extra-embryonic endoderm stem cells reveals distinct differentiation capacities of pluripotent stem cell states. Development 139, 2866–2877.

Cohen, M.A., Markoulaki, S., and Jaenisch, R. (2018). Matched developmental timing of donor cells with the host is crucial for chimera formation. Stem Cell Reports 10, 1445–1452.

Dawlaty, M.M., and van Deursen, J.M. (2006). Gene targeting methods for studying nuclear transport factors in mice. Methods 39, 370–378.

Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science 343, 193–196.

Faddah, D.A., Wang, H., Cheng, A.W., Katz, Y., Buganim, Y., and Jaenisch, R. (2013). Single-cell analysis reveals that expression of nanog is biallelic and equally variable as that of other pluripotency factors in mouse ESCs. Cell Stem Cell 13, 23–29.

Goolam, M., Scialdone, A., Graham, S.J.L., Macaulay, I.C., Jedrusiak, A., Hupalowska, A., Voet, T., Marioni, J.C., and Żernicka-Goetz, M. (2016). Heterogeneity in Oct4 and Sox2 targets biases cell fate in 4-cell mouse embryos. Cell 165, 81–74.

Hammoud, S.S., Low, D.H., Yi, C., Carrell, D.T., Guccione, E., and Cairns, B.R. (2014). Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. Cell Stem Cell 15, 239–253.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors drive cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589.

Home, P., Ray, S., Dutta, D., Bronshhteyn, I., Larson, M., and Paul, S. (2009). GATA3 is selectively expressed in the trophectoderm of peri-implantation embryo and directly regulates Cdx2 gene expression. J. Biol. Chem. 284, 28729–28737.

Jaber, M., Sebba, S., and Buganim, Y. (2017). Acquisition of the pluripotent and trophoderm states in the embryo and during somatic nuclear reprogramming. Curr. Opin. Genet. Dev. 46, 37–43.

Kubaczka, C., Senner, C., Araújo-Bravo, M.J., Sharma, N., Kuckenberg, P., Becker, A., Zimmer, A., Brustle, O., Peitz, M., Hemberger, M., and Schorle, H. (2014). Derivation and maintenance of murine trophoblast stem cells under defined conditions. Stem Cell Reports 2, 232–242.

Kubaczka, C., Senner, C.E., Cierlitz, M., Araújo-Bravo, M.J., Kuckenberg, P., Peitz, M., Hemberger, M., and Schorle, H. (2015). Direct induction of trophoblast stem cells from murine fibroblasts. Cell Stem Cell 17, 557–568.

Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Curr. Protoc. Bioinformatics Chapter 17, Unit 11.7.

Latos, P.A., Goncalves, A., Oxley, D., Mohammed, H., Turro, E., and Hemberger, M. (2015). Fgf and Esmr integrate epigenetic and transcriptional networks that regulate self-renewal of trophoblast stem cells. Nat. Commun. 6, 7776.

Lengner, C.J., Camargo, F.D., Hochdelinger, K., Welstead, G.G., Zaidi, S., Gokhale, S., Scholer, H.R., Tomlin, A., and Jaenisch, R. (2007). Oct4 expression is not required for mouse somatic stem cell self-renewal. Cell Stem Cell 1, 403–415.

Li, X., Liu, D., Ma, Y., Du, X., Jing, J., Wang, L., Xie, B., Sun, D., Sun, S., Jin, X., et al. (2017). Direct reprogramming of fibroblasts via a chemically induced XEN-like state. Cell Stem Cell 21, 264–273.e7.

Lin, S.C., Wani, M.A., Whitsett, J.A., and Wells, J.M. (2010). Klf5 regulates lineage formation in the pre-implantation mouse embryo. Development 137, 3953–3963.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Luoh, S.W., Bain, P.A., Polakiewicz, R.D., Goodheart, M.L., Gardner, H., Jaenisch, R., and Page, D.C. (1997). Zfx mutation results in small animal size and reduced germ cell number in male and female mice. Development 124, 2275–2284.

McConnell, J., Petrie, L., Stennard, F., Ryan, K., and Nichols, J. (2005). Eomesodermin is expressed in mouse oocytes and pre-implantation embryos. Mol. Reprod. Dev. 71, 399–404.

McDonald, A.C., Biechele, S., Rossant, J., and Stanford, W.L. (2014). Sox17-mediated XEN cell conversion identifies dynamic networks controlling cell-fate decisions in embryo-derived stem cells. Cell Rep. 9, 780–793.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 28, 495–501.

Mendenhall, E.M., Williamson, K.E., Reyon, D., Zou, J.Y., Ram, O., Joung, J.K., and Bernstein, B.E. (2013). Locus-specific editing of histone modifications at endogenous enhancers. Nat. Biotechnol. 31, 1133–1136.

Ng, A.Y., Jordan, M.I., and Weiss, Y. (2001). On spectral clustering: analysis and an algorithm. In Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic (MIT), pp. 849–856.

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

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophoderm differentiation. Cell 123, 917–929.

Parenti, A., Halbisen, M.A., Wang, K., Latham, K., and Raiston, A. (2016). OSKM induce extraembryonic endoderm stem cells in parallel to induced pluripotent stem cells. Stem Cell Reports 6, 447–455.

Ramirez, F., Ryan, D.P., Grünig, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44 (W1), W160–W165.

Sommer, C.A., Stadtfeld, M., Murphy, G.J., Hochdelinger, K., Kotton, D.N., and Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells 27, 543–549.
Stark, R., and Brown, G. (2011). DiffBind: differential binding analysis of ChIP-Seq peak data. http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf.

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

Thomas, K., Wu, J., Sung, D.Y., Thompson, W., Powell, M., McCarney, J., Gibbs, R., and Walker, W. (2007). SP1 transcription factors in male germ cell development and differentiation. Mol. Cell. Endocrinol. 270, 1–7.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111.

Wernig, M., Lengner, C.J., Hanna, J., Lodato, M.A., Steine, E., Foreman, R., Staerk, J., Markoulaki, S., and Jaenisch, R. (2008). A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. Nat. Biotechnol. 26, 915–924.

Yu, G., and Schöler, H.R. (2016). Lineage segregation in the totipotent embryo. Curr. Top. Dev. Biol. 117, 301–317.

Zhao, Y., Zhao, T., Guan, J., Zhang, X., Fu, Y., Ye, J., Zhu, J., Meng, G., Ge, J., Yang, S., et al. (2015). A XEN-like state bridges somatic cells to pluripotency during chemical reprogramming. Cell 163, 1678–1691.
##STAR METHODS

###KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| **Antibodies**      |        |            |
| Esrrb               | Perseus Proteomics | PP-H6705-00; RRID: AB_1964232 |
| H3K4me2             | Millipore | 07-030; RRID: AB_310342 |
| H3K27ac             | Abcam   | ab4729; RRID: AB_2118291 |
| CD-40               | Abcam   | ab22469; RRID: AB_447074 |
| Nanog               | Bethyl  | A300-379A; RRID: AB_2632098 |
| Elf5                | Santa cruz biotechnology | sc-9645; RRID: AB_640106 |
| Gata6               | Abcam   | ab22600; RRID: AB_732529 |
| Sox17               | Santa cruz biotechnology | sc-130295; RRID: AB_2286667 |
| Utfl                | Abcam   | ab24273; RRID: AB_778767 |
| Cdx2                | Biogenenex | CDX2-88; RRID: AB_2650531 |
| Eomes               | Abcam   | ab23345; RRID: AB_778267 |
| Tfp2c               | Abcam   | ab110635; RRID: AB_10858471 |
| Gata3               | Abcam   | ab106625; RRID: AB_10887935 |
| Igg                 | Santa cruz biotechnology | sc-2025; RRID: AB_737182 |
| EGFP                | Abcam   | Ab6556; RRID: AB_305564 |
| Myc                 | Abcam   | ab32072; RRID: AB_731658 |
| Actin               | Santa cruz biotechnology | sc-1616; RRID: AB_630836 |
| **Bacterial and Virus Strains** |        |            |
| **Biological Samples** | N/A   |            |
| **Chemicals, Peptides, and Recombinant Proteins** |        |            |
| mFgf4               | In-house production | N/A |
| mLIF                | In-house production | N/A |
| Heparin             | Sigma-Aldrich | H3149 |
| Recombinant human-TGFβ| PeproTech | 100-21c-10UG |
| PDG325901           | Biogems  | 3911091-25MG |
| CHIR99021           | Biogems  | 2520691-10MG |
| Protease inhibitor cocktail | Sigma-Aldrich | P8340 |
| SYBR green Fast qPCR Mix | Applied Biosystems | A25742 |
| Transfection Reagent | Mirus Bio | TransIT-LT1 |
| Propidium iodide    | Sigma-Aldrich | P4864 |
| β-mercaptoethanol   | GIBCO    | 31350-010 |
| RNaseA              | Sigma-Aldrich | R6513-10MG |
| PMS hormone         | AIP veterinary | 59450-02 |
| hCG hormone         | CalbioChem | mbs230734 1mg |
| FHM medium          | Zenith Biotech | ZEHP-050 |
| Puromycin           | TOKU-E   | P001 |
| Zeocin              | invivoGen | ant-zn-1 |
| Neomycin            | Mercury  | 345812-20ml |
| Hygromycin          | Invitrogen | 10687-010 |
| NP-40               | Mercury  | 492018 |
| **Critical Commercial Assays** |        |            |
| adaptor-loaded Nextera Tn5 | Illumina, Fc-121-1030 | 15028212 |
| SENSE mRNA-Seq library prep kit V2 | Lexogen | k00124 |

*(Continued on next page)*
### Deposited Data

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| ATAC-Seq            | This study | GEO: GSE98124 |
| RNA-Seq             | This study | GEO: GSE98124 |
| ChIP-Seq            | This study | GEO: GSE98124 |

### Experimental Models: Cell Lines

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| KH2 ESCs            | (Beard et al., 2006) | N/A |
| V6.5 ESCs           | (Beard et al., 2006) | N/A |
| R26-tdTomato ESCs   | (Cohen et al., 2018) | N/A |
| Oct4-GFP MEFs       | (Lengner et al., 2007) | N/A |
| BYKE ESCs           | This study | N/A |
| ZHBTc4 ESCs         | (Niwa et al., 2005) | N/A |

### Recombinant DNA

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| FUW-TetO-Sox15      | This study | N/A |
| FUW-TetO-Klf5       | This study | N/A |
| FUW-TetO-Sox17      | This study | N/A |
| FUW-TetO-Gata6      | This study | N/A |
| FUW-TetO-Esrrb-2A-EGFP | This study | N/A |
| FUW-TetO-Eomes-2A-EGFP | This study | N/A |
| FUW-Esrrb-2A-EGFP   | This study | N/A |
| FUW-Eomes-2A-EGFP   | This study | N/A |
| FUW-TetO-Eomes-2A-tdTomato | This study | N/A |
| FUW-TetO-Eomes-2A-tdTomato | This study | N/A |
| FUW-TetO-mOKSM STEMCCA | (Sommer et al., 2009) | N/A |
| FUW-TetO-Gata3      | (Benchetrit et al., 2015) | N/A |
| FUW-TetO-Tfap2c     | (Benchetrit et al., 2015) | N/A |
| FUW-TetO-Eomes      | (Benchetrit et al., 2015) | N/A |
| FUW-TetO-Tead4      | (Benchetrit et al., 2015) | N/A |
| FUW-TetO-Nanog      | (Buganim et al., 2012) | Addgene #40800 |
| FUW-TetO-Sall4      | (Buganim et al., 2012) | N/A |
| FUW-TetO-Uhf1       | (Buganim et al., 2012) | N/A |
| FUW-TetO-Lin28      | (Buganim et al., 2012) | Addgene #60345 |
| FUW-TetO-Esrhb      | (Buganim et al., 2012) | Addgene #40798 |
| FUW-TetO-Oct4       | (Brambrink et al., 2008) | Addgene #20323 |
| FUW-TetO-Sox2       | (Brambrink et al., 2008) | Addgene #20326 |
| FUW-TetO-Klf4       | (Brambrink et al., 2008) | Addgene #20322 |
| FUW-TetO-Myc        | (Brambrink et al., 2008) | Addgene #20324 |
| FUW-2A-EGFP         | This study | N/A |

### Software and Algorithms

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| MATLAB              | MathWorks | R2015b |
| Cufflinks (v2.0.2)  | (Trapnell et al., 2010) | https://cole-trapnell-lab.github.io/cufflinks/releases/v2.0.2/ |
| Cutadapt            | https://doi.org/10.14806/ej.17.1.200 | https://cutadapt.readthedocs.io/ |
| TopHat (v2.0.14)    | (Trapnell et al., 2009) | https://ccb.jhu.edu/software/tophat/ |
| HTSeq-count (v0.6.0) | (Anders et al., 2015) | https://htseq.readthedocs.io/en/release_0.11.1/ |
| DESeq2 (v1.16.1)    | (Love et al., 2014) | https://bioconductor.org/packages/release/bioc/html/DESeq2.html |
| Bowtie              | (Langmead, 2010) | http://bowtie-bio.sourceforge.net/index.shtml |
| HOMER’s annotatePeaks.pl program | (Heinz et al., 2010) | http://homer.ucsd.edu/homer/ngs/annotation.html |
| deepTools           | (Ramirez et al., 2016) | https://github.com/deeptools/deepTools |

(Continued on next page)
CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Yosef Buganim (yossibu@ekmd.huji.ac.il)

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice and Cell Culture

MEFs were grown in DMEM supplemented with 10%FBS, 2mM L-glutamine, and antibiotics. ESCs and iPSCs (male sex) were grown in DMEM supplemented with 10%FBS, 1%non-essential amino acids, 2mM L-glutamine, in-house mouse Leukemia inhibitory factor (mLif), 0.1mM β-mercaptoethanol (Sigma), and antibiotics with or without 2i-PD0325901 (1 μM) and CHIR99021 (3 μM) (PeproTech). BYKE ESCs or 5F-iPSCs were injected into blastocysts and at 13.5dpc embryos were isolated and MEFs were extracted and used for primary infection. All infections were performed on MEFs (passage 0 or 1) that were seeded at 60%–80% confluency two days prior to the first infection. Blastocyst-derived TSC (bdTSC) lines were isolated previously (Benchetrit et al., 2015). bdTSCs and stable iTSCs were grown in TSC culturing medium, combined of 30%TSC medium containing RPMI supplemented with 0.1mM β-mercaptoethanol (Sigma), 2mM L-glutamine, in-house mouse FGF4 (equivalent to 25ng/ml) and 1 mg/ml heparin (Sigma-Aldrich), and 70%MEF conditioned TSC media (MEF-CM) with the same supplements. For culturing bdTSCs and stable iTSCs in defined medium, cells were grown on GFR-Matrigel-coated dishes in TX medium as described previously (Kubaczka et al., 2014). Extraembryonic endoderm (XEN) cells and induced XENs were grown in RPMI supplemented with 15%FBS, 2mM L-glutamine, 1mM β-mercaptoethanol (sigma), in-house mouse FGF4 (equivalent to 25ng/ml) and 1 μg/ml heparin (Sigma-Aldrich). For differentiation experiment DMEM, without Fgf4 and heparin, supplemented with 10%FBS, 2mM-glutamine and antibiotics was used. The joint ethics committee (IACUC) of the Hebrew University and Hadassah Medical Center approved the study protocol for animal welfare. The Hebrew University is an AAALAC international accredited institute.

Primary MEFs and secondary MEF system production

Mouse embryonic fibroblasts (MEFs) were isolated as previously described (Wernig et al., 2008). Briefly, BYKE-ESCs, Oct4-GFP or OSKM/GETMS-iPSCs were injected into blastocysts and chimeric embryos were isolated at E13.5 (see Chimeric embryo and placenta contribution section for more details). For MEF production, embryos were dissected under the binocular removing internal organs and heads. The remaining body was chopped thoroughly by scalpels and exposed to 1ml Tripsin-EDTA (0.25%, GIBCO) for 30 minutes at 37°C. Following that, 10 mL of DMEM medium containing 10%FBS was added to the plate and the chopped tissue was subjected to thorough and intensive pipetting resulting in a relatively homogeneous mix of cells. Each chopped embryo was seeded in 15cm plate and cells were cultured in DMEM supplemented with 10%FBS, 2mM L-glutamine, and antibiotics until the plate was full. Puromycin (2 μg/ml) was added to each 15cm plate for positive selection for BYKE or Oct4-GFP MEFs (the M2rtTA cassette that resides inside the Rosa26 locus of the injected cells contains a resistance gene for puromycin), killing only the host cells.

METHOD DETAILS

BYKE Quadruple Reporter System

To establish dox-inducible quadruple fluorescent knock-in reporter system, the mouse ESC line KH2 containing the dox-inducible activator M2rtTA in the Rosa26 locus and a flip-in system in the collagen (Col1a1) locus (Beard et al., 2006) was Sequentially targeted with targeting constructs either containing RBGpA or Neomycin selection cassette. RBGpA was introduced into pNTKV-frt-loxP (Dawlaty and van Deursen, 2006) vector using Hpal and Hindll restriction sites using specific primers (For Sequences see Methods S1). Four different fluorescent reporters (Nanog-2A-EGFP, Utf1-2A-tdTomato, Esrrb-2A-TagBFP and Elf5-2A-EYFP-NLS) were introduced into the 3'UTR of the Nanog, Utf1, Esrrb or Elf5 genes and targeted clones were isolated either by FACS or Neomycin selection. The targeting vectors were designed by cloning the 5'arms and 3'arms using Hpal and SacII or Smal respectively. 2A-tdTomato for Ut1 and 2A-TagBFP for Esrrb were introduced into pNTKV-RBGpA using Hpal and Hindll restriction sites. gRNA for all loci was synthesized and cloned into pX330 (addgene #42230) using Bbsl sites. The targeting constructs together with corresponding gRNA were transfected into the targeting cells using TransIT-LT1 Transfection Reagent (Mirus Bio) according to manufacturer's
instructions. Correctly targeted clones for Utf1 and Esrrb were selected by FACS and verified by Southern blot and PCR, respectively. For Southern blot (Utf1 locus), DNA was digested with Nhel and probed for the 3’ region and PCR for Esrrb was done using primer pair A (Figures S1F and S1G). Lacq gene together with constitutively active Ubiquitin promoter (Ubb) were amplified by PCR (for primer Sequence see Methods S1) and cloned into pBS31-RBGpa (Beard et al., 2006) plasmid, using EcoRI site. The resulting construct was co-transfected together with flippase containing plasmid into KH2-NgUrEb ESCs and Lacq expressing clones were isolated following Hygromycin selection as previously described (Beard et al., 2006). The targeting construct for Elf5 locus was designed by cloning the 5’ arm and 3’ arm with Hpal and SacII/Smal respectively and 2A-EYFP-NLS with Hpal and BglII into pNTKV-frt-loxP (Dawlaty and van Deursen, 2006). The positive colonies for Elf5 correctly targeted locus were picked after 10-12 day of G418 selection at 500μg/ml (Millipore) and verified by primer pair B (Figures S1J–S1L). The correctly targeted clone was transfected with Cre-expressing vector (FUW-zeo-Cre) for 10-12 days of selection of Zeocin (Invivogen) at 500μg/ml and correctly targeted clones were verified by primer pair C (Figures S1J–S1L) and Sequenced.

**Molecular Cloning, Lentiviral Infection, and ITSC and iPSC reprogramming**

All dox-inducible factors were generated by cloning the open reading frame of each factor, obtained by reverse transcription with specific primers (for primer Sequences see Methods S1), into the pMINI vector (NEB) and then restricted with EcoRI or MfeI and inserted into the FUW-TetO expression vector. The mouse Eomes and Esrrb-2A-EGFP or 2A-tdtTomato constructs were generated by cloning ORF of the genes (without stop codon) into FUW-TetO-2A-EGFP or FUW-TetO-2A-tdTomato expression vector with EcoRI site (for primer Sequences see Methods S1). For infection, replication-incompetent lentiviruses containing the various reprogramming factors and rations (GETM 3:3:3:1, GTMS: 3:3:1:3, GETMS 2:2.5:2:1:2.5, OSKM 3:3:3:1 or STEMCCA cassette) were packaged with a lentiviral packaging mix (7.5 μg psPAX2 and 2.5 μg pMDG.2) in 10cm plate of 293T cells and collected 48, 60, and 72 hr after transfection. The supernatants were filtered through a 0.45μm filter, supplemented with 8 μg/ml of polybrene (Sigma), and then used to infect MEFs. Six hours following the third infection, medium was replaced with fresh DMEM containing 10%FBS. Eighteen hours later, medium was replaced to either TSC reprogramming medium (RPMI supplemented with 20%FBS, 0.1mM β-mercaptoethanol, 2mM L-glutamine, in-house mouse recombinant FGF4 (equivalent to 25ng/ml), 1 μg/ml heparin (Sigma-Aldrich), and 2 μg/ml doxycycline) or ESC reprogramming medium (DMEM supplemented with 10%FBS, 0.1mM β-mercaptoethanol, 2mM L-glutamine, 1%non-essential amino acids, in-house mouse Leukemia inhibitory factor (mLIF), and 2 μg/ml doxycycline). TSC or ESC reprogramming medium was replaced every other day for 20 days, followed by 10 days in TSC culturing medium or 2i/L culturing medium, respectively. Ten days (or as indicated in the figure) after dox removal, plates were screened for primary ITSC or iPSC colonies. For OSKM reprogramming, excluding Figure 1 that shows reprogramming to iPSCs with separate OSKM plasmids, all OSKM reprogramming experiments were performed using STEMCCA cassette. Reprogramming efficiency was measured by FACS and colony number. For iPSC or ITSC clone isolation, a single iPSC/iTSC colony was trypsinized (0.25%), and plated in a separate well in a 6-well plate on feeder cells. Wells were followed and medium was replaced every other day for five to ten passages, until stable colonies developed.

**FACS analysis**

Cells were washed twice with PBS and trypsinized (0.25%) and filtered through mesh paper. Flow cytometry analysis was performed on a Beckman Coulter and analyzed using Kaluza Software. All FACS experiments were repeated at least three times, and the bar graph results are presented as a mean ± standard deviation of two biological duplicate from a typical experiment. For cell cycle analysis, GETMS-iTSCs grown in differentiation media for the indicated time points were trypsinized and fixed with 95% ice-cold ethanol. Cells were stained with propidium iodide (PI) staining solution (50 μg/ml PI [BD]; 0.1% [v/v] Triton X-100 [Sigma-Aldrich]; 0.2mg/ml RNaseA [Sigma-Aldrich] in PBS) for 30 min at room temperature. Flow cytometry analysis was performed on a Beckman Coulter and analyzed using Kaluza Software.

**Chimeric embryo and placenta contribution**

Blastocyst injections were performed using CB6F1 host embryos. After priming with PMSG (M.I.P. Veterinary) and hCG (Merck) hormones and mating with CB6F1 males, embryos were obtained at 3.5dpc (blastocyst stage), and then injected with 10–20 ES/iPS/iTSC cells with a flat tip microinjection pipette with an internal diameter of 16 mm (Origio) in a drop of FHM medium (Zeuthen Biotech, ZEHP-050) covered by mineral oil. Shortly after injection, blastocysts were transferred to 2.5dpc pseudopregnant CD1/ICR females (10-15 blastocysts per female). Chimeric embryos and placentas were isolated at E13.5 and observed by fluorescent microscope (Nikon Eclipse T).
A total of 5x10^6 GETMS-iTSCs were resuspended in 100 μL CM containing in house mouse recombinant Fgf4 (equivalent to 25ng/ml) and 1 μg/ml heparin (Sigma-Aldrich) and injected subcutaneously into male athymic nude mice. 6-7 days later, lesions were dissected, fixed in 4% paraformaldehyde overnight, embedded in paraffin, and sectioned (4mm). Sections were stained with H&E and analyzed by a certified pathologist.

**Generation of ZHBTc4 ESCs Esrrb hypomorph using CRISPR/cas9**
ZHBTc4 ESCs were plated in ESC medium (without CHIR & PD). The next day, cells were transfected with 2.5μg PX330 vector containing Esrrb gRNA (for gRNA Sequences see Methods S1), and 0.5μg Puromycin resistance plasmid, using TransIT-LT1 transfection reagent (Mirus). 48 hours later, medium was changed to selective ESC medium with CHIR, PD and Puromycin (2μg/μl). Resistant colonies were picked, and Esrrb knockdown was validated by Sequencing, immunostaining, sm-FISH and Western Blot.

**Quantitative PCR**
Total RNA was isolated using the Macherey-Nagel kit (Ornat). 500–2000 ng of total RNA was reverse transcribed using iScript cDNA Synthesis kit (Bio-Rad). Quantitative PCR analysis was performed in duplicates using 1/100 of the reverse transcription reaction in a StepOnePlus (Applied Biosystems) with SYBR green Fast qPCR Mix (Applied Biosystems). Specific primers flanking an intron were designed for the different genes (for primer Sequences see Methods S1). All quantitative real-time PCR experiments were repeated at least three times, and the results were normalized to the expression of GAPDH and presented as a mean ± standard deviation of two duplicate runs from a typical experiment.

**Immunostaining and western blot**
Cells were fixed in 4% paraformaldehyde (in PBS) for 20 minutes. The cells were rinsed 3 times with PBS and blocked for 1hr with PBS containing 0.1% Triton X-100 and 5% FBS. The cells were incubated overnight with primary antibodies (1:200) in 4°C. The antibodies are: anti-Gata6 (Abcam, ab22660), anti-sox17 (Santa Cruz, sc-130295), anti-Esrrb (Perseus Proteomics, PP-H6705-00), anti-Eomes (Abcam, ab23345), anti-Cdx2 (Biogenex, CDX2-88), anti-Nanog (Bethyl, A300-379A) and anti-Elf5 (Santa Cruz, sc-9645) in PBS containing 0.1% Triton X-100 and 1%FBS (1:200 dilution). The next day, the cells were washed 3 times and incubated for 1hr with relevant (Alexa) secondary antibody in PBS containing 0.1% Triton X-100 and 1% FBS (1:500 dilution). DAPI (1:1000 dilution) was added 10 minutes before end of incubation. For western blot, cell pellets were lysed on ice in lysis buffer (20 mM Tris-HCl, pH 8, 1mM EDTA pH 8, 0.5% Nonidet P-40, 150mM NaCl, 10% glycerol, 1mM, protease inhibitors (Roche Diagnostics)) for 10 min, supernatant were collected and 40μg protein were suspended with sample buffer and boiled for or 5 min at 100°C, and subjected to western blot analysis. Primary antibodies: anti-Gata3 (Abcam, ab106625), anti-Tfap2c (Abcam, ab32072), anti-Eomes (Abcam, ab3345), and anti-actin (Santa cruz, Sc-1616). Blots were probed with anti-mouse, anti-goat, or anti-rabbit IgG-HRP secondary antibody (1:10,000) and visualized using ECL detection kit.

**Chromatin immunoprecipitation (ChiP)**
Chromatin immunoprecipitation (ChiP) assay was performed as previously described (Mendenhall et al., 2013). Briefly, cells were fixed for 10min at RT with a final formaldehyde concentration of 0.8%. Formaldehyde was quenched with glycine at a final concentration of 125mM. The cells were then lysed with lysis buffer (100mM Tris-HCl, 300mM NaCl, 2% Triton X-100, 0.2%v sodium deoxycholate and 10mM CaCl2) supplemented with EDTA-free protease inhibitor (Roche, 11873S0001) for 20min in Ice. The chromatin was digested by adding MNase (Thermo Scientific, 88216) for 20min at 37°C and MNase was inactivated by adding 20mM EGTA. The fragmented chromatin was added to pre-bounded Dynabeads (A and G mix, Invitrogen, 10004D/10002D) using H3K27ac antibody (Abcam, ab4729), H3K4me2 antibody (Millipore, 07-030), Esrrb antibody (Perseus Proteomics, PP-H6705-00), anti-Eomes (Abcam, ab3345), and anti-actin (Santa cruz, Sc-1616). Blots were probed with anti-mouse, anti-goat, or anti-rabbit IgG-HRP secondary antibody (1:10,000) and visualized using ECL detection kit.

**RNA libraries and Sequencing**
Total RNA was isolated using the QIAGEN RNasey kit. mRNA libraries were prepared using the SENSE mRNA-Seq library prep kit V2 (Lexogen), and pooled libraries were Sequenced on an Illumina NextSeq 500 platform to generate 75-bp single-end reads.

**ATAC libraries and Sequencing**
ATAC-Seq library preparation was performed as previously described (Buenrostro et al., 2013). Briefly, 50,000 cells per replicate (two biological replicates per line) were incubated with 0.1% NP-40 to isolate nuclei. Nuclei were then transposed for 30 min at 37°C with
adaptor-loaded Nextera Tn5 (Illumina, Fc-121-1030). Transposed fragments were directly PCR amplified and Sequenced on an Illumina NextSeq 500 platform to generate 2 × 36-bp paired-end reads.

Mapping and analysis of RNA-Seq data
Reads were mapped to the mouse genome (mm9) using TopHat (Trapnell et al., 2009) with default parameters. Expression levels were estimated in six conditions (MEFs, OSKM, GTMS, GETM, GETMS_ESM, and GETMS_TSM, in biological duplicates) using cufflinks (Trapnell et al., 2010) with default parameters. ESC and bdTSC RNA-Seq data were similarly reanalyzed from (Benchetrit et al., 2015) (GEO accession, database: GSE64684).

Principal component analysis
Gene expression levels (FPKM) for 23,309 genes were analyzed to remove genes with very low (< 10) or very high (> 1000) variance, retaining 7,446 genes. These were then standard normalized and PCA was applied using MATLAB (R2015b, “pca” function). For a closer examination of the three-day reprogramming dynamics, PCA was rerun on the same genes, using six libraries (MEFs, OSKM, GTMS, GETM, and GETMS in ESC and TSC media, in duplicates).

For transcriptional analysis of the various iPSC/ESC clones, raw reads (fastq files) were quality-trimmed and adapters removed with cutadapt (version 1.12). The processed fastq files were mapped to the mouse transcriptome and genome using TopHat (v2.0.14). The genome version was GRCm38, with annotations from Ensembl release 84. Quantification was done using htSeq-count (version 0.6.0). Genes with a sum of counts less than 10 over all samples were filtered out, retaining 21568 genes. Normalization was done with the DESeq2 package (version 1.16.1). PCA was applied and visualized in R (version 3.4.1).

Cluster analysis
Cluster analysis. Gene expression levels (FPKM) for 23,309 genes in eight RNA-Seq libraries were analyzed. Genes with FPKM lower than 2 in all eight conditions were removed, remaining with 13,309 genes. These were log2 transformed, and then multiplicatively normalized to maximal expression of 1 per gene. Genes were then clustered using spectral clustering algorithm with K = 19 (Ng et al., 2001). Briefly, this clustering algorithm computes the distance between each pair of genes, and embeds the data in a graph, whose nodes corresponds to genes and edges correspond to the similarity (or adjacency) of their expression patterns. Then, K densely connected components (= K clusters of similarly expressed genes) are identified (using K-means, with 100 random starting points). K was selected as an optimal tradeoff between generalization (fewer clusters) and specificity – as shown in Figure 4C each cluster offers a unique expression profile.

Mapping and analysis of ATAC-Seq data
Paired-end reads were mapped using Bowtie (Langmead, 2010) to the mouse genome (mm9), using max insert size of 2000. Only unique hits with up to 3 mismatches were retained. Genome-wide paired-end read coverage was then calculated and normalized to a total primary data size of 150M bases (using UCSC’s bigWigInfo program). ATAC-Seq peaks were called using the MACS2 function bdgpeakcall (Zhang et al., 2008), with min-length and max-gap parameters set to 500 bp. The top 50K peaks were then selected for each experiment, and annotated using HOMER’s annotatePeaks.pl program (Heinz et al., 2010), using default parameters. Metagene plot was created from the ChIP-Seq bigwig files using deepTools plotProfile.

Cluster analysis of ATAC-Seq peaks
For each peak, we used UCSC’s bigWigAverageOverBed program with -minMax option to compute the maximal ATAC-Seq signal for each of the eight conditions. These peaks were then united and cluster analysis was performed as described above, with K = 18. Metagene and heatmap plots were created from the ATAC-Seq bigwig files using deepTools (Ramírez et al., 2016).

Motif analyses and functional annotations
The clustered ATAC-Seq peaks (Intronic+distal peaks among top 50K ATAC-Seq peaks, Figure 4D) were systematically analyzed for enriched transcription factor binding sites using HOMER (Heinz et al., 2010). Specifically, we compared the genomics Sequences at GETMS accessible regions against accessible Sequences from the GETM ATAC-Seq peaks as background. This was done using HOMER’s findMotifsGenome.pl program. In addition, we compared the motif enrichment of each cluster versus the overall enrichment in all other classes of intronic and distal ATAC-Seq peaks. This analysis highlighted several motifs, associated with various transcription factors including Esrrb (1e-128) and others (Table S4). We then identified the ATAC-Seq peaks that contain the Esrrb motif (HOMER’s “findMotifsGenome.pl –find esrrb.motif” option), and used GREAT to analyze the specific annotations (GO and others) enriched with those genes.

Mapping and analysis of ChIP-Seq
Paired-end reads were mapped to the genome (mm9) using Bowtie (Langmead, 2010) using max insert size of 1000, mapping unique hits with up to 3 mismatches. Genome-wide paired-end read coverage was calculated and normalized to 20M reads. ChIP-Seq peaks were called using MACS (Zhang et al., 2008), using a p value threshold of 1e-3, shift size of 2Kb, short local window of
10Kb and long local window of 50Kb. Differential peaks were identified using DiffBind version 2.4.8 (Stark and Brown, 2011). Annotation was performed using ChIPseeker version 1.12.1 (Yu et al., 2015). Data for MEF control were taken from database: GSE36292 (Chang et al., 2014).

**Venn diagram and heatmap of XEN signature**

Reads from our study and three studies (Li et al., 2017; Parenti et al., 2016; Zhao et al., 2015) describing a XEN-like signature (database: GSE73631, GSE77550, and GSE97721) were aligned to the Ensembl mouse genome version mm9 (NCBI37) using STAR 2.5.2b (default parameters). Uniquely aligned reads in BAM format were annotated against the protein-coding mRNA regions using SeqMonk v 1.38.2 platform (Babraham Bioinformatics, Cambridge UK). Raw counts per mRNA, strand-specific, merging isoforms were quantified using SeqMonk. Differential expression analysis was using R (version 3.4.1) and packages DESeq2 and EdgeR.

The Venn diagram and heatmap were generated by comparing day 3 RNA-Seq data of MEFs, GETMS, GETM and GTMS induced cells and aligning them with genes related to OSKM-induced XEN cells (Parenti et al., 2016) or chemically-induced XEN cells. Each study was processed independently to obtain upregulated differentially expressed genes of MEF and chemically-induced XEN cells or MEF and OSKM-induced XEN cells. In our study we tried to exclude all genes that were upregulated in GETM-induced cells. For each study, upregulated differentially expressed genes identified using DESeq2 and EdgeR packages were merged as one list. A Venn diagram was constructed from the four lists to explore the overlap between the four studies. We identified 96 genes that were upregulated in our study and shared at least once with another study. The significance of the overlaps between our study and the other three studies were explored by the hypergeometric test and showed p value < 2.834E-16, 2.874E-07, and 7.835E-11.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

For experiments comparing differences between two groups, we used unpaired Student’s t test. Differences were considered significant when p value < 0.05. All experiments were repeated at least 3 times. For quantitative PCR experiments the results were normalized to the expression of the housekeeping control gene, Gapdh and presented as a mean ± standard deviation of two duplicate runs from a typical experiment. Unless indicated otherwise a representative experiment is shown for each figure. The statistical analyses for the high throughput analyses are depicted in the Method Details section. All

**Identification of a XEN-like signature**

The significance of the overlaps between our study and the other three studies were explored by the hypergeometric test and showed p value < 2.834E-16, 2.874E-07, and 7.835E-11.

|                      | Our Study | Zhao et al. (2015) | Parenti et al. (2016) | Li et al. (2017) |
|----------------------|-----------|--------------------|-----------------------|-----------------|
| The present Study    |           |                    |                       |                 |
| Zhao et al. (2015)   | 2.83E-16  |                    |                       |                 |
| Parenti et al. (2016)| 2.87E-07  | 0                  |                      | 0               |
| Li et al. (2017)     | 7.84E-11  | 0                  | 0                     |                 |

**DATA AND SOFTWARE AVAILABILITY**

All ATAC-Seq, ChIP-Seq and RNA-Seq were deposited in the Gene Expression Omnibus database (GEO) under accession number database: GSE98124. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.B. (yossibug@ekmd.huji.ac.il).
Supplemental Information

Direct Induction of the Three Pre-implantation Blastocyst Cell Types from Fibroblasts

Hana Benchetrit, Mohammad Jaber, Valery Zayat, Shulamit Sebba, Avital Pushett, Kirill Makedonski, Zvi Zakheim, Ahmed Radwan, Noam Maoz, Rachel Lasry, Noa Renous, Michal Inbar, Oren Ram, Tommy Kaplan, and Yosef Buganim
Supplementary Materials

Figure S1

A

BYKE ESCs

BYKE E14 TCs

Nanog-2A-EGFP

Nanog

Untagged allele

β-Tubulin

B

BYKE-ESC line

Count

BYKE-iPSC line

Count

C

Internal control

Targeted allele

OSKM-derived BYKE-iPSC line

GETM-derived BYKE-iTSC line

D

Infected MEFs

Eomes

Tlap2c

Esrb

Gata3

Myc

Actin

F

Transgene integration in stable GETMS-iPSC and iTSC clones

Gata3

Eomes

Tlap2c

Myc

Esrb

Relative DNA enrichment

Control

5F-iPSC clones

5F-iTSC clones

E

GETMS reprogramming (14d DMEM + 7d TSC medium + 14 d TSC culture medium without dox)

Nanog

Cdx2

Dapi

Gata3

Myc

Fibroblast

J

GTMS reprogramming (ESC medium)

Eosin

Nanog-2A-EGFP

U111-2A-tdTomato

Esrb-2A-TgBFP

ES-2A-EYFP-488

G

2° system K1H2-NgEtb-GETMS#1 (ESC medium)

Relative mRNA levels

No dox

dox (72h)

H

ESC medium

Nanog-2A-EGFP

U111-2A-tdTomato

Esrb-2A-TgBFP

I

2° system K1H2-NgEtb-GETMS#1 (TSC medium)

+Eomes

+Nanog-2A-EGFP

+U111-2A-tdTomato

+Esrb-2A-TgBFP
Figure S1. Related to Figure 1.

Esrrb overexpression drives pluripotency fate while Eomes overexpression drives TE fate in GETMS reprogramming combination. (A) Western Blot analysis depicting the protein levels of Nanog, Utf1, Esrrb and β-tubulin in the BYKE ESC system and V6.5 ESC control. (B) Bright field and fluorescence images of a 13.5 dpc chimeric male gonad, following BYKE ESC injection, showing the activity of the 3 pluripotency reports. (C, Left) Bright field, fluorescence images and FACS analysis of the 3 pluripotent reporters on isolated BYKE-iPSC clone. (C, Right) Bright field, fluorescence images and FACS analysis of the 3 TSC reporters on isolated BYKE-iTSC clone. (D) Western blot analysis of the five reprogramming factors (Gata3, Ttfap2c, Eomes, Myc and Esrrb) and the housekeeping control protein β-actin, 72h post dox exposure in MEFs infected with GETMS. (E) Cells were reprogrammed with GETMS in a protocol that promotes the generation of both iPSCs and iTSCs (Figure 1F) in the same dish. Bright field and immunostaining images of endogenous Nanog (green), endogenous Cdx2 (red) and Dapi (blue) showing mutually exclusive expression of Nanog and Cdx2 in their corresponding iPSC and iTSC colonies. Bottom row shows a magnification of the depicted rectangles from the top row. (F) qPCR showing the genomic integrations of each of the five factors into seven 5F-iPSC colonies and seven 5F-iTSC colonies. Single asterisk indicates a colony without Gata3 integration, two asterisks indicate colonies without Esrrb integration and three asterisks indicate a colony without Esrrb and Myc integration. Uninfected MEFs were used as control. Error bars presented as a mean ± standard deviation of two duplicate runs from a typical experiment out of 3 independent experiments. (G) qPCR of the GETMS (5F) genes normalized to the Gapdh housekeeping gene in KH2-NgEbUr-GETMS#1 2°MEFs after 72 hours of dox induction. Error bars presented as a mean ± standard deviation of two duplicate runs from a typical experiment out of 3 independent experiments. (H) Graph summarizing FACS analysis for the 3 pluripotency reporters in KH2-NgEbUr-GETMS#1 2°MEFs that were infected with additional single factor and were reprogrammed for 20 days followed by 10 days of dox removal in ESC medium. (I) FACS analysis for Nanog-2A-EGFP and Utf1-2A-tdTomato reporters in KH2-NgEbUr-GETMS#1 2°MEFs that were infected with additional single factor and were reprogrammed for 20 days followed by 10 days of dox removal in TSC medium. For Eomes-infected cells, bright field and fluorescence images of the 3 reporters are shown for one representative iTSC colony. (J, Left) FACS analysis for the 4 reporters in BYKE-MEFs infected with GTMS factors that were reprogrammed for 20 days followed by 10 days of dox removal in ESC medium. (J, Right) Representative bright field and fluorescence images of GTMS reprogrammed BYKE-MEFs showing triple-positive iPSC colony.
Figure S2

A

| Experiment number | Cell line            | Culture medium | # of recipient mice | # of injected embryos | # of 13.5 d.p.c recovered embryos/placentas (%) | # of 13.5 d.p.c chimeric embryos/placentas (%) |
|-------------------|----------------------|----------------|---------------------|-----------------------|-----------------------------------------------|-----------------------------------------------|
| 1                 | GETMS{Prl3d1} | S/Lf            | 2                   | 22                    | 11(50)                                        | 7(32)                                         |
| 2                 | GETMS{Prl3b1} | S/Lf            | 2                   | 20                    | 2(5)                                         | 1(5)                                          |
| 3                 | GETMS{Prl3b2} | S/Lf            | 1                   | 15                    | 4(26)                                        | 4(26)                                         |
| 4                 | GETMS{Prl3c1} | S/Lf            | 3                   | 34                    | 12(27)                                       | 3(14)                                         |
| 5                 | GETMS{Prl3c2} | S/Lf            | 2                   | 24                    | 11(46)                                       | 3(12.5)                                       |
| 6                 | GETMS{Prl3c3} | S/Lf            | 2                   | 21                    | 10(48)                                       | 10(48)                                        |
| 7                 | GETMS{Prl3c4} | S/Lf            | 1                   | 15                    | 12(80)                                       | 7(47)                                         |
| 8                 | GETMS{Prl3c5} | S/Lf            | 2                   | 56                    | 25(66)                                       | 17(30)                                        |
| 9                 | GETMS{Prl3c6} | S/Lf            | 2                   | 27                    | 9(30)                                        | 4(15)                                         |
| 10                | GETMS{Prl3c7} | S/Lf            | 1                   | 13                    | 5(38)                                        | 5(38)                                         |
| 11                | GETMS{Prl3c8} | S/Lf            | 1                   | 18                    | 9(50)                                        | 8(44)                                         |
| 12                | GETMS{Prl3c9} | S/Lf            | 2                   | 26                    | 2(8)                                         | 2(8)                                          |
| 13                | GETMS{Prl3c10}| S/Lf            | 2                   | 72                    | 4(15)                                        | 1(4)                                          |
| 14                | GETMS{Prl3c11}| S/Lf            | 2                   | 26                    | 4(15)                                        | 1(4)                                          |
| 15                | GETMS{Prl3c12}| S/Lf            | 1                   | 18                    | 9(50)                                        | 8(44)                                         |
| 16                | GETMS{Prl3c13}| S/Lf            | 2                   | 22                    | 12(55)                                       | 10(45)                                        |
| 17                | ESC{Prl3d1}    | S/Lf            | 3                   | 43                    | 9(21)                                        | 6(14)                                         |
| 18                | ESC{Prl3b1}    | S/Lf            | 2                   | 26                    | 2(8)                                         | 1(4)*                                         |
| 19                | ESC{Prl3b2}    | S/Lf            | 3                   | 39                    | 17(34)                                       | 12(51)*                                       |
| 20                | ESC{Prl3c1}    | S/Lf            | 3                   | 21                    | 12(57)                                       | 8(38)*                                        |
| 21                | ESC{Prl3c2}    | S/Lf            | 2                   | 25                    | 9(36)                                        | 2(8)* + 1 (4)                                 |
| 22                | ESC{Prl3c3}    | S/Lf            | 7                   | 100                   | 23(23)                                       | 10(10)* + 1 (1)                               |

B

idTomato (Placenta)

C

GETMS-TSCM1

D

Prl3d1  Prl3b1  Tpbbp  Prl2c2  Gcm1  Syna

E

GETMS-TSCM3

Prl3d1  Prl3b1

F

Day 0  Day 3  Day 6  Day 9  Day 12

Cell Count

0  200  400  600  800  1000

P1 (FL3)  P1 (FL3)  P1 (FL3)  P1 (FL3)  P1 (FL3)
Figure S2. Related to Figure 2.

5F-iTSCs are functional and are capable of differentiating into the various trophoblast lineages and of generating hemorrhagic lesions. (A) A table summarizing the generation of chimeric embryos/placentas with GETMS-iPSC, GETMS-iTSC clones and controls following blastocyst injection. All embryos and placentas were examined at 13.5 dpc. S- stands for serum, RI- stands for Rock inhibitor, 70:30- stands for TSC culturing medium, 2i- stands for two inhibitors (PD0325901 and CHIR99021) and *-stands for low contribution. (B) Contribution of tdTomato ESCs to 13.5 dpc placenta. A clear tdTomato signal in endothelial cells of placental blood vessels was detected by immunohistochemistry. (C, left) Hemorrhagic lesions 6-7 days following subcutaneous injection of GETMS-iTSC#1 line into nude mice. (C, right) H&E staining of paraffin sections of hemorrhagic lesions, showing necrotic tissue with blood and scattered trophoblastic cells. (D) qPCR of the indicated trophoblast lineage markers normalized to the Gapdh housekeeping gene in the indicated hemorrhagic lesions and controls (i.e. MEFs, bdTSCs and 6 days of bdTSC differentiation). Error bars presented as a mean ± standard deviation of two duplicate runs from a typical experiment out of 3 independent experiments. (E) qPCR of the indicated trophoblast lineage markers normalized to the Gapdh housekeeping control gene in GETMS-iTSCs#3, grown in differentiation media for the indicated time points. The expression level of these trophoblast differentiation markers was comparable to that of bdTSCs after differentiation (Benchetrit et al. 2015). Error bars presented as a mean ± standard deviation of two duplicate runs from a typical experiment out of 3 independent experiments. (F) Flow cytometric analysis of GETMS-iTSCs#3 grown in differentiation medium for the indicated time points, followed by propidium iodide (PI) staining. Scale bars show staining intensities; representative of chromosome copy number. The percentage of cells from each sample in every phase is indicated.
Figure S3. Related to Figure 3.

Distal and intronic ATAC-Seq peak analysis in the various reprogramming combinations revealed a strong biased toward the PrE/XEN lineage in combination that harbor Esrrb. (A) Eighteen ATAC-Seq clusters (Figure 3D) were grouped into 9 blocks, based on their similarity, and tested for functional annotations using GREAT. Graphs showing FDR values of GO annotation of biological processes that are enriched for each group in the indicated clusters. (B) Metagene plots (top) and heatmaps (bottom) show the ATAC-Seq signal of the 19 gene clusters found in (Figure 3C), for MEFs, bdTSC, iPSC, and indicated reprogramming factor combination lines. Each plot shows a 5Kb fragment at the promoter region (i.e. 2.5Kb upstream and downstream to the transcription start site (TSS)). Cluster ID is color coded. As shown, the promoter region is characterized by strong ATAC-Seq signal, which is generally ubiquitous to all conditions. However, some clusters, (e.g. cluster #17, 18 and 19), do show some matching between promoter ATAC-Seq signal and expression levels (Figure 3C). (C) Motifs were identified by comparing the top distal and intronic ATAC-Seq peaks in GETMS and GTMS to those in GETM treated cells (see also Table S4). (D) Average read count over the entire XEN-signature genes (n=225) for H3K27ac and H3K4me2 histone modifications in the indicated samples at the depicted genomic regions. TSS- transcription start site, TES- Transcription end site. (E) Genomic visualization of ATAC-Seq profiles (top), ChIP-Seq data of H3K4me2 and H3K27ac (middle) and RNA-Seq levels in FPKM (bottom; color coded: white- no expression, blue- intermediate expression and red- strong expression) in MEFs overexpressing either Esrrb alone (denoted “MEFs + S”), GTMS, or GETMS for 3 days. Red rectangles indicate H3K4me2 peak regions and green rectangles indicate H3K27ac peak regions. Myh6 and Oct4 loci are presented.
Figure S4. Related to Figure 4.

Esrrb induces a XEN-like state during GETMS reprogramming and following ESC-TSC transdifferentiation. (A) qPCR of the indicated genes normalized to the Gapdh housekeeping gene in MEFs, iPSCs, bdTSCs, bdXENs and GETMS-iXENs. For each gene the highest expressing sample was set to 1. Error bars presented as a mean ± standard deviation of two duplicate runs from a typical experiment out of 3 independent experiments. (B) Oct4-GFP MEFs were infected with dox-inducible GETMS or OSKM factors. 12 days post dox induction, Oct4-GFP-positive and negative cells were sorted from each combination of factors and used for qPCR analysis. A graph depicting the mRNA levels of Myh6 normalized to the Gapdh housekeeping gene in the indicated samples is shown. Error bars presented as a mean ± standard deviation of two duplicate runs from a typical experiment out of 3 independent experiments. (C) Oct4-GFP MEFs were infected with dox-inducible GETMS factors, with Eomes marked by GFP and Esrrb marked by tdTomato. 6 days post dox induction, Eomes-GFP positive and Esrrb-TdTomato-Low cells (GETMS E+/S\textsuperscript{Low}) and Eomes-GFP positive and Esrrb-TdTomato-High cells (GETMS E+/S\textsuperscript{High}), were sorted and used for qPCR analysis. A graph depicting the mRNA levels of Myh6 normalized to the Gapdh housekeeping gene in the indicated samples is shown. Error bars presented as a mean ± standard deviation of two duplicate runs from a typical experiment out of 3 independent experiments. (D) FACS analysis showing the percentage of Oct-GFP-positive cells for Oct4-GFP MEFs that were infected with either GETM or GETMS in combination with either empty vector (EV) or Gata6 and Sox17 and were reprogrammed for 14 days followed by 10 days of dox removal in 2i/L ESC medium. (E) qPCR of the indicated genes normalized to the Gapdh housekeeping gene during transdifferentiation of ZHBTc4 ESCs into TS-like cells after infection with either constitutively active Eomes or Esrrb or non-infected cells control. d0 refers to pre-dox addition. Error bars presented as a mean ± standard deviation of two duplicate runs from a typical experiment out of 3 independent experiments. (F) qPCR of the indicated genes normalized to the Gapdh housekeeping gene in stable colonies that were generated during ESC-TSC transdifferentiation in ZHBTc4 control (Esrrb\textsuperscript{WT}) and in ZHBTc4 Esrrb-overexpressing cells (Esrrb\textsuperscript{OE}) and in MEFs, ESCs, bdTSCs and bdXENs. (G) Quantification of the number of stable TS-like colonies and stable Xen-like colonies in transdifferentiated ZHBTc4 control and Esrrb-overexpressing cells. Error bars presented as a mean ± standard deviation of two duplicate runs from a typical experiment out of 3 independent experiments. (H) Bright field images depicting the morphology of stable TS-like cells (circled) and Xen-like cells (squared) generated during ESC-TSC transdifferentiation of ZHBTc4 Esrrb-overexpressing (Esrrb\textsuperscript{OE}) and control cells. (I) Immunofluorescence images for Cdx2 (Green) and Gata6...
(Red) in stable TS-like and XEN-like colony generated during ESC-TSC transdifferentiation in ZHBTc4 Esrrb-overexpressing cells (Esrrb\textsuperscript{OE}).