

# Comparative Parallel Multi-Omics Analysis During the Induction of Pluripotent and Trophectoderm States

Mohammad Jaber<sup>1,5</sup>, Ahmed Radwan<sup>1,5</sup>, Netanel Loyfer<sup>2,5</sup>, Mufeed Abdeen<sup>1</sup>, Shulamit Sebban<sup>1</sup>, Thorsten Kolb<sup>3</sup>, Marc Zapatka<sup>4</sup>, Kirill Makedonski<sup>1</sup>, Aurelie Ernst<sup>3</sup>, Tommy Kaplan<sup>2,\*</sup>, and Yosef Buganim<sup>1,\*</sup>

1. Department of Developmental Biology and Cancer Research, Institute for Medical Research Israel-Canada, The Hebrew University-Hadassah Medical School, Jerusalem 91120, Israel
2. School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel
3. Group Genome Instability in Tumors, DKFZ, Heidelberg, Germany and German Cancer Consortium (DKTK), 69120 Heidelberg, Germany
4. Division of Molecular Genetics, German Cancer Research Consortium (DKTK), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany
5. These authors contributed equally to this work
  - Correspondence should be addressed to Y.B. ([yossibug@ekmd.huji.ac.il](mailto:yossibug@ekmd.huji.ac.il)) or T.K. ([tommy@cs.huji.ac.il](mailto:tommy@cs.huji.ac.il)).
  - Lead contact: Yossi Buganim [yossibug@ekmd.huji.ac.il](mailto:yossibug@ekmd.huji.ac.il)

Short title: The trajectories to Pluripotency and Trophectoderm

21 Following fertilization, totipotent cells divide to generate two compartments in the early embryo: the  
22 inner cell mass (ICM) and trophoctoderm (TE). It is only at the 32-64 -cell stage when a clear segregation  
23 between the two cell-types is observed, suggesting a 'T'-shaped model of specification. Here, we  
24 examine whether the acquisition of these two states *in vitro* by nuclear reprogramming share similar  
25 dynamics/trajectories. We conducted a comparative parallel multi-omics analysis on cells undergoing  
26 reprogramming to Induced pluripotent stem cells (iPSCs) and induced trophoblast stem cells (TSCs), and  
27 examined their transcriptome, methylome, chromatin accessibility and activity and genomic stability.  
28 Our analysis revealed that cells undergoing reprogramming to pluripotency and TSC state exhibit specific  
29 trajectories from the onset of the process, suggesting 'V'-shaped model. Using these analyses, not only  
30 we could describe in detail the various trajectories toward the two states, we also identified previously  
31 unknown stage-specific reprogramming markers as well as markers for faithful reprogramming and  
32 reprogramming blockers. Finally, we show that while the acquisition of the TSC state involves the  
33 silencing of embryonic programs by DNA methylation, during the acquisition of pluripotency these  
34 specific regions are initially open but then retain inactive by the elimination of the histone mark,  
35 H3K27ac.

## 36 Cellular Specification during Early Embryogenesis

37 Fertilization of an oocyte by a sperm initiates robust epigenetic reprogramming of the DNA content  
38 within the newly formed cell, resulting in a totipotent zygote that holds the potential to produce all  
39 embryonic and extra-embryonic tissues<sup>1</sup>. Several divisions later, an early blastocyst is formed, containing  
40 two compartments that are more committed: an inner cell mass (ICM) which contains pluripotent cells  
41 (epiblast (Epi)) that will form the embryo proper, and an outer layer of trophoblast (TE) cells, which  
42 will give rise to components of extra-embryonic tissues such as the placenta<sup>2-4</sup>.

43 Exactly how the specification between the ICM and TE cells is made during embryogenesis is not fully  
44 understood, although several models have been suggested<sup>2-4</sup>. Recently, the transcriptional trajectory  
45 from zygote to blastocyst has been described using single-cell transcriptomic data<sup>5-11</sup>. Interestingly,  
46 while clear transcriptional changes are found between the different stages (i.e. zygote, 2-cell stage, 4-  
47 cell stage, 8-cell stage, morula and blastocyst), the transcriptional heterogeneity within the cells of each  
48 group before blastocyst formation is relatively mild. Although some genes, like *Sox21*, were shown to  
49 exhibit transcriptional heterogeneity even within the 4-cell stage<sup>6</sup>, the overall transcriptome is relatively  
50 similar between the two groups of cells. This suggests a 'T'-shaped model, where cells at each stage  
51 undergo relatively similar transcriptional changes before segregation, and separate into two distinct  
52 cells types, the ICM and TE, only at the morula/early blastocyst stage. This notion is supported by  
53 multiple evidence; first, each of the 2-8 first cells of the embryo can give rise to both TE and ICM.  
54 Second, cells of the outer layer of the morula stage have been observed to migrate into the inner layer  
55 and become pluripotent cells. This suggests a dynamic chromatin landscape and transcriptome that are  
56 relatively analogous between the cells before the final specification<sup>3,4</sup>.

57

## 58 Somatic Epigenetic Reprogramming to Pluripotency and Trophoblast Stem Cell State

59 Epigenetic reprogramming of a somatic nucleus to pluripotency or to a TE state has been achieved *in*  
60 *vitro* by somatic cell nuclear transfer (SCNT<sup>12</sup>) or by forced expression of a defined number of  
61 transcription factors<sup>13-19</sup>. While ectopic expression of Oct4, Sox2, Klf4 and Myc (OSKM) induces the  
62 formation of pluripotent stem cells (PSC, the *in vitro* counterpart of the ICM-Epi)<sup>17</sup>, we and others have  
63 shown that ectopic expression of Gata3, Eomes, Tfap2c and Myc (GETM, or Ets2 instead of Myc) induces  
64 the formation of trophoblast stem cells (TSC, the *in vitro* counterpart of the TE)<sup>13,16</sup>. Importantly, in both  
65 reprogramming systems, the resulting cells are equivalent to their *in vitro* blastocyst-derived  
66 counterparts in their transcriptome, epigenome and function<sup>13,15-17,20</sup>.

67 However, while nuclear reprogramming to pluripotency and TE state during fertilization or in SCNT  
68 occurs within 2-3 days<sup>21</sup>, nuclear reprogramming by defined transcription factors is a long and inefficient  
69 process<sup>13,15-17</sup>. Intrigued by these fundamental differences, scientists have devoted the last decade to  
70 monitoring and describing the various mechanisms, stages and pathways that underlie somatic cells  
71 undergoing reprogramming to pluripotency<sup>14</sup>. These major efforts have revealed key aspects in nuclear  
72 reprogramming which also explain, at least partially, the low efficiency of the process and describe in  
73 detail the trajectories somatic cells undergo in their way to become iPSCs.

74 However, while some properties of the reprogramming process to pluripotency have been illuminated,  
75 the characterization of the reprogramming process to the TSC state, or more intriguing, concomitant  
76 comparative multi-omics analysis of cells undergoing reprogramming to the TSC and pluripotent states  
77 has never been performed.

78 Here, we describe the trajectories and key elements that regulate and characterize cells undergoing  
79 reprogramming to iTSCs and iPSCs in depth. We show that in contrast to early embryonic cells,  
80 fibroblasts transduced with GETM or with OSKM mostly follow a 'V'-shaped model where cells acquire,  
81 from the onset of the reprogramming process, a unique and specific chromatin and transcriptional  
82 programs that are mostly mutually exclusive and important for the induction of each state. Surprisingly,  
83 this 'V'-shaped behavior was also evident at the methylation levels, where correlation with  
84 transcription is relatively low. Using single-cell analysis, we revealed unique and previously unknown  
85 markers for each reprogramming system. Moreover, chromatin accessibility and activity analyses  
86 identified many global reprogramming blockers such as Usf1/Usf2, Nrf2 and MafK along with other  
87 oxidative stress response genes that significantly hinder both reprogramming systems but with a  
88 different dynamic.

89 Lastly, by integrating all the data together we illuminated key aspects that characterize each fate.  
90 Remarkably, by comparing both systems we show that from the onset of the reprogramming process  
91 OSKM define regions that are important for the development of the heart and brain, two most essential  
92 organs for the developing embryo. Moreover, we show that while GETM shut off the embryonic  
93 program by DNA methylation, OSKM open these regions but retain them as inactive by eliminating the  
94 histone mark H3K27ac. In conclusion, besides providing the first multi-layer characterization of cells  
95 undergoing reprogramming to the TSC state, our approach of conducting concomitant and comparative  
96 multi-omics analysis of cells acquiring both pluripotency and TSC state allowed us to identify previously  
97 unknown properties for OSKM reprogramming as well.



## 98 Results

### 99 The Trajectory from the Zygote to the Blastocyst Stage is following a 'T'- Shaped behavior

100 The trajectory from the zygote to the blastocyst stage (**Fig. 1A**) has recently been described by several  
101 studies using single-cell transcriptomic data<sup>5-11</sup>. Using principal component analysis (PCA), Deng et al.  
102 suggested that the trajectory from zygote to blastocyst follows a 'U' shape<sup>5</sup>, in which PC1 separates  
103 between the zygote/early 2-cell stage and blastocyst and PC2 separates between the other stages,  
104 namely the 2-16-cell stage and the zygote/blastocyst (**Fig. 1B**). However, since the zygote and the 2-cell  
105 stage are considered totipotent and thus harbor a unique transcriptome, we reanalyzed the data by  
106 excluding these two stages. Interestingly, the re-analyzed PCA revealed a clear 'T'-like shape where PC1  
107 separates between the 4-cell stage and the blastocyst and PC2 between the ICM and TE (**Fig. 1C and**  
108 **Extended Data Fig. 1A**). More importantly, while both analyses (**Fig. 1B-C**) suggest that a clear  
109 transcriptional shift between different stages occurs during early embryogenesis, in both analyses the  
110 heterogeneity within each group was mild, indicating that the cells undergo relatively similar changes  
111 during embryogenesis and before specification. A 'T'-shaped behavior of early embryonic cells was  
112 similarly observed in the datasets of Guo et al.<sup>7</sup>, strengthening the notion that only at the morula/early  
113 blastocyst stage, a clear transcriptional segregation between cells of the same developmental stage can  
114 be witnessed (**Fig. 1D and Extended Data Fig. 1B**).

115 We sought to determine whether this 'T'-like behavior also characterizes somatic cells undergoing  
116 reprogramming to pluripotency and TSC state. In general, the reprogramming process of fibroblasts is  
117 characterized by multiple steps: (1) loss of the somatic cell identity, (2) rapid proliferation, (3)  
118 mesenchymal to epithelial transition (MET), (4) metabolic shift, (5) stochastic gene expression, (6)  
119 epigenetic switch, (7) silencing of the exogenous factors by methylation and finally (8) the stabilization  
120 of the core cellular circuitry<sup>14,32</sup>. We proposed three possible models, 'T', 'Y' and 'V', which may  
121 represent the trajectory of fibroblasts undergoing reprogramming into iPSCs and iTSCs (**Extended Data**  
122 **Fig. S1C**). The 'T'-shaped model predicts that fibroblasts undergoing reprogramming into iPSCs or iTSCs  
123 will undergo comparable transcriptional and epigenetic changes during the conversion and that a  
124 separation between the two cell types will occur only at the end of the reprogramming process, similar  
125 to the cells of the early embryo. The 'Y'-shaped model predicts that only genes and regulatory elements  
126 that are responsible for early and general processes like loss of somatic cell identity, proliferation and  
127 MET will be shared between the two systems, after which each process will take a different path toward  
128 its own fate. The 'V'-shaped model predicts that although some early processes are the same between

129 the two systems, each process will mostly employ different set of genes and regulatory elements to  
130 achieve its own unique fate.

131 To understand which of the three proposed models most accurately represents the reprogramming  
132 process toward the two states, we performed a parallel comparative multi-omics analysis on fibroblasts  
133 undergoing reprogramming to iPSCs by OSKM or to iTSCs by GETM factors simultaneously (**Fig. 1E**). We  
134 utilized our previously developed system for distinguishing between pluripotency and TSC  
135 reprogramming, BYKE mouse embryonic fibroblasts (MEFs), which contain 4 unique knock-in fluorescent  
136 reporters: (1) Nanog-2A-EGFP, a cytoplasmic reporter that marks specifically pluripotent cells, (2) Elf5-  
137 EYFP-NLS, a nuclear reporter that marks specifically TSCs and (3) Utf1-2A-tdTomato and (4) Esrrb-2A-  
138 TagBFP, cytoplasmic reporters that mark both cell types (**Fig. 1F-I**, <sup>42</sup>). We then measured the  
139 transcriptome (bulk RNA-seq and single-cell RNA-seq (SC-RNA-seq)), methylome (reduced  
140 representative bisulfite sequencing, RRBS), chromatin accessibility (ATAC-seq), chromatin activity (ChIP-  
141 seq for H3K4me2 and H3K27ac) and genomic stability (CNVs) at different time points along the  
142 reprogramming processes and compared between the two systems (**Extended Data Fig. 1D**).

143

#### 144 **Bulk Transcriptomic Analysis Depicts Unique Transcriptional Profiles for Each Reprogramming system**

145 The transcriptional landscape of reprogrammable cells is the easiest and most robust examination to  
146 compare two parallel reprogramming systems (i.e. iPSC vs iTSC reprogramming). To explore the global  
147 transcriptomic changes of cells undergoing reprogramming to the pluripotent and TSC states, we  
148 reprogrammed BYKE MEFs into iPSCs with OSKM and into iTSCs with GETM and collected cell pellets in  
149 duplicates every three days along the reprogramming process (**Extended Data Fig. 1D**). The parental  
150 BYKE MEFs, the final cells (i.e. iTSCs and iPSCs) and their blastocyst-derived control counterparts (i.e.  
151 bdTSCs and ESCs) were analyzed as well.

152 We used full-transcript RNA-seq to estimate expression levels and transcribed isoforms, with a total  
153 depth of 20M reads per replicate. Initially we conducted principal component analysis (PCA) to observe  
154 the trajectory cells undergo during the reprogramming process to iTSCs and iPSCs (**Fig. 2A-F**). In  
155 addition, we extracted the gene loadings associated with the first two principal components in each PCA  
156 plot to reveal those genes that drive the distinction between the different stages/steps of each  
157 reprogramming process and between the two reprogramming systems (**Extended Data Fig. 1E-J**).  
158 Notably, cells undergoing reprogramming to a TSC or pluripotent state exhibited a very different  
159 transcriptional landscape, as analyzing both processes together in the same PCA plot, generated a 'V'-

160 like shape, starting from the beginning of the process (**Figs. 2C and 2F and Supplementary Figs. 1G and**  
161 **1J**). This suggests that major transcriptional changes separate the two systems.  
162 Interestingly, while the reprogramming process toward iPSCs showed constant and gradual  
163 transcriptional changes along the process until the stabilization of the final cells (**Figs. 2A and 2D and**  
164 **Supplementary Figs. 1E and 1H**), the reprogramming process toward iTSCs showed two main waves of  
165 transcriptional change where the first wave occurred very early (i.e. already at day 3, PC2, **Figs. 2B and**  
166 **Extended Data Fig. 1F**), followed by subtle changes in transcription until day 21 (**Fig. 2E and Extended**  
167 **Data Fig. 1I**) and a second wave which is initiated following transgene expression removal and is  
168 important for the activation of the core TSC circuitry (i.e. PC1, **Fig. 2B and Extended Data Fig. 1F**).  
169 We believe that these differences between OSKM and GETM reprogramming are partially due to the  
170 nature of each reprogramming process. In our experience, while iPSC colonies may be stabilized during  
171 the reprogramming process and in the presence of transgenes, iTSC colonies cannot. Only when  
172 transgenes expression is shut off (i.e. removal of dox) stable iTSC colonies emerge.  
173 Next, we took the ~10,000 most differentially expressed genes among all samples and clustered them  
174 together, yielding 27 unique clusters (**Fig. 2G and Supplementary Table 1**). Clusters 1, 4, 6, 10, 11, 20  
175 and 23 contain MEF-specific genes (i.e. gene ontology (GO) terms of connective tissue development,  
176 actin filament organization, extracellular matrix organization and angiogenesis) that are downregulated  
177 during GETM and OSKM reprogramming but with unique dynamics for each cluster and system  
178 (**Supplementary Table 1**). Clusters 7, 16, 17, 18, 19, 24, 25 and 27 are specific to the TSC reprogramming  
179 process. Interestingly, most of these clusters involve genes which are important to metabolism and cell  
180 cycle regulation (e.g. GO terms of ribonucleoprotein complex biogenesis, tRNA/rRNA/ncRNA metabolic  
181 processes and regulation of mitotic cell cycle, **Supplementary Table 1**). Clusters 2, 8, 12 and 15 are  
182 specific to iPSC reprogramming and contain genes that participate in cell junction organization, Ras  
183 protein signal transduction, regulation of vasculature development, histone modification and Wnt  
184 signaling (**Supplementary Table 1**). Cluster 3, 5 and 9 are shared between the two processes and  
185 compose of genes that regulate cell cycle, DNA repair and Wnt signaling pathway (**Supplementary Table**  
186 **1**). We noticed that most genes behaved differently between the two reprogramming systems, even in  
187 early and shared dynamics such as proliferation, chromatin remodeling and mesenchymal to epithelial  
188 transition (MET, **Extended Data Fig. 2A-D**). While key mesenchymal genes and regulators of epithelial to  
189 mesenchymal transition (EMT) are downregulated in both systems as expected (**Extended Data Fig. 2C,**  
190 bottom of the heatmap), indicating loss of fibroblastic identity, particular mesenchymal and MET-  
191 specific genes are uniquely expressed in each reprogramming system (**Extended Data Fig. 2C-D**).

192 Another example for an important difference which can be observed already in early stages of  
193 reprogramming between the two systems is a metabolic shift which occurs in two waves (i.e. day 3-9  
194 and days 12-21) in iTSC reprogramming. This shift, which plays a role in translation regulation activity  
195 and RNA processing, is completely absent in iPSC reprogramming (**Extended Data Fig. 2E-G and**  
196 **Supplementary Table 1**). Moreover, even when similar GO terms were annotated between the two  
197 systems, each reprogramming system utilized a different set of genes to execute the process. Figure 2H  
198 shows an example where both iTSC and iPSC reprogramming systems activated placenta/trophoblast-  
199 specific genes with ‘placenta development’ GO annotation, though while OSKM activated trophoblast  
200 differentiation genes, GETM activated trophoblast stem cell genes (marked by green, **Fig. 2H and**  
201 **Supplementary Table 1**). In conclusion, bulk transcriptomic analyses suggest a ‘V’-shaped behavior  
202 where GETM and OSKM operate distinctively to reprogram the somatic nucleus and already at the onset  
203 of the process take completely different transcriptomic routes, with minimal overlapping genes and  
204 signaling pathways.

205

#### 206 **Single-Cell Transcriptomic Analysis Suggests No Overlap between GETM and OSKM Reprogramming**

207 Bulk transcriptional analysis is a powerful tool to describe the global transcriptional changes that occur  
208 in most cells during the reprogramming process. However, it is lacking the sensitivity to identify the  
209 small fraction of cells that are destined to become iPSCs or iTSCs (typically only 1-5% of the cells for iTSC  
210 reprogramming and 10-20% for iPSC reprogramming acquire the final fully reprogrammed state<sup>13,42</sup>, **Fig.**  
211 **1F-G**). Moreover, due to population averaging, one cannot detect a small group of cells that might  
212 harbor a transcriptional profile that is similar between the two reprogramming systems, suggesting a ‘T’  
213 or ‘Y’ behavior. To overcome this limitation, we conducted single-cell analysis on GETM and OSKM  
214 reprogrammable cells. While multiple studies employed the single-cell RNA-seq (SC-RNA-seq)  
215 technology to probe the transcriptome of individual cells undergoing reprogramming to  
216 pluripotency<sup>34,35,43,44</sup>, similar characterization of the reprogramming process toward the TSC state or a  
217 comparative analysis between the two processes has never been done before.

218 To evaluate the transcriptomes of individual cells undergoing reprogramming into iPSCs and iTSCs, we  
219 exploited the 10X Genomics platform and profiled the transcriptome of ~16,000 single cells at two time  
220 points (days 6 and 12) along the reprogramming process toward iPSCs and iTSCs. We chose these time  
221 points as they represent the stochastic gene expression phase that occurs in the two reprogramming  
222 systems and thus are expected to show the highest variation between individual cells. UMAP analysis  
223 for both days, 6 and 12, demonstrated two distinct clusters of cells; one for GETM and one for OSKM

224 reprogramming (**Fig. 3A-B**). No overlapping cells were found between the two systems, indicating that  
225 each process acquired a completely different transcriptional profile, again suggesting a 'V'-shaped  
226 model. Using EnrichR mouse gene atlas we identified the different transcriptional fates reprogrammable  
227 cells undergo during each reprogramming process. As the reprogramming process to iTSCs is much less  
228 efficient compared to OSKM reprogramming, we mostly identified non-reprogrammable MEFs ( $p \leq 8.9e-$   
229  $58$ ) in the GETM reprogramming process with a very small fraction of cells with MEF-like transcriptional  
230 profile in OSKM reprogramming ( $p < 4.1e-7$ , cluster 1 in **Extended Data Fig. 3A-B**). Interestingly, both  
231 processes contained cells with a transcriptional profile partially similar to that of the placenta (Cluster 3  
232 for GETM and cluster 14 for OSKM, **Fig. 3A-B and Extended Data Fig. 3A**). This is in agreement with the  
233 bulk transcriptomic analysis that showed gene ontology of placenta development with unique gene set  
234 for each system (**Fig. 2H**). Other than MEFs and placental cells, different group of reprogrammable cells  
235 activated genes that are enriched in epidermis, dorsal root ganglia, mast cells and bone marrow in  
236 OSKM reprogramming and umbilical cord, bladder and NK cells in GETM reprogramming (**Fig. 3A-B and**  
237 **Extended Data Fig. 3A-B**). Epidermis, placenta and neuronal fates have previously been observed in  
238 OSKM reprogramming, strengthening our findings<sup>43</sup>. We next identified known and unknown stage-  
239 specific markers for each reprogramming process (**Fig. 3C-F**). For iTSC reprogramming, non-infected  
240 MEFs or refractory MEFs were identified using the known mesenchymal markers Thy1, Col1a2 and  
241 Postn. Bgn, Tagln and Scmh1 mark both MEFs and cells that have succeeded to initiate the  
242 reprogramming process. Dusp9, Dlk1, Cdca3 and Bex4 mark most reprogrammable cells that are in the  
243 midst of the reprogramming process prior to any fate decision, and Cbs, Wnt6, Pgf, Peg10, Cd82, Adssl1,  
244 Stard10 and Ppp2r2c mark cells that are either differentiated trophoblasts (Cluster #3) or those that are  
245 probably destined to become iTSCs, based on their unique intermediate stemness gene signature (part  
246 of cluster 5 and the junction between 5 and 3, **Fig. 3C-D and Extended Data Fig. 3C**).

247 For OSKM reprogramming, Ly6g6c, Cd13, Krt16, Ecam1 and Cd34 represent cells that underwent MET and  
248 have acquired a partial epidermal fate. Krtdap, Epcam, and Tdh are markers for most OSKM  
249 reprogrammable cells but cannot predict successfully which cells are destined to become iPSCs.  
250 Interestingly, while Anxa3 marks all the cells that took a failed route toward pluripotency, high levels of  
251 TdGF1 and Cenpf represent a successful trajectory to reprogramming based on gene expression (**Fig. 3E-F**  
252 **and Extended Data Fig. 3D**). As we and others have previously noted<sup>14,23,45</sup>, while Sox2, Dppa5a and to a  
253 lesser extent Tet1 stringently mark fully reprogrammed cells, Oct4 fails to do so (**Fig. 3E-F and Extended**  
254 **Data Fig. 3D**).

255 We next searched for genes that can robustly distinguish between GETM-reprogrammable cells and  
256 OSKM-reprogrammable cells. We identified four unique and specific markers for each system. While  
257 *Arhgdib*, *Id3*, *Tm4sf1* and *Egfl7* mark specifically most GETM-reprogrammable cells, *Shisa8*, *Fetub*, *Slc7a3*  
258 and *Tdh* are uniquely expressed in almost all OSKM-reprogrammable cells (**Fig. 3G-H**). Interestingly, the  
259 proliferation rate of the two systems was different as well, though both reprogramming combinations  
260 contained *Myc*. Based on the expression of proliferation gene signature, it was clear that OSKM-  
261 reprogrammable cells proliferate much faster than GETM-reprogrammable cells, while non-  
262 reprogrammable cells from both systems expressed these genes to a very low level (**Extended Data Fig.**  
263 **3E**). This is in agreement with the proliferation rates of ESCs, whereby ESCs exhibit a very rapid doubling  
264 time of every 10-14 hours<sup>46</sup>.

265 In conclusion, these results illuminate the various fates, stage-specific markers and unique identifiers for  
266 each reprogramming system and strengthen the notion that each reprogramming process takes a  
267 distinct trajectory toward its own fate.

268

### 269 **Unique Methylation Dynamics for GETM and OSKM Reprogramming**

270 One of the crucial aspects of nuclear reprogramming is the ability to erase the epigenetic landscape of  
271 the somatic nucleus and construct a new epigenetic profile that is similar to the target cells (e.g. ESCs for  
272 iPSCs and bdTSCs for iTSCs). One important epigenetic mark is DNA methylation, which allows chromatin  
273 condensation and silencing of specific loci along the genome<sup>14</sup>.

274 To assess the methylation landscape of somatic cells undergoing reprogramming to iTSCs and iPSCs, we  
275 applied the reduced representation bisulfite sequencing (RRBS) technique to capture the CpG  
276 methylation landscape of the cells as a representation for the global methylation changes. GETM and  
277 OSKM-reprogrammable cells from different time points (**Extended Data Fig. 1D**), as well as the parental  
278 fibroblasts, bdTSCs, iTSCs, ESCs and iPSCs were subjected to RRBS and analyzed.

279 We used the K-means algorithm to classify ~130,000 genomic regions (blocks) shared amongst all  
280 samples during reprogramming to a TSC or pluripotent state, and generate 100 unique clusters where  
281 some of the clusters contained tiles that are specific to TSCs and ESCs, other to MEFs and the vast  
282 majority to reprogrammable cells (**Extended Data Fig. 4A**). Average DNA methylation levels per sample  
283 per cluster were then projected onto the first two principal components which generated gradual and  
284 time-dependent methylation dynamics for each reprogramming system with a clear 'V'-shaped  
285 trajectory, where PC1 represents the OSKM trajectory and PC2 GETM trajectory (**Fig. 4A**). The accuracy  
286 of the time-dependent methylation trajectory in the two reprogramming systems was surprising, given



287 that there is often poor correlation between methylation degree and gene activation, and the unique  
288 transcriptional profiles that characterize intermediate reprogrammable cells (**Fig. 2A and 2C**).

289 One clear difference between OSKM and GETM reprogramming was the overall dynamic of methylation  
290 changes during the reprogramming process (**Fig. 4B**). While OSKM-reprogrammable cells predominately  
291 lose methylation on CpG-enriched sites during the reprogramming process, GETM-reprogrammable cells  
292 mostly gain methylation on CpG-enriched sites either in the middle or gradually until the end of the  
293 reprogramming process (**Fig 4B-D and Extended Data Fig. 4B-D**).

294 Mammalian placentas are unique in their methylation landscape as they contain regions in the genome  
295 that are highly methylated in gene bodies and regions that are only intermediately methylated  
296 (40=60%)<sup>47</sup>. In accordance with that, our unbiased analysis identified two unique clusters that contain  
297 intermediately methylated regions only in the final and stabilized iTSCs/TSCs (**Extended Data Fig. 4E**).  
298 These results suggest a unique mechanism of methylation/demethylation that occurs only when the  
299 core TSC circuitry is activated.

300 We then associated the various tiles from the different clusters to their neighboring genes and  
301 performed GO analysis using GREAT<sup>48</sup> (**Supplementary Table 2**). Interestingly, while clusters associated  
302 with gradual loss of methylation in OSKM reprogramming involve genes that participate in the  
303 maintenance of the fibroblastic identity, apoptosis and multiple somatic cell properties (e.g. GO terms of  
304 actin cytoskeleton organization, focal adhesion, regulation of small GTPase mediated signal  
305 transduction, fat differentiation, muscle development, vasculogenesis and keratinocyte differentiation),  
306 the singular cluster among these which exhibits early demethylation in both systems is composed of  
307 genes that participate in somatic stem cell maintenance, immune system development and regulation of  
308 growth (**Supplementary Table 2**). While demethylation of regions that play a role in stemness and  
309 growth can be expected in the two systems, the identification of a set of genes that is enriched in the  
310 immune system, in methylation pattern and RNA in both systems, is intriguing (**Extended Data Fig. 3A-B**  
311 **and Supplementary Table 2**).

312 Clusters that are associated with gradual gain of methylation specifically in GETM reprogramming mostly  
313 include genes that negatively regulate metabolic processes involved in RNA production and transcription  
314 (e.g. RNA biosynthesis, macromolecule metabolic processes, nitrogen compound metabolic process). In  
315 accordance with their identity as extraembryonic cells, clusters that involve methylation only at the final  
316 step of the reprogramming process and in the fully reprogrammed cells are comprised of genes that are  
317 essential for embryo development, neuronal lineage development and somatic cell differentiation at  
318 large (**Supplementary Table 2**).

319 These data suggest that while GETM factors utilize DNA methylation to shut off all master genes  
320 important for executing the embryonic development program, OSKM first open these regions and  
321 subsequently regulate their expression by histone modifications (as will be discussed in the next  
322 section). Of special note is the neuronal lineage: While OSKM activates this program, GETM induces its  
323 silencing.

324 Given the fact that pluripotent cells and TSCs share many stemness genes (e.g. Sall4, Esrrb, Sox2, Lin28  
325 etc.<sup>13,42</sup>), we next asked whether we can identify methylation differences in their regulatory elements  
326 during reprogramming and in the final cells. We selected 6 genomic loci that contain tiles for genes that  
327 are either specific to pluripotent cells (Slc15a1 and Tex19.2), specific to TSCs (Eomes and Bmp8b) or  
328 shared between the two cell types (Sall4 and Stmn2). Interestingly, only few tiles on regulatory elements  
329 (e.g. tile block number 2 and 3 in Sall4 locus) were methylated/hypomethylated similarly between  
330 iPSCs/ESCs and bdTSCs/iTSCs and different from MEFs, weakening the notion of widespread shared  
331 regulatory elements between the two cell types (**Fig. 4E**). Most tiles on regulatory elements were  
332 methylated oppositely between the two cell states (**Fig. 4E and Extended Data Fig. 4F-G**), indicating  
333 tight and cell type-specific regulation for each reprogramming process.

334 Taken together, these data suggest that although the acquisition of the final methylation landscape of  
335 GETM and OSKM reprogrammable cells is gradual and time-dependent, the methylation level and  
336 deposition is unique for each reprogramming process, even in genes that are expressed in both cell  
337 types.

338

### 339 **Chromatin Accessibility and Activity of Cells Undergoing Reprogramming to Pluripotent and TSC states** 340 **Demonstrate Unique Chromatin Dynamics for each Reprogramming Process**

341 One of the properties of reprogramming factors is their ability to open closed chromatin by recruiting  
342 chromatin remodelers and transcriptional machinery to heterochromatin<sup>49,50</sup>. One approach to assess  
343 chromatin accessibility and activity is to employ Assay for Transposase-Accessible Chromatin using  
344 sequencing (ATAC-seq)<sup>51</sup> in conjunction with chromatin immunoprecipitation and sequencing (ChIP-seq)  
345 for specific histone marks.

346 To assess which sites are open and active and which are closed during early stages of reprogramming to  
347 iPSCs and iTSCs, we collected cells (50,000 cells per replicate for ATAC-seq and 50 million cells per  
348 replicate for ChIP-seq for H3K4me2 and H3K27ac) at day 3, 6 and 9 of the reprogramming process to  
349 iTSCs and iPSCs (**Extended Data Fig. 1D**). As a control, we profiled chromatin accessibility and activity of  
350 the parental MEFs, the final reprogrammed cells (i.e. iTSCs and iPSCs) and their blastocyst-derived



351 controls. We chose H3K27ac and H3K4me2 because H3K27ac marks both active promoters and distal  
352 enhancers while H3K4me2 is enriched in *cis*-regulatory regions, particularly in promoters, of  
353 transcriptionally active genes but more importantly it also marks genes primed for future expression<sup>52,53</sup>.  
354 Overall, we analyzed 170,658 peaks for ATAC-seq, 498,376 for H3K27ac and 770,274 for H3K4me2.  
355 These experiments allowed us to map the regions that are open and active early in the reprogramming  
356 process and those that are primed to be active later as well as closed regions in each reprogramming  
357 system.

358 In accordance with the transcriptome and methylome results, PCA on datasets of chromatin accessibility  
359 (ATAC-seq) and activity (i.e. H3K27ac and H3K4me2) revealed two separate 'V'-shaped trajectories  
360 distinguishing OSKM from GETM reprogramming, already at the beginning of the process (**Fig. 5A-C**).  
361 These results suggest that OSKM and GETM remodel the chromatin at different regions. Following this  
362 observation, we asked whether the distribution of the various peaks along the genome (i.e. promoters,  
363 exons, introns, UTRs, TSS and intergenic) is unique to each reprogramming system or a global  
364 reprogramming phenomenon. We found that while the location of the various peaks along the genome  
365 is mostly different between the two reprogramming systems (**Extended Data Fig. 5A-C**), the distribution  
366 of the peaks is very similar (**Fig. 5D-F**).

367 Interestingly, while shared OSKM and GETM ATAC-seq peaks and, to a lesser extent H3K4me2 peaks,  
368 exhibit a significant enrichment in promoters and exons, peaks that are unique to each reprogramming  
369 process are mostly localized to introns and intergenic regions (**Figs. 5D and 5F**). EnrichR GO analysis on  
370 the genes associated with the shared open promoters suggested that many of these active promoters  
371 are associated with response to the lentiviral infection itself ( $p \leq 0.001$ ). No significant differences in the  
372 distribution of H3K27ac is found between the various samples. These data suggest that the  
373 reprogramming process at large follows the same rules of genomic remodeling (i.e. beginning with  
374 robust opening of intronic and intergenic regions in conjunction with promoter closing), but each  
375 reprogramming system remodels the chromatin at different loci along the genome in accordance with  
376 its final cellular fate.

377 We then classified the various peaks (mean ATAC-seq, H3K27ac and H3K4me2 at  $\pm 5$ Kb) based on their  
378 behavior in the two reprogramming processes (**Extended Data Fig. 6A-D**). We identified 4 distinct  
379 patterns: (1) 1,605 genomic regions which appear predominantly in OSKM reprogramming in which their  
380 H3K27ac signal is typically the strongest at day 6 and is accompanied by a matching H3K4me2 signal, but  
381 with no dynamic change in DNA accessibility (**Fig. 5G**). (2) 1,716 GETM-specific regions that are marked  
382 by H3K4me2 and H3k27ac already at day 3 but chromatin accessibility is gained only later in the process

383 (i.e. day 9). Intriguingly, a mirror image can be seen in these regions during OSKM reprogramming.  
384 There, chromatin accessibility is mildly open and remained unchanged but a significant increase in  
385 H3K27ac and H3K4me2 signals is observed at later stages of reprogramming (**Fig. 5H**). (3) 2,848 regions  
386 that are open and active in both reprogramming systems but lose activity (i.e. H3K27ac and H3K4me2  
387 signal) over time exclusively in OSKM reprogramming (**Fig. 5I**). (4) 464 genomic regions that are open in  
388 both reprogramming processes, but while in GETM reprogramming they are active all along the process,  
389 in OSKM reprogramming they gain activity exclusively at later stages (**Fig. 5J**). We used GREAT to test  
390 these genomic regions for enriched annotations. We found that group 1 is associated with cellular  
391 response to leukemia inhibitory factor ( $p \leq 8.12e-23$ ), Ras guanyl-nucleotide exchange factor activity  
392 ( $p \leq 5.0e-9$ ) and represents a phenotype of embryonic lethality between implantation and placentation in  
393 the mouse ( $p \leq 1.5e-12$ ). Group number 2 is associated with cell migration and motility ( $p \leq 2.6e-20$ ), cell  
394 adhesion ( $p \leq 4e-14$ ), extracellular matrix ( $p \leq 1.0e-8$ ), insulin-like growth factor binding ( $p \leq 2.7e-8$ ) and  
395 heparin binding ( $p \leq 3.0e-7$ ), all relevant to trophoblast differentiation and placentation. Interestingly,  
396 once again, this group of genes is significantly enriched in cells of the immune system, giving rise to the  
397 GO term of mouse phenotype of autoimmune response ( $p \leq 3.4e-13$ ). These results suggest a mechanism  
398 by which GETM induce a TSC fate by gradually opening and activating trophoblast-specific regions that  
399 are important for TSC function. In contrast, OSKM do not change the accessibility of these regions, which  
400 remain mildly open, but then gradually activate them, which explains the small fraction of differentiated  
401 trophoblast cells present in OSKM reprogramming.

402 Group number 3 contains regions that are related to platelet-derived growth factor receptor signaling  
403 pathway ( $p \leq 1.8e-5$ ), ERBB signaling pathway ( $p \leq 8.5e-5$ ), epidermal growth factor receptor signaling  
404 pathway ( $p \leq 1.5e-4$ ), with a mouse phenotype of placental labyrinth hypoplasia ( $p \leq 2.6e-5$ ). Group  
405 number 4 is associated with genes involve in cell motility and cell migration ( $p \leq 8.1e-8$ ), focal adhesion  
406 ( $p \leq 1.3e-14$ ) and actin cytoskeleton ( $p \leq 3.6e-13$ ), and abnormal extraembryonic tissue morphology  
407 ( $p \leq 3.1e-10$ ).

408 Overall, OSKM and GETM factors open and activate regions that are essential for their function (e.g. Wnt  
409 signaling for OSKM and cell migration, motility and reaction to heparin and insulin for GETM) as well as  
410 regions that are important for the induction of MET and cellular transformation (e.g. focal adhesion and  
411 actin cytoskeleton).

412 Next, we subtracted all the peaks that were overlapped with MEFs to identify transcription factor  
413 binding sites that are enriched in peaks (ATAC-seq, H3K27ac and H3K4me2) associated with each  
414 reprogramming process at various reprogramming time points. (**Fig. 6A-B and Supplementary Figs. 5A-C**)

415 **and 6E-F).** We observed a highly significant P-value for binding motifs of OSK factors in OSKM  
416 reprogramming peaks and GET binding motifs in GETM reprogramming peaks, supporting our analysis  
417 **(Figs. 6A-B, Supplementary Figs. 5A-C and 6E-F).** Interestingly, the binding sites of the AP1/CREB/ATF  
418 families of proteins, which act as somatic cell identity safeguards and block the reprogramming process  
419 to pluripotency<sup>41</sup>, are significantly more enriched in GETM reprogramming peaks compared to OSKM  
420 reprogramming peaks **(Figs. 6A-B and Extended Data Fig. 5A-C).** In contrast, regions that are open in the  
421 fibroblasts and closed upon reprogramming, the binding sites of the AP1/CREB/ATF family of proteins  
422 are significantly more enriched in OSKM reprogramming compared to GETM reprogramming **(Extended**  
423 **Data Fig. 6G-H).** These results explain the potent ability of OSKM to erase somatic cell identity, as well as  
424 the continued presence of MEF-like cells in GETM reprogramming even at day 12 of the process **(Fig. 3A-**  
425 **B and Extended Data Fig. 3A).** Besides the expected Gata, Tfp2c and Eomes/Tbet motifs and the  
426 binding sites for AP1/CREB/ATF families, GETM-specific peaks were enriched with genes involved in  
427 oxidative stress response such as Nrf2<sup>54</sup>, Nfe2<sup>54</sup>, MafK<sup>54</sup> and Bach1/2<sup>55</sup> **(Fig. 6A-B, Supplementary Figs.**  
428 **5A-C and 6E).** In contrast, OSKM-specific peaks were enriched with pluripotency binding sites such as Klf,  
429 Sox, Oct and Nanog as expected, but also with genes involved in neuronal differentiation, such as E2A<sup>56</sup>,  
430 Ascl1<sup>57</sup>, and with trophoblast such as Cdx2 and Znf263 **(Figs. 6A-B and Supplementary Figs. 5A-C and**  
431 **6F),** explaining the generation of trophoblast-like cells and neuronal fate observable in OSKM  
432 reprogramming **(Figs. 2H and 3B).** In agreement with their role as reprogramming factors, GETM and  
433 OSKM shared peaks were enriched with genes that are important for remodeling the chromatin such as  
434 Ctcf<sup>58</sup>, BORIS<sup>59</sup>, E2f6<sup>60</sup>, Elf1<sup>61</sup>, Usf1/2<sup>62</sup> and YY1<sup>63</sup>.

435 We then chose 13 factors whose binding sites were significantly enriched in either GETM  
436 reprogramming (Nrf2, Nfe2, Fos, MafK, Atf3, Fosl2, Tead2) or OSKM reprogramming (Klf4, Cdx2, Pdx1) or  
437 shared between the two systems (Ctcf, Usf1, Usf2,). We performed reprogramming experiments to  
438 iPSCs and iTSCs with BYKE cells transduced with either OSKM or GETM, together with an empty vector  
439 (EV) control or with one of the 13 selected factors **(Fig 6C-F).** Strikingly, all examined factors (besides  
440 Cdx2 in GETM) either hindered the reprogramming process or had a mild effect in both systems,  
441 suggesting that both OSKM and GETM initially open regions that are highly regulated by somatic identity  
442 safeguards that counteract the reprogramming process at large **(Fig. 6C-F).**

443 Surprisingly, a very strong reprogramming inhibition was noted when Klf4 was overexpressed in both  
444 GETM and OSKM reprogramming. As Klf4 is relatively highly expressed in MEFs, it is tempting to  
445 speculate that overexpression of Klf4 on top of OSKM alters the stoichiometry of the reprogramming  
446 factors and counteracts reprogramming by maintaining fibroblastic identity **(Fig. 6C-F).** Another very

447 strong reprogramming blocker that we found, especially for iTSC reprogramming, is Usf2, which is a  
448 known tumor suppressor and Myc inhibitor<sup>64</sup>. Moreover, Usf2 is a strong regulator of iron metabolism  
449 and oxidative stress response<sup>65,66</sup>, proposing an explanation as to why Usf2 has a stronger effect on iTSC  
450 reprogramming.

451 In contrast to the global reprogramming blockers mentioned above, Ctf, which significantly hindered  
452 the reprogramming to iTSCs, only mildly affected the reprogramming to iPSCs (**Figs. 6C and 6E**). As Ctf  
453 is highly expressed in iTSCs/TSCs and acts as a very important chromatin insulator that controls gene  
454 expression<sup>67</sup>, this result emphasizes the importance of retaining normal levels of Ctf for the induction  
455 of the TSC fate. As expected, Cdx2 facilitated the reprogramming to the TSC state and hindered  
456 reprogramming to the pluripotent state.

457 We then analyzed the binding sites of closed regions; peaks that were open in MEFs and disappeared  
458 during the reprogramming process with GETM or OSKM (**Extended Data Fig. 6G-H**). Interestingly, while  
459 OSKM closed peaks are enriched with binding sites of AP1/CRE/ATF family, TEAD, Pdx1, RUNX and  
460 Mef2a, indicating the initial loss of the fibroblast identity, GETM closed peaks are enriched as well with  
461 RUNX and TEAD but also with interferon response genes such as STAT5, ISRE, RXR and IRF1/2 and  
462 apoptosis-related genes such as p53 and p63. This might suggest that GETM overcome viral infection-  
463 induced apoptosis by closing regions that control interferon response genes and master regulators of  
464 cell death.

465 Next, we sought to determine whether the regions that begin to open up via GETM and OSKM during  
466 the initial phase of reprogramming are active or not. Using scatter plots, we probed all ATAC-seq peaks  
467 from both GETM and OSKM reprogramming (**Extended Data Fig. 7A**). In accordance with the PCAs and  
468 Venn diagrams, the vast majority of the peaks were unique to each reprogramming process. We then  
469 plotted all the H3K4me2 peaks on top of the ATAC-seq peaks (**Extended Data Fig. 7B**) and performed GO  
470 analysis on OSKM or GETM-specific peaks using EnrichR (**Extended Data Fig. 7C-D**). By comparing OSKM  
471 and GETM H3K4me2 peaks, we were able to focus on all the unique regions that are remodeled by  
472 OSKM and by GETM, as any global regions that are involved in the identity of the fibroblast or are  
473 important for reprogramming at large will overlap between the two reprogramming systems.  
474 Remarkably, besides regions that are involved in the regulation of epithelial cell migration, analyzing  
475 OSKM-specific peaks revealed significant enrichment for regions important for the development of the  
476 heart (e.g. GO terms of regulation of heart contraction and regulation of cardiac conduction), the first  
477 and arguably most crucial organ to form during embryogenesis (**Extended Data Fig. 7C**). Moreover, a  
478 significant enrichment was found for the formation of the brain and liver as well (e.g. GO terms of

479 neuron projection maintenance, axon guidance and determination of liver asymmetry, **Extended Data**  
480 **Fig. 7C**). In contrast, GETM-specific H3K4me2 peaks are enriched for regions that involve metabolic  
481 processes and proliferation, as well as regions that are essential for trophoblast function such as  
482 migration and attraction of blood vessels (e.g. GO terms of regulation of blood vessel cell migration,  
483 regulation of cell migration in angiogenesis, **Extended Data Fig. 7D**).

484 We then plotted the active histone mark H3K27ac on top of the ATAC-seq peaks (**Extended Data Fig.**  
485 **7E**). Interestingly, we noted that while OSKM-specific ATAC-seq peaks tend to lose H3K27ac during the  
486 reprogramming process, GETM-specific peaks gain H3K27ac (**Extended Data Fig. 7F**). OSKM-specific  
487 H3K27ac peaks are mainly enriched within regions that play a role in neuron development and Wnt and  
488 calcium signaling pathways (**Extended Data Fig. 7G**), while GETM-specific peaks are enriched within  
489 regions that involve the regulation of MAPK activity, response to reactive oxygen species (ROS) as well  
490 as various metabolic processes (e.g. GO terms of metal ion homeostasis, Camp-mediated signaling and  
491 protein phosphorylation, **Extended Data Fig. 7H**).

492 Overall, these results agree with the transcriptome and methylation data, suggesting that OSKM initially  
493 open (ATAC-seq peaks) and define (H3K4me2 peaks) regions along the genome that are important for  
494 their differentiation potential but then eliminate their activity by removing the active mark, H3K27ac. In  
495 contrast, GETM specifically open and activate regions that are essential for trophoblast function as the  
496 process progresses, while closing and methylating regions which participate in the embryonic  
497 development program.

498

#### 499 **Data Integration Analysis Illuminates Key and Unique Aspects for the Induction of Pluripotency and** 500 **TSC State**

501 We then performed data integration analysis to correlate gene expression to chromatin accessibility and  
502 activity (**Figs. 7A-E and Extended Data Fig. 8A-K**) as well as to methylation (**Figs. 7F-I**) To that end, we  
503 initially performed cluster analysis on 18,420 GETM and OSKM-specific ATAC-seq peaks. This gave rise to  
504 14 clusters (**Fig. 7A-B**) that show unique patterns of chromatin accessibility, activity, peak distribution  
505 and DNA binding motifs in the two reprogramming systems. Then, we associated the ATAC-seq peaks of  
506 each cluster to their neighboring genes and tested whether their expression is highest in GETM  
507 reprogrammable cells, OSKM reprogrammable cells or in MEFs (depicted as a pie graph in the bottom  
508 part of each cluster). Finally, for each group of genes (i.e. highest in GETM, OSKM or MEFs) we  
509 performed GO annotation.

510 Clusters 1-4 are GETM-specific clusters, as they harbor genomic regions with a higher chromatin  
511 accessibility and activity in GETM samples compared to OSKM samples (**Figs. 7C and Extended Data Fig.**  
512 **8A-C**). While clusters 1,2,3 are enriched for binding motifs of the GETM reprogramming factors Gata3,  
513 Tfap2c and Eomes as well as Tbx6 and FOS/Atf3/AP-1, cluster 4 is enriched for binding motifs for AP-1  
514 and for the master TSC regulator, TEAD<sup>68</sup>. Moreover, while clusters 1,2,3 contain mostly intronic and  
515 intergenic ATAC-seq peaks, cluster 4 encompasses a large fraction of transcription start site (TSS) ATAC-  
516 seq peaks. GO annotation analysis revealed that clusters 1,2,3 include genes that are associated with fat  
517 differentiation ( $p \leq 0.005$ ), p38 MAPK pathway ( $p \leq 0.001$ ) and Glycogen metabolism ( $p \leq 0.0002$ ), while  
518 cluster 4 contains genes involved in apoptosis ( $p \leq 0.0007$ ).

519 In accordance with that, it has been shown that lipid droplets formation and glycogen storage are crucial  
520 for proper trophoblast function and that p38 MAPK pathway controls the invasiveness capability of  
521 trophoblastic cells<sup>69-71</sup>. Genes that are upregulated in clusters 1,2,3 to the highest level in OSKM  
522 reprogrammable cells are related to Wnt signaling ( $p \leq 0.008$ ), IGF-1 pathway ( $p \leq 0.01$ ) and cardiocyte  
523 differentiation ( $p \leq 0.0004$ ), while cluster 4 is enriched with genes involve in BMP signaling pathway  
524 ( $p \leq 0.0001$ ). Interestingly, regions that control Wnt and IGF-1 pathways which are implicated in  
525 pluripotency maintenance<sup>72</sup> are also more activated in GETM reprogramming but the associated genes  
526 are higher in OSKM reprogramming suggesting that these regions are negatively regulated by GETM. In  
527 contrast, genes that are expressed to the highest level in the parental MEFs in clusters 1,2,3 are  
528 connected to prostaglandin synthesis ( $p \leq 0.0001$ ), integrin binding ( $p \leq 0.00001$ ) and Tgf- $\beta$  signaling  
529 ( $p \leq 0.004$ ), while cluster 4 is enriched with genes that involve in focal adhesion ( $p \leq 7e-12$ ), all implicated  
530 in fibroblastic identity maintenance<sup>73,74</sup> and are negatively regulated by GETM factors.

531 Clusters 5-11 are GETM and OSKM shared clusters as they harbor genomic loci with either high or low  
532 chromatin accessibility and activity in both reprogrammable cells (**Figs. 7D and Extended Data Fig. 8D-I**).  
533 Clusters 6,8,9,10, are highly enriched with strong peaks around the TSS and as such, contain binding  
534 sites for transcription factors that are implicated in transcription initiation, such as NFY, SP1/KLF, ETS,  
535 E2A and Oct2<sup>75-77</sup>. Intriguingly, genes that are upregulated in these clusters to the highest level in GETM  
536 reprogramming are implicated in mRNA processing (e.g. mRNA catabolic process ( $Pv \leq 5.0e-9$ ), mRNA  
537 processing ( $p \leq 5.0e-8$ ), spliceosome ( $p \leq 9.1.0e-10$ )), while genes that exhibit the highest levels in OSKM  
538 reprogramming involve mitotic cell cycle regulation ( $p \leq 4.4e-11$ ), estrogen signaling ( $p \leq 0.00002$ ) and  
539 insulin signaling ( $p \leq 0.0001$ ). These processes are general and characterize both states, however, OSKM  
540 reprogramming also upregulated genes implicated in terpenoid backbone biosynthesis ( $p \leq 0.0002$ ) which  
541 is involved in the conversion of the pluripotency primed state to naïve state<sup>78</sup>. Genes with the highest



542 expression levels in MEFs participate in protein processing in ER ( $p \leq 0.00003$ ), EGFR1 signaling pathway  
543 ( $p \leq 0.000007$ ), cyclic nucleotide catabolic process ( $p \leq 0.00007$ ) and cell junction assembly ( $p \leq 0.000002$ ).  
544 In contrast, clusters 5,7,11, which are also shared between OSKM and GETM reprogramming but  
545 enriched with smaller peaks that are located around the TSS as well as in intergenic and intronic regions,  
546 contain binding sites for transcription factors that are key drivers of the fibroblastic identity (Zeb1,  
547 Tcf12, Tbx5 and Sox6) and more significantly, with binding sites for the insulator gene Ctf. In agreement  
548 with that, GO analysis for genes that are highest expressed in MEFs in these clusters exhibited gene  
549 ontologies of extracellular matrix organization ( $p \leq 1.6e-10$ ) and integrin binding ( $p \leq 0.00001$ ). In contrast,  
550 genes that are expressed to the highest levels in GETM reprogrammable cells are involved in RNA  
551 processing ( $p \leq 0.0001$ ), translation ( $p \leq 7.4e-9$ ) and regulation of cytokines biosynthesis ( $p \leq 0.0001$ ), while  
552 genes that are highest in OSKM reprogrammable cells are enriched for axonal transport ( $p \leq 0.0002$ ),  
553 negative regulation on bone remodeling ( $p \leq 0.00007$ ) and ubiquitin conjugation binding ( $p \leq 0.000008$ ),  
554 suggesting the initial opening of the different lineages.

555 Clusters 12,13,14 are OSKM-specific with open and active peaks in OSKM reprogrammable cells that are  
556 enriched for OSK reprogramming factor binding sites, Oct, Sox, Klf as well as for Nanog (**Figs. 7E and**  
557 **Extended Data Fig. 8J-K**). Similar to GETM-specific clusters, these OSKM-specific clusters are also  
558 enriched with peaks that are mostly localized to intronic and intergenic regions. Genes that are  
559 associated with these peaks and harbor the highest expression level in OSKM reprogrammable cells are  
560 involved in Rap1 signaling pathways ( $p \leq 0.0001$ ), positive regulation of the non-canonical Wnt signaling  
561 pathway ( $p \leq 0.00009$ ) and neuron development ( $p \leq 0.00004$ ), all implicated in neuronal cell activity<sup>79,80</sup>,  
562 further explaining why OSKM reprogramming can induce neuronal fate. Genes that are expressed to the  
563 highest levels in GETM reprogrammable cells play a role in ERBB signaling pathway ( $p \leq 0.000002$ ),  
564 glycogen metabolism ( $p \leq 0.006$ ) and mRNA surveillance pathway ( $p \leq 0.04$ ), processes which are  
565 important for trophoblast differentiation. Finally, genes that are associated with these peaks but exhibit  
566 the highest levels in MEFs are implicated in regulation of actin cytoskeleton ( $p \leq 0.001$ ) and regulation of  
567 cell migration ( $p \leq 3.4e-10$ ), all important for normal fibroblastic function.

568 In conclusion, these data describe how GETM reprogramming differs from OSKM reprogramming in the  
569 induction and dynamics of various signaling pathways, metabolomic processes and well as in their ability  
570 to erase the somatic identity and to induce alternative cell fates during reprogramming. Moreover, it  
571 demonstrates that from the onset of the reprogramming process, GETM reprogramming is directed  
572 toward the TSC fate and activates various processes and pathways that are essential for TSC  
573 maintenance and differentiation.

574

575 We then sought to correlate gene expression to methylation. We focused on one interesting cluster in  
576 which demethylation occurs in most regions of both systems only at the final step of the reprogramming  
577 process (**Fig. 7F**). We associated the 3,992 tiles of this cluster to their neighboring genes and examined  
578 their expression in ESCs and TSCs (**Fig. 7G**). We identified 525 genes that are upregulated and 453 genes  
579 that are downregulated specifically in TSCs when compared to MEFs. 770 genes are upregulated and  
580 772 genes are downregulated specifically in ESCs, while 300 genes are upregulated and 317 genes are  
581 downregulated in both ESCs and TSCs when compared to MEFs.

582 GO annotation analysis revealed that genes that are upregulated specifically in TSCs play a role in EGFR1  
583 signaling, insulin signaling and fat differentiation and contain transcription factor binding sites of GATA2,  
584 SUZ12 and TP63, all important players in trophoblast formation and differentiation. Indeed, Jensen  
585 tissue analysis associated these genes with the placenta with the most significant P-value ( $p \leq 6.5e-5$ , **Fig**  
586 **7H**).

587 Analysis of the genes that are specifically upregulated in ESCs identified pluripotency and notch signaling  
588 pathways as the most significant biological processes, with transcription factor binding sites that are  
589 enriched for many pluripotency genes such as Tcf3, Rest, Stat3, Sox2, Klf4, Nanog and Oct4.  
590 Interestingly, Jensen tissue analysis associated these genes with Cerebral cortex prefrontal, cerebellum  
591 as well as ESC lines (**Fig. 7H**). Genes that are upregulated in both ESCs and TSCs are involved in stemness  
592 at large, as pluripotency network and DNA replication were identified as the most significant biological  
593 processes and ESC lines and placenta the most significant tissues. In accordance with that, E2F4 and to a  
594 lesser extent other E2F family members, which play a key role in stem cell proliferation<sup>81</sup>, are found to  
595 be the most significantly enriched binding sites in these tiles, in addition to binding motifs of  
596 pluripotency genes such as Nanog, Tcf3, Klf4, Sall4 and Oct4 (**Fig. 7H**). When we analyzed the genes that  
597 are downregulated either in TSCs, ESCs or both, they were all implicated in fibroblastic identity and  
598 function, ranking osteoblasts, MEFs and macrophages as the most significant cell types (**Fig. 7I**). In  
599 accordance with the downregulation of these genes, their associated tiles were enriched with binding  
600 motifs of factors known to induce strong transcriptional repression such Klf4, AR, Zbtb7a, UBTF, Suz12  
601 and Nfe2l2, suggesting how these hypomethylated regions are associated with gene repression.

602 Taken together, this analysis allowed us to examine the final stage of demethylation that occurs in both  
603 GETM and OSKM reprogramming, which characterizes the stabilization stage of both cell types.  
604 Surprisingly, this final stabilization step still involves the erasure of the fibroblastic identity.

605



606 **Genomic Stability Analysis during OSKM and GETM Reprogramming Reveals Equivalent Frequency of**  
607 **Copy Number Variations (CNVs) during the Initial Phase of Reprogramming**

608 One of the characteristics of TSCs is a unique methylation landscape that allows the activation of many  
609 repetitive elements within the trophoblast genome<sup>82</sup>. This unique property is believed to induce  
610 genomic instability<sup>83,84</sup> and indeed, multiple genomic aberrations are found in both iTSCs and bdTSCs  
611 following prolonged culture<sup>13</sup>. Therefore, we next aimed to understand whether GETM activation  
612 induces genomic instability already at the onset of the reprogramming process toward the TSC state.  
613 Since Myc is a known driver of genomic instability<sup>85</sup>, we reprogrammed fibroblasts into iTSCs by GET or  
614 GETM and to iPSCs by OSK or OSKM as a control. Cells were collected for copy number variation (CNV)  
615 analysis immediately following the infection (day 0), at day 3 of reprogramming and at day 6 of  
616 reprogramming. In addition, we examined ten iPSC clones and two previously characterized partially  
617 reprogrammed cells<sup>23</sup> as reference cells. All genomic reads were aligned against the parental MEF  
618 genome. As can be seen in Figure S7L, while many CNVs were identified in one of the two partially  
619 reprogrammed iPSC clones and few CNVs in chromosome 1 or 8 in three out of ten fully reprogrammed  
620 iPSC clones, this analysis could not identify a significant amount of CNVs in the initial phase of both  
621 OSK/M and GET/M reprogramming (**Extended Data Fig. 8L**).

622 As the sensitivity of bulk whole-genome sequencing is limited and does not allow the detection of CNVs  
623 at single cell resolution, further examination is needed to fully address this question. However, we can  
624 confidently conclude that substantial genomic instability is not induced by GETM at the initial phase of  
625 reprogramming.

626

627 **Discussion**

628 Nuclear reprogramming by defined factors is a powerful tool in understanding cellular plasticity and cell  
629 fate decision<sup>86</sup> and for the generation of various cell types from somatic cells. Since the discovery of  
630 iPSCs by Takahashi and Yamanaka in 2006<sup>17</sup>, the reprogramming process of fibroblasts to iPSCs by the  
631 OSKM factors has been investigated extensively by many groups<sup>22-37</sup>. In contrast, the reprogramming  
632 process of fibroblasts to iTSCs by GETM, described for the first time in 2015<sup>13,16</sup>, has never been done  
633 before.

634 Considering the fact that pluripotency and trophectoderm fates arise simultaneously during blastocyst  
635 development, we hypothesized that performing a parallel and comparative multi-omics analysis on both  
636 reprogramming processes concomitantly will yield knowledge that cannot be revealed otherwise, such  
637 as when each reprogramming process is analyzed separately.

638 To that end, we reprogrammed MEFs to iPSCs by OSKM and to iTSCs by GETM and examined their  
639 transcriptome (i.e. Bulk RNA-seq and SC-RNA-seq), methylome (i.e. RRBS), chromatin accessibility and  
640 activity (i.e. ATAC-seq and CHIP-seq for H3K4me2 and H3K27ac) and genomic stability (i.e. CNVs) at  
641 various time points along the process.

642 Initially, we asked whether the reprogramming process toward pluripotent and TSC states follows the  
643 same dynamics as of the forming cells during early embryogenesis. While clear transcriptional changes  
644 are found between the different stages (i.e. zygote, 2-cell stage, 4-cell stage, 8-cell stage, morula and  
645 blastocyst), the transcriptional heterogeneity within the cells of each group before blastocyst formation  
646 is relatively mild. This suggests a 'T'-shaped model, where cells at each stage undergo relatively similar  
647 epigenetic and transcriptional changes before segregation, and dispersed into two distinct cells types,  
648 the ICM and TE, only at the morula/early blastocyst stage.

649 Our comparative and parallel multi-layer analysis revealed that, in contrast to cells during early  
650 embryogenesis that mostly resemble each other in each stage, cells undergoing reprogramming to  
651 pluripotent and TSC states exhibit unique and specific trajectories from the beginning of the process till  
652 the end, suggesting 'V'-like behavior. Although similar processes such as somatic identity loss,  
653 proliferation, MET and metabolic shift occur in the two systems, each of the processes mostly uses  
654 different sets of genes and regulatory elements to induce its own fate. This 'V'-shaped behavior was  
655 observed at all levels, starting from transcription and chromatin accessibility and activity and ending  
656 with DNA methylation.

657 We show that each reprogramming process uses different genomic regions and various strategies to  
658 silence fibroblastic identity. While the OSKM combination is very potent in inducing identity loss by  
659 interacting, from the onset of the process, with key regions that safeguard fibroblastic identity (i.e.  
660 regions that are enriched with ATF/CREB/AP1 sites), GETM open regions that are enriched with  
661 ATF/CREB/AP1 binding sites, which counteract their ability to silence the fibroblastic identity. This is in  
662 agreement with the SC-RNA-seq data that demonstrate a big fraction of cells with MEF-like identity even  
663 at day 12 of the reprogramming process by GETM.

664 By exploiting single-cell analysis, we demonstrate two unique and distinct populations of  
665 reprogrammable cells, suggesting that neither of the reprogrammable cells harbor a transcriptional  
666 profile that is shared during GETM and OSKM reprogramming. Moreover, we could also illuminate  
667 previously unknown reprogramming stages and markers for a faithful (Tdgf1 for OSKM and Cd82 for  
668 GETM) and failed (Anx3 for OSKM) reprogramming process for the two reprogramming systems.

669 These results clearly demonstrate that the reprogramming process of somatic cells toward pluripotency  
670 and TSC state takes completely different routes from the onset of reprogramming, and that somatic  
671 nuclear reprogramming and reprogramming during early embryonic development toward pluripotency  
672 and TE state are characterized by different properties and follow diverse paths.

673 However, we believe that key features that characterize the process of nuclear reprogramming by OSKM  
674 and GETM are shared with the reprogramming process that occurs before lineage specification in the  
675 early embryo. As such, by comparing OSKM to GETM reprogramming we could subtract all the general  
676 regions that are being remodeled in the fibroblastic nucleus during reprogramming at large and to focus  
677 on regions that are being remodeled specifically by OSKM factors. Remarkably, we identified a clear and  
678 significant embryonic development program that is executed by OSKM and involves chromatin  
679 remodeling of two of the most important organs of the developing embryo, the brain and the heart.  
680 Initially, OSKM demethylate, define and open these regions and subsequently limits their activity by  
681 decreasing the levels of the active histone mark, H3K27ac. In contrast, GETM factors induce DNA  
682 methylation on key developmental genes and thus shutting off this early embryonic development  
683 program. By inducing chromatin accessibility and activity and by increasing the levels of the active  
684 histone mark H3K27ac, GETM activate the trophoblastic program that involves the activation of  
685 metabolic processes that participate in transcription and translation, as well as migration and  
686 endothelial cell attraction, which are all known properties of trophoblast cells.

687 Overall, this study describes and illuminates key features that characterize the reprogramming process  
688 toward pluripotent and TSC states at all levels of regulation (i.e. DNA methylation, chromatin  
689 accessibility and activity, transcriptome and CNVs). By comparing two reprogramming processes  
690 simultaneously we were able to reveal new properties for the induction of pluripotent and TSC states,  
691 elements which could not have been elucidated had each reprogramming process been analyzed  
692 separately. We believe that the generation of such a database is a powerful tool to study cellular  
693 plasticity and cell fate decision.

694

#### 695 **Acknowledgment**

696 Y.B. is supported by research grants from the European Research Council (ERC, #676843), the Israel  
697 Science Foundation (ISF, #823/14), EMBO Young Investigator Programme (YIP), DKFZ-MOST (CA 177),  
698 Howard Hughes Medical Institute International Research Scholar (HHMI, #55008727) and by a generous  
699 gift from Ms. Nadia Guth Biasini. T.K. is supported by the Israel Science Foundation (ISF, #913/15 and

700 #1250/18). A.R. is supported by the British GROWTH fellowship. N.L. is supported by CIDR Data Science  
701 and Leibniz fellowships

702

### 703 **Author Contribution**

704 Y.B. conceived the study, prepared the figures and wrote the manuscript. Y.B. and M.J. designed the  
705 experiments. M.J. performed all the reprogramming experiments toward iTSCs and iPSCs, collected and  
706 prepared cells for all analyses and performed ChIP-seq, RRBS and cloned the various factors for  
707 reprogramming efficiency experiments. A.R. performed the bioinformatics analyses including bulk RNA,  
708 SC-RNA-seq, RRBS (together with N.L. and T.K.), ATAC-seq and ChIP-seq (together with N.L. and T.K.).  
709 M.A. performed the reprogramming efficiency experiments with the various identified factors and  
710 scored and counted the colonies. S.S. prepared all the libraries for the RNA-seq and ATAC-seq and  
711 performed the reprogramming experiments for CNV analysis. T.K., M.Z. and A.E. performed and  
712 analyzed the CNV experiments on reprogrammable cells and fully reprogrammed iPSC colonies. T.K.  
713 supervised A.R. and N.L. and contributed to the design of the analyses and performed data integration  
714 analysis (together with A.R.). Y.B., A.R. and T.K. contributed to the interpretation of the results. K.M.  
715 injected BYKE cells to blastocysts for the purpose of MEF isolation.

716

### 717 **Author Information**

718 All ATAC-seq, ChIP-seq, RRBS, SC-RNA-seq and bulk RNA-seq data is currently being deposited in the  
719 Gene Expression Omnibus database (GEO). We will be happy to share the data with the reviewers.  
720 Please let us know if you request it for the review process. Correspondence and requests for materials  
721 should be addressed to Y.B. ([yossibug@ekmd.huji.ac.il](mailto:yossibug@ekmd.huji.ac.il)).

722

### 723 **Methods**

#### 724 **Cell culture and primary MEFs production.**

725 All ESCs and iPSCs were cultured in mouse embryonic stem cell medium containing 500ml DMEM  
726 supplemented with (15% FBS, 2mM L-Glutamine, 1% non-essential amino acid, in-house mouse  
727 Leukemia inhibitory factor (mLif), 0.1mM  $\beta$ -mercaptoethanol (Sigma), 1% penicillin-streptomycin and  
728 with or without 2i- PD0325901 (1  $\mu$ M) and CHIR99021 (3  $\mu$ M) (PeproTech). All TSCs and iTSCs were  
729 grown in TSC medium containing a combination of 70% MEF conditioned medium and 30% freshly  
730 prepared medium, (RPMI supplemented with 20% FBS, 0.1 mM  $\beta$ -mercaptoethanol, 2 mM L-Glutamine,

731 1% penicillin-streptomycin, 25 ng/ml human recombinant FGF4 (PeproTech) and 1 µg/ml heparin  
732 (Sigma-Aldrich).

733 Mouse embryonic fibroblasts (MEFs) were isolated as previously described<sup>87</sup>. Briefly, Pluripotent mouse  
734 ESCs and iPSCs were injected into E3.5 blastocysts, chimeric embryos were isolated at E13.5 and then  
735 dissected under the binocular to remove any internal organs and heads. The tissue was chopped by  
736 scalpels and incubated 30 minutes with 1ml Trypsin-EDTA (0.25%, Gibco) at 37°C. Next, trypsin  
737 activation was neutralized by 10ml DMEM containing 10% serum and the chopped embryos underwent  
738 intensive pipetting until homogeneous mixture of cells was noted. Each embryo was seeded into one  
739 15cm plate and cultured with DMEM containing 10%FBS, 1% penicillin-streptomycin and 2mM L-  
740 glutamine. The cells were grown till the plate being full. Puromycin (2µg/ml) was added for selection for  
741 BYKE MEFs (the M2rtTA cassette that resides inside the rosa26 locus of the injected cells contains a  
742 resistance gene for puromycin), killing only the host cells. All cells were maintained in a humidified  
743 incubator at 37°C and 6% CO<sub>2</sub>.

744 **Molecular Cloning, Lentiviral Infection, and Reprogramming.**

745 The open reading frame of the examined genes (i.e. Ctcf<sup>23</sup>, Cdx2<sup>13</sup>, Atf3<sup>88</sup>, Tead2, Fosl2, Pdx1, Nrf2, Usf1,  
746 Usf2, NE-F2, Fos, Mafk) was cloned into pMINI vector (NEB) and then restricted with EcoRI or MfeI and  
747 transferred into FUW-TetO expression vector. Lentiviruses were generated by transfecting vector DNA,  
748 (hGETM 3:3:3:1) or STEMCCA cassette for hOSKM, with a mix of lentiviral packaging vectors (7.5 µg  
749 psPAX2 and 2.5 µg pGDM.2) into 293T cells, the viruses were collected at 48, 60 and 72h after  
750 transfection, the medium containing the viruses was supplemented with 8 µg/ml of polybrene (Sigma)  
751 and filtered by 0.45 µm filter, the viruses were then added to MEFs (passage 0) that were seeded at 70%  
752 confluency two days prior to the first infection. Six hours following the third infection, medium was  
753 changed into DMEM containing 10% FBS. Eighteen hours later, medium was changed into  
754 reprogramming medium; ESC reprogramming medium (DMEM supplemented with 10%FBS, 0.1mM β-  
755 mercaptoethanol, 2mM L-glutamine, 1%non-essential amino acids, in-house mouse Leukemia inhibitory  
756 factor (mLif), and 2 µg/ml doxycycline) or TSC reprogramming medium (RPMI supplemented with 20%FBS,  
757 0.1mM β-mercaptoethanol, 2mM L-glutamine, in house mouse recombinant FGF4 (equivalent to  
758 25ng/ml), 1 µg/ml heparin (Sigma-Aldrich), and 2µg/ml doxycycline). The two reprogramming mediums  
759 were changed every other day. For iPSC reprogramming, the MEFs were exposed to doxycycline for 15  
760 days, followed by 5 days of dox withdrawal in ESC culturing medium. For iTSC reprogramming, the MEFs  
761 were exposed to doxycycline for 20 days, followed by 10 days of dox removal in TSC culturing medium.  
762 iTSCs colonies were then isolated, trypsinized, and plated in a well in a 6-well plate on feeder cells and  
763 passaged until stable colonies emerged.

764

765 **FACS analysis.**

766 Cells were trypsinized, washed with PBSx1 and filtered through mesh paper. Flow cytometry analysis  
767 was performed on a Beckman Coulter and cell sorting was performed on FACS-Aria III.

768

769 **RNA libraries and sequencing.**

770 Total RNA was isolated using the Qiagen RNeasy kit. mRNA libraries were prepared using the SENSE  
771 mRNA-seq library prep kit V2 (Lexogen), and pooled libraries were sequenced on an Illumina NextSeq  
772 500 platform to generate 75-bp single-end reads.

773 **Reduced Representation Bisulfite sequencing (RRBS):**

774 RRBS assay was performed as previously described<sup>89</sup>, briefly, 20ng of genomic DNA were digested with  
775 Msp1 restriction enzyme (NEB, R0106L), DNA fragments were end-repaired and A-Tailed using Klenow  
776 fragment (3'-5'-exo-) (NEB, M0212L), the DNA fragments were ligated to illumina adaptors (Illumina, PE-  
777 940-2001) using T4 ligase (NEB, M0202M) and then size selected using AMPure XP beads (Beckman  
778 Coulter Genomics, A63881) The samples were then subjected to two consecutive bisulfite conversions  
779 using EpiTect Bisulfite Kit (QIAGEN, 59104) and PCR using PfuTurbo Cx hotstart DNA polymerase (Agilent  
780 Technologies, 600412). The RRBS libraries were sequenced by Illumina HiSeq 2000 platform.

781

782 **Single-cell RNA seq:**

783 Reprogrammable cells at day 6 or 12 were prepared as instructed in the 10X Genomics cell preparation  
784 guidelines. Briefly, cells were trypsinized and centrifuged at 1000 RPM for 3 minutes, then were washed  
785 twice with PBSx1 containing 0.04% BSA and cleaned from cell debris and large clumps by filtering through  
786 mesh paper. Next, resuspended cells were subjected to dead cell removal kit (MACS, 130-090-101) to  
787 remove any non-viable cells. Cell viability were estimated using trypan blue staining. Cells were then  
788 resuspended in PBSx1 with 0.04% BSA at the concentration 1000 cells/ $\mu$ l and 4000 cells from each  
789 condition were subjected to 10x Genomics. Single-cell RNA libraries were prepared using Chromium  
790 Single Cell 3' Library Kit v2 (10X genomics, 120234) and the generated libraries were sequenced using  
791 Illumina NextSeq 500 platform.

792

793 **Chromatin immunoprecipitation (ChIP).**

794 Chromatin immunoprecipitation (ChIP) assay was performed as previously described<sup>90</sup>. Briefly, cells  
795 were fixed for 10 min at RT with a final concentration of 0.8% formaldehyde. Formaldehyde was  
796 quenched with glycine for a final concentration of 125mM. The cells were then lysate with lysis buffer  
797 (100mM Tris-HCl, 300mM NaCl, 2% Triton<sup>®</sup> X-100, 0.2%v sodium deoxycholate, 10mM CaCl<sub>2</sub>)  
798 supplemented with EDTA free protease inhibitor Roche- 11873580001 for 20 min at Ice and the  
799 chromatin was digested by MNase (micrococcal nuclease)- Thermo Scientific™- 88216 for 20 min at  
800 37°C. MNase was inactivated by 20mM EGTA. The fragmented chromatin was incubated with pre-  
801 bounded Dynabeads (A and G mix) - Invitrogen 10004D/ 10002D using H3K27ac antibody (Abcam,  
802 ab4729) and H3K4me2 antibody (Millipore, 07-030). Samples were then washed twice with RIPA buffer,  
803 twice with RIPA high salt buffer (NaCl 360mM), twice with LiCl wash buffer (10mM Tris-Hcl, 250mM LiCl,  
804 0.5% DOC, 1mM EDTA, 0.5% IGEPAL), twice with 10mM Tris-HCl pH=8. DNA was purified by incubating



805 the samples with RNase A (Thermo Scientific™ EN0531) for 30 min at 37°C followed by a 2 hours  
806 incubation with Proteinase K (Invitrogen™ 25530049). DNA was eluted by adding 2X concentrated  
807 elution buffer (10mM Tris-HCl, 300mM NaCl, 1% SDS, 2mM EDTA) and then reverse crosslinked  
808 overnight at 65°C. Finally, DNA was extracted using AMPure XP beads (Beckman Coulter Genomics,  
809 A63881). Chip sample libraries were prepared according to Illumina Genomic DNA protocol as  
810 described<sup>91</sup>.

811

### 812 **ATAC libraries and sequencing.**

813 ATAC-seq library preparation was performed as previously described<sup>92</sup>. Briefly, cells were trypsinized  
814 and 50,000 cells were counted and incubated in lysis buffer to isolate nuclei. Nuclei were then  
815 resuspended in transposase reaction mix for 30 min at 37 °C (Illumina, Fc-121-1030). The samples were  
816 purified using Qiagen MiniElute kit (QIAGEN, 28204), Transposed fragments were directly PCR amplified  
817 and sequenced on an Illumina NextSeq 500 platform to generate 2 × 36-bp paired-end reads.

818

### 819 **Data processing:**

#### 820 **A) Bulk RNA-seq**

821 Low quality bases and sequencing adaptors of 36 raw fastq files RNA-seq containing single-end 61bp-  
822 long reads were trimmed using Trim Galore (V 0.6.0, <https://github.com/FelixKrueger/TrimGalore>) and  
823 then mapped to the mm9 reference genome using HISAT2 (V 2.1.0, <sup>93</sup>) with default parameters. Read  
824 counting was performed using featureCounts (V 1.6.2, <sup>94</sup>) with (Mus\_musculus.NCBI37.gtf annotation).  
825 Differential gene expression analysis was performed using DESeq2\_1.26.0 package<sup>94</sup>. Unsupervised  
826 hierarchical clustering was performed for 10,000 most variable genes among ESCs, bdTSCs, fibroblasts  
827 and cells during reprogramming. R package dynamicTreeCut <sup>95</sup> was used to perform adaptive branch  
828 pruning detecting 27 prominent clusters. R packages Enrichr (V 2.1, <sup>96</sup>) and ClusterProfiler (V 3.14.3, <sup>97</sup>)  
829 were used to query Biological processes, Mouse gene atlas and KEGG pathways analysis of significantly  
830 over-represented genes for each cluster. A second aligner TopHat [4] (v2.0.6, <sup>98</sup>) was used to map reads  
831 to mm9 reference genome. Mapped reads were then processed using cufflinks [4] (v2.0.2, <sup>99</sup>), and gene  
832 expression levels (FPKM) were calculated for each replicate.

833

#### 834 **B) 10x Single-cell data**

835 SC-RNA-seq libraries were generated from each time point using the 10X Genomics. The cellranger-  
836 3.0.2, <https://github.com/10XGenomics/cellranger> was used for mapping of the 10x single-cell RNA-seq



837 data. Read1 data of pooled cells were split into single-cell data using the barcode sequences contained  
838 in the first 16 bps. The next 10 bps were recorded as unique molecular identifiers (UMIs). Read2 with 75  
839 bp were aligned to the mm10 reference genome. We used Seurat (V 3.1.4, <sup>100</sup>) to pre-processing the  
840 data and perform clustering. The function 'FindAllMarkers' to identify the marker genes for each of the  
841 clusters in the UMAP representation. For day 6 OSKM and GETM reprogramming, we excluded cells with  
842 detected genes less than 2000 or cells with sum of the non-normalized UMI counts less than 10,000, or  
843 cells with percentage of mitochondrion UMI values larger than 10%. For day 12 OSKM and GETM  
844 reprogramming, we excluded cells with detected genes less than 1000 or cells with sum of the non-  
845 normalized UMI counts less than 5000, or cells with percentage of mitochondrion UMI values larger  
846 than 15%. R package DoubletFinder (<https://github.com/chris-mcginnis-ucsf/DoubletFinder>) was used to  
847 identify and exclude potential doublets. Overall, 5899 cells of day 6 OSKM-GETM reprogramming and  
848 5752 cells of day 12 OSKM-GETM passed the quality control criteria. On median, there were  
849 16232/13488 UMI counts and 4079/3735 detected genes for each cell of day6/day12 OSKM-GETM  
850 dataset.

851

## 852 **DNA methylation**

853 Low quality bases and sequencing adaptors of 45 raw fastq files were trimmed using Trim galore (V  
854 0.6.0, <https://github.com/FelixKrueger/TrimGalore>) and then mapped to the mm9 reference genome  
855 using Bsmmap (V 2.90, <sup>100</sup>) with flags -S 10 -R -p 8 -D C-CGG. Bam files belonging to same reprogramming  
856 system and day were merged to ensure maximum overlap between all samples. Methylation beta values  
857 were extracted from the BAM files using wgbs\_tools ([https://github.com/nloyfer/wgbs\\_tools](https://github.com/nloyfer/wgbs_tools)).  
858 Methylation markers were identified using in-house developed script find\_markers.py to generate Bed  
859 files with p-value < 0.05 between different conditions summarized in different groups. 130,000 blocks  
860 were identified with significant methylation alteration that occurs during reprogramming in both OSKM  
861 and GETM reprogramming. In order to minimize noise and extract significant trends, we used the K-  
862 means algorithm to classify ~130,000 blocks that are shared amongst all samples during reprogramming  
863 to a TSC or pluripotent states and obtained 100 clusters. A new table was constructed by averaging DNA  
864 methylation levels per sample per cluster and then projected the processed data onto the first two  
865 principal components. Clusters loading plot showed significant clusters contributed to the first two  
866 principal components and clusters that are near to each other showed similar trends of methylation  
867 allowing us to extract 15 different trends shown as heatmaps in Fig. 4A and Fig. S4A. Genomic regions

868 associated with all blocks belonging to each of the 15 clusters were annotated using GREAT (V 4.0.4<sup>48</sup>)  
869 and were summarized in Supplementary Table 2.

## 870 **ATAC-seq**

871 Fastq files were mapped to the mm9 reference using bwa (<https://arxiv.org/abs/1303.3997>, version  
872 0.7.17-r1188). The mapped reads were converted to BAM format and filtered by mapping quality  
873 (MAPQ) of  $\geq 10$ , retaining only properly aligned pairs (samtools -F 1796 flag). The BAM files were then  
874 sorted and indexed using samtools (v1.9<sup>101</sup>).

875 Bigwig coverage tracks were generated using deepTools bamCoverage (v3.4.1<sup>102</sup>) with the following  
876 flags: --normalizeUsing RPGC -bs 50 -e 500 --effectiveGenomeSize 2150570000.

877 Coverage peaks were called using MACS (v2.1.2<sup>103</sup>) with flags -g mm --slocal=2000 --llocal=20000 --  
878 nomodel --extsize=300 -f BAMPE.

879 Peaks of multiple replicates were retained only if identified by 30% of the replicates, or more.

880

## 881 **ChIP-seq**

882 Fastq, BAM and bigwig files were processed in a similar way to the ATAC-seq files.

## 883 **Annotation of genomic regions**

884 Peaks from each experiment were then divided into subsets, including peaks that appear in both OSKM  
885 and GETM (3, 6, or 9 days after induction) but not in MEFs, peaks from GETM (days 3,6,9) not  
886 identifiable in MEFs, OSKM peaks (days 3, 6, 9) not identifiable in MEFs, and disjoint sets of cell-type  
887 specific peaks (e.g. GETM day 3 peaks not found in MEFs or in OSKM day 3, etc.). We also analyzed peaks  
888 from ESC, TSC or MEF cells.

889 Genomic regions from each group of peaks were then annotated using annotatePeaks.pl (HOMER suite,  
890 <http://homer.ucsd.edu/homer/ngs/annotation.html>, UCSC mm9 genome version) as Promoter, TTSs, 5'  
891 and 3' UTRs, or as Exonic, Intronic, or Intergenic regions.

892

## 893 **Motif analysis**

894 ATAC-seq peaks were called in each replicate separately, and overlapping peaks from replicates were  
895 then merged. Peaks overlapping MEF peaks (top 50K) were then removed. Finally, the center 250bp of  
896 each peak was considered for further analysis (peaks shorter than 250bp were removed).

897 We then further divided the peaks of each time point into disjoint groups, including peaks identified in  
898 both GETM and OSKM ATAC-seq (e.g. GETM&OSKM D03), GETM-only peaks (e.g. GETM\OSKM D03) or  
899 OSKM-only peaks (e.g. OSKM\GETM D03).

900 Finally, we used “findMotifsGenome.pl -nomotif” (HOMER suite) to identify occurrences of known  
901 motifs within those sequences. A similar approach was applied to H3K27ac and H3K4me2 ChIP-seq  
902 peaks.

### 903 **CNV analysis**

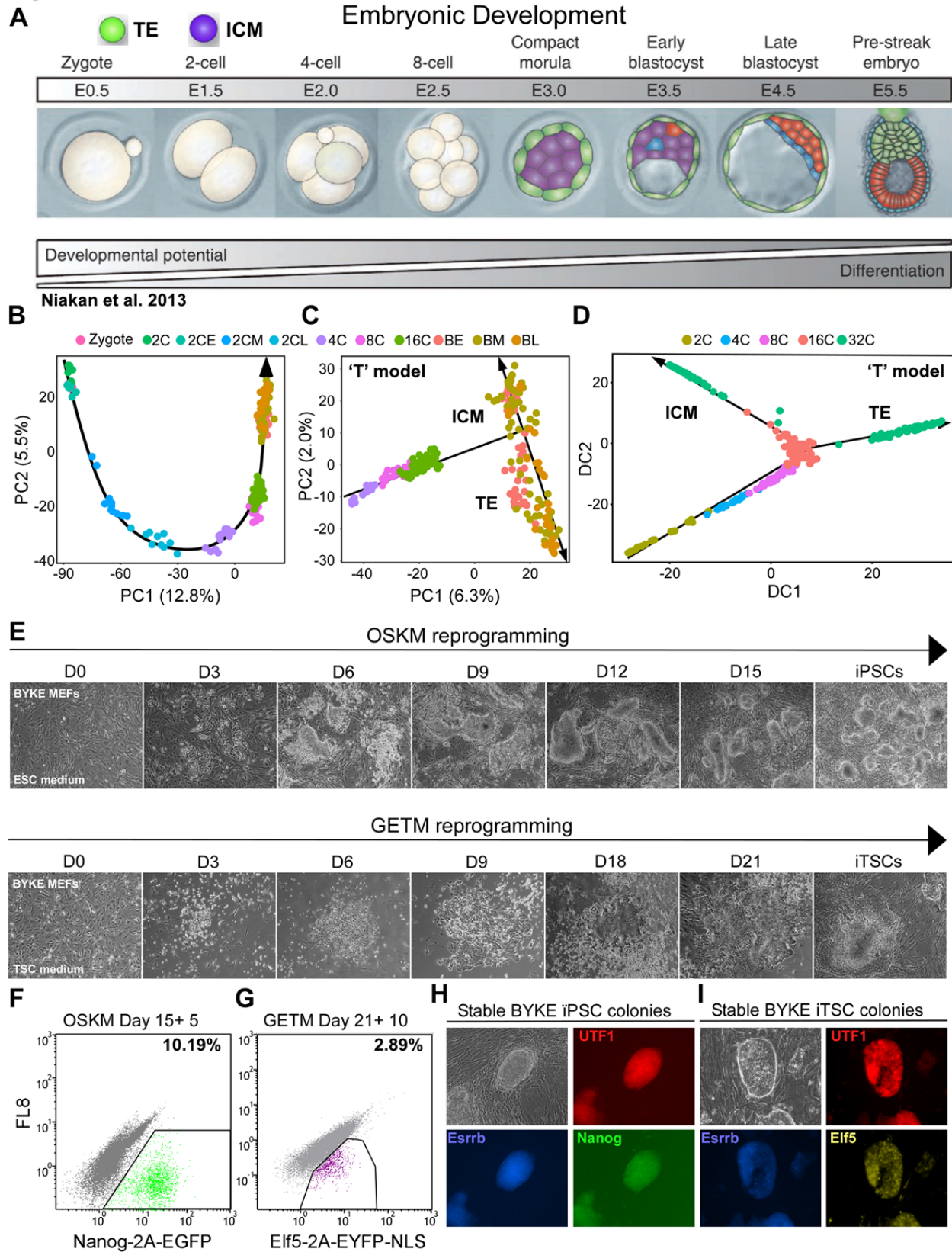
904 Read alignment was done with BWA mem 0.7.15<sup>104</sup> to the mouse reference genome including PhiX174.  
905 Copy number analysis was performed using library cn.mops<sup>105</sup> in paired mode for 4kb windows and  
906 applying DNACopy for segmentation with R (R Core Team (2015). R: A language and environment for  
907 statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [https://www.R-](https://www.R-project.org/)  
908 [project.org/](https://www.R-project.org/)).

909 **Table 1. List of primers**

<b>Gene</b>	<b>Primers</b>
Mafk <sup>cDNA</sup>	F- 5' CCGGGTTATGACGACTAATCC 3' R- 5' GAGCCTGGGATAGGCATGAG 3'
Tead2 <sup>cDNA</sup>	F- 5' GGAATCGGGATCCTGCTTGG 3' R- 5' CCGGTTCTTTCTAAGAGGAG 3'
Fosl2 <sup>cDNA</sup>	F- 5' AAACCACCCTGTTTCCTCTC 3' R- 5' ACCAGTGTCTCACCCTAAG 3'
Nrf2 <sup>cDNA</sup>	F- 5' CAGTTGCCACCCAGGATGTC 3' R- 5' GGGTTTACTCGTCAGTAGTG 3'
Fos <sup>cDNA</sup>	F- 5' AGCTCCCACCAGTGTCTACC 3' R- 5' TTGCCTTCTCTGACTGCTCAC 3'
Pdx1 <sup>cDNA</sup>	F- 5' ATGAACAGTGAGGAGCAGT 3' R- 5' TCACCGGGGTTCTGCGGT 3'
Nfe2 <sup>cDNA</sup>	F- 5' GGCTTTTCTGCTGGCACAGTAG 3' R- 5' GGCTTTGAGGGAGTCTCTAGC 3'
Usf1 <sup>cDNA</sup>	106
Usf2 <sup>cDNA</sup>	106
Atf3 <sup>cDNA</sup>	107
hGATA3 <sup>cDNA</sup>	F- 5' ATGGAGGTGACGGCGGACCAG 3' R- 5' CTAACCCATGGCGGTGACCATGC 3'
hTFAP2C <sup>cDNA</sup>	F- 5' ATGTTGTGGAAAATAACCGATA 3' R- 5' TTATTTCTGTGTTTCTCCATT 3'
hEOMES <sup>cDNA</sup>	F- 5' ATGCAGTTAGGGGAGCAGCTCTTG 3' R- 5' TTAGGGAGTTGTGTA AAAAGC 3'

910

Figure 1

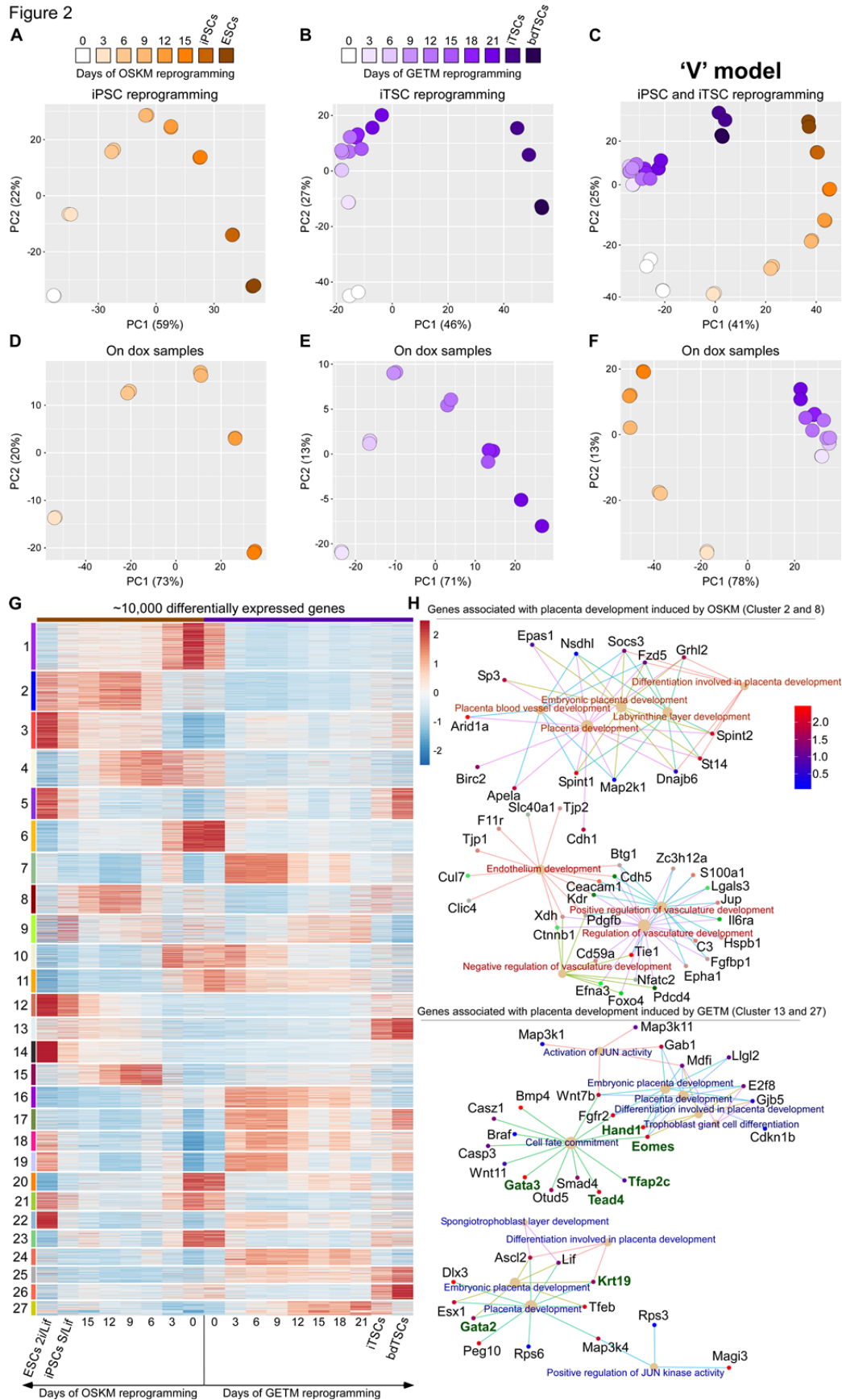


911

912 **Fig. 1. Establishment of the pluripotent and trophectoderm states in the embryo and during somatic**  
913 **nuclear reprogramming. (A)** An illustration of the various early embryonic stages and forming cell types  
914 during embryogenesis (adapted from<sup>108</sup>). Inner cell mass (ICM, purple) and trophectoderm (TE, green)  
915 are the first compartments to show a clear transcriptional specification. **(B + C)** Single-cell RNA  
916 sequencing data obtained from different stages of developing embryo<sup>5</sup> demonstrating the trajectory  
917 from zygote to blastocyst. PCA graphs showing gene expression profiles among 252 single cells  
918 projected onto the first two principal components. The trajectory from the zygote to the blastocyst is  
919 following a U-like shape (B) while the exclusion of totipotent cells (zygote and 2-cell stage (2C)) allows  
920 the visualization of a T-like shape progression segregating the TE from the ICM. **(D)** Single-cell RNA  
921 sequencing data obtained from different stages of developing embryo<sup>7</sup> demonstrating the trajectory  
922 from the zygote to the blastocyst stage. Diffusion map was constructed by MERLOT package using 48  
923 genes in 433 individual cells obtained from 2C through blastocyst. Again a clear T-like shape progression  
924 is noted separating the ICM from the TE. **(E)** Representative bright field images showing cell morphology  
925 and cell density during OSKM reprogramming toward iPSC formation (top) and during GETM  
926 reprogramming toward iTSC generation (bottom). **(F)** FACS analysis for Nanog-2A-EGFP reporter on BYKE  
927 MEFs undergoing reprogramming for 15 days with OSKM factors followed by 5 days of dox removal. **(G)**  
928 FACS analysis for Elf5-2A-EYFP-NLS reporter on BYKE MEFs undergoing reprogramming for 21 days with  
929 GETM factors followed by 10 days of dox removal. **(H)** Bright field and fluorescence images of a stable  
930 BYKE iPSC colony demonstrating the activation of the 3 pluripotent reporters (Utf1-2A-tdTomato/Esrrb-  
931 2A-TagBFP/Nanog-2A-EGFP). **(I)** Bright field and fluorescence images of a stable BYKE iTSC colony  
932 demonstrating the activation of the 3 TSC reporters (Utf1-2A-tdTomato/Esrrb-2A-TagBFP/Elf5-2A-EYFP-  
933 NLS).



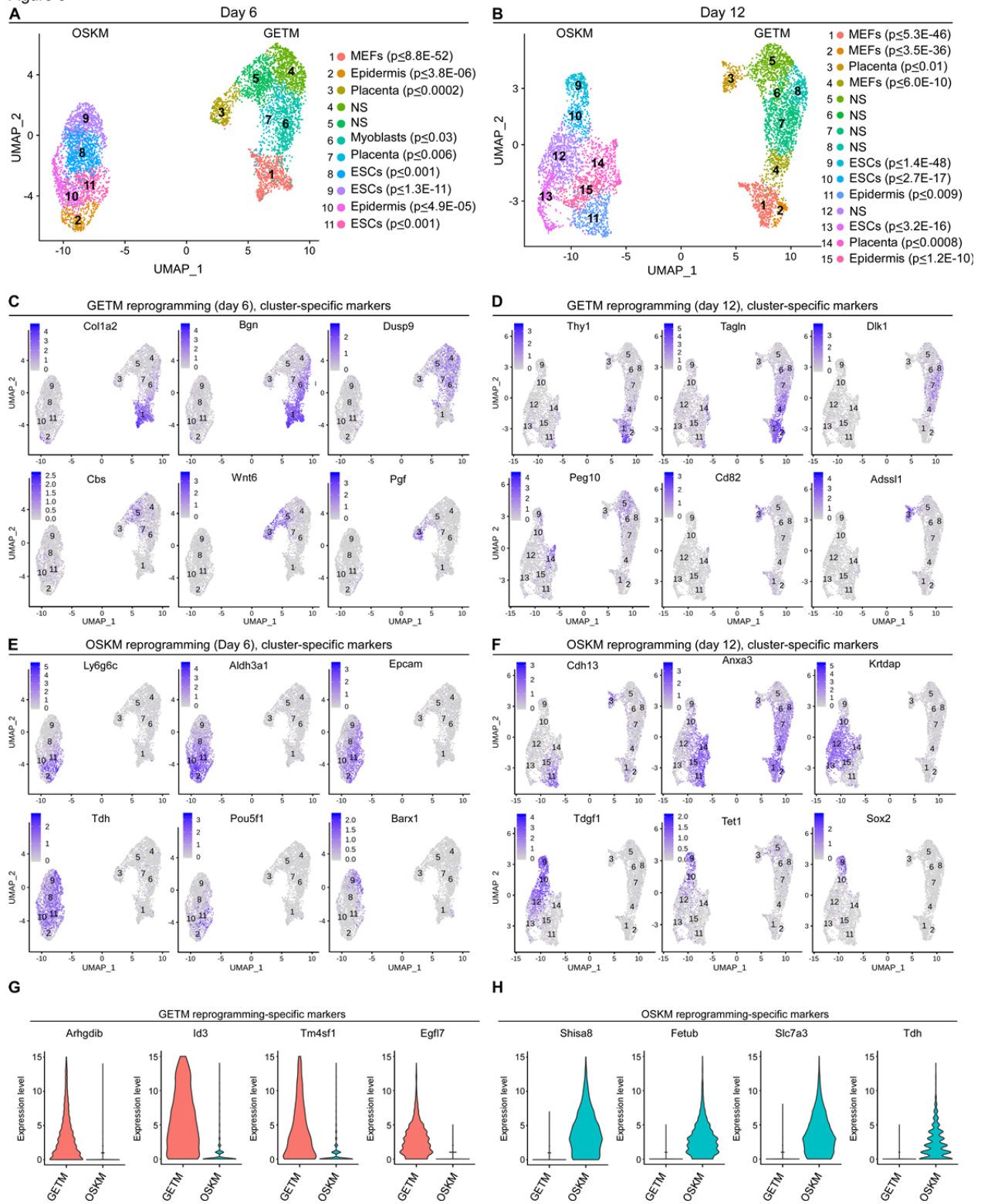
Figure 2





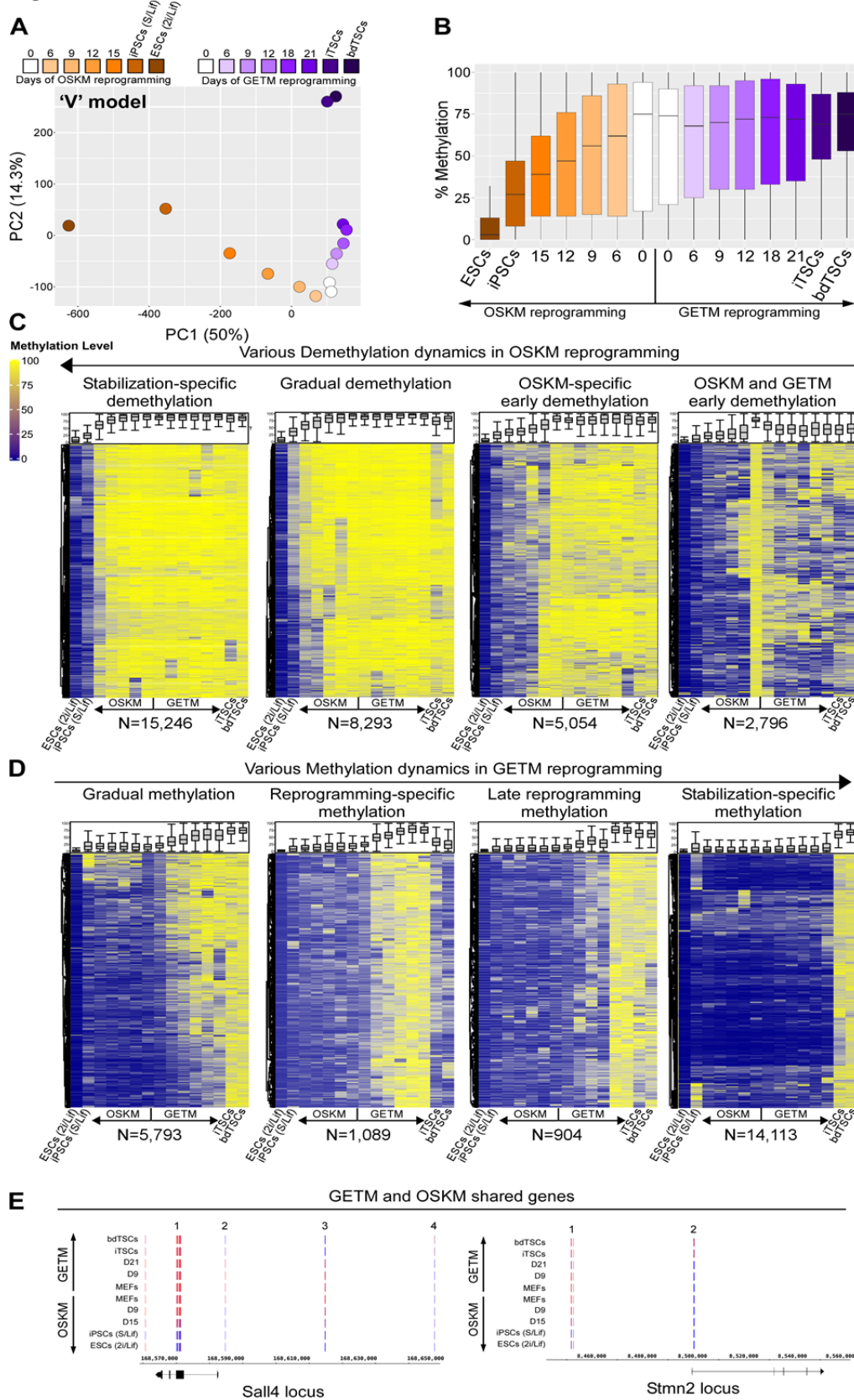
935 **Fig. 2. Bulk RNA-seq analysis on cells undergoing reprogramming to iPSCs and iTSCs suggests a unique**  
936 **transcriptional profile for each reprogramming system. (A-C)** PCA plots describing the trajectory MEFs  
937 undergo during reprogramming to either iPSCs (A), iTSCs (B) or both (C) as assessed by gene expression  
938 profiles of bulk RNA-seq data projected onto the first two principal components. **(D-F)** same as in (A-C)  
939 but here only reprogrammable cells (cells on dox) are plotted. **(G)** Heatmap showing gene expression  
940 levels detected by bulk RNA-seq of 10,000 most variable genes among ESCs, bdTSCs, MEFs and cells  
941 during reprogramming. Unsupervised hierarchical clustering was performed and adaptive branch  
942 pruning was used to identify 27 prominent clusters. **(H)** Gene-concept network of GO terms associated  
943 with placenta development induced by OSKM (upper panel) or by GETM (lower panel) reprogramming.  
944 Key regulators of the trophoblast stem cell state such as Gata3, Gata2, Tfap2c, Tead4 and Eomes  
945 (marked by green) are only specific to the GETM reprogramming.

Figure 3



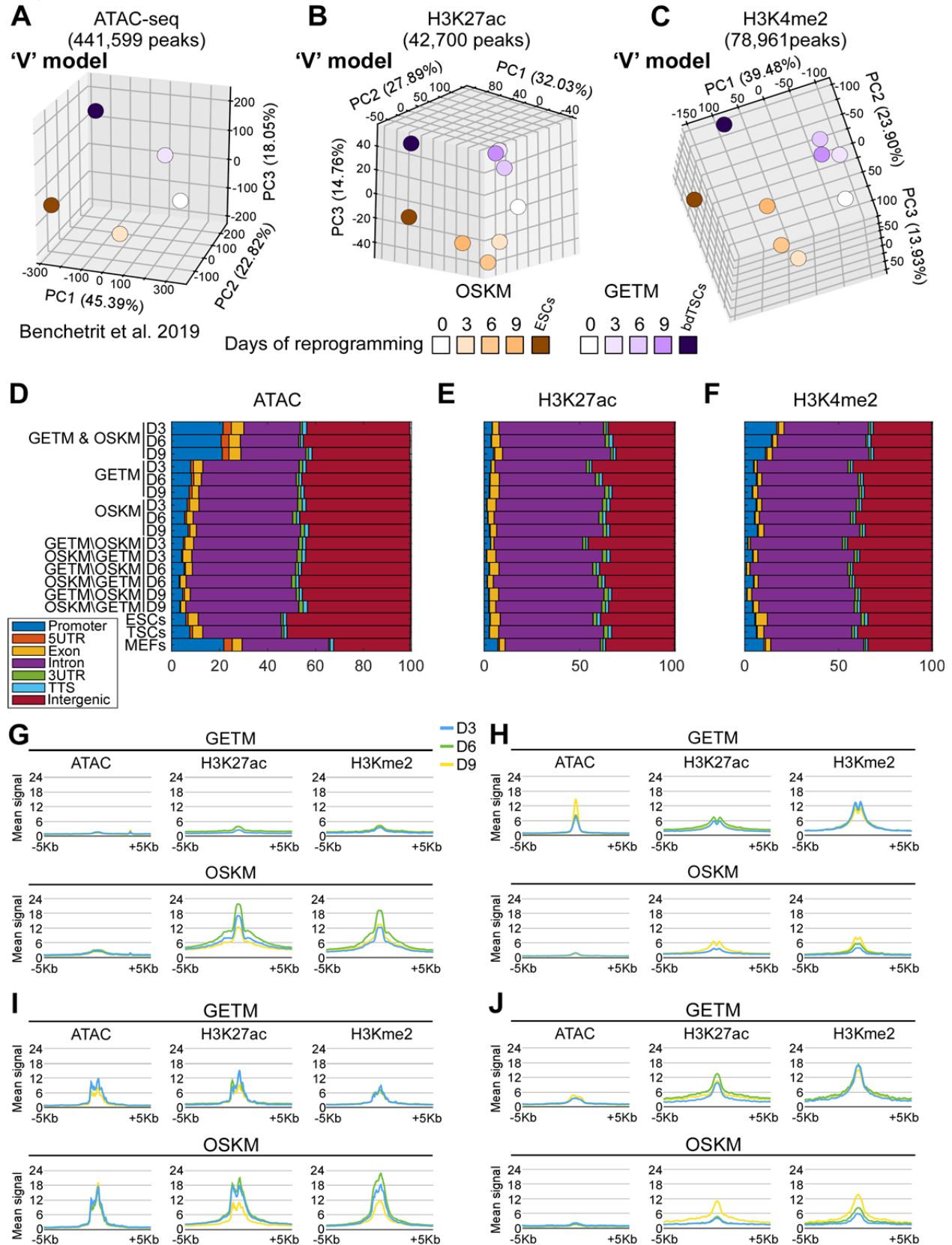
947 **Fig. 3. Single-cell RNA-seq analysis separates OSKM from GETM reprogramming. (A)** Uniform Manifold  
948 Approximation and Projection (UMAP) visualization analysis of 5,899 single cells at day 6 of both OSKM  
949 and GETM reprogramming. Each point represents a single cell and each color represents a unique  
950 community among the population. 1000 marker genes were used to characterize all subpopulation and  
951 using the Mouse Gene Atlas (“ARCHS4”: <https://amp.pharm.mssm.edu/archs4/>; “MGA”: Mouse Gene  
952 Atlas), the closest significant cell type was assigned to each subpopulation. **(B)** UMAP visualization  
953 analysis of 5,752 single cells at day 12 of both OSKM and GETM reprogramming. 1217 marker genes  
954 were used to characterize all subpopulation and using the Mouse Gene Atlas, the closest significant cell  
955 type was assigned to each subpopulation. **(C + D)** Expression level of selected cluster-specific markers of  
956 GETM reprogramming at day 6 and day 12, respectively. The expression level of the specified markers is  
957 visualized by a range of intensities of a purple color. **(E + F)** Expression level of selected cluster-specific  
958 markers of OSKM reprogramming at day 6 and day 12, respectively. **(G + H)** Violin plots summarizing  
959 single-cell expression level of GETM-specific markers (G) and OSKM-specific markers (H) in GETM and  
960 OSKM reprogramming processes.

Figure 4



962 **Fig. 4. RRBS analysis demonstrates methylation specific dynamics between OSKM and GETM**  
963 **reprogramming. (A)** Average bulk DNA methylation data of cells undergoing reprogramming toward  
964 pluripotency and TSC state projected onto the first two principal components. A clear V-like shape  
965 progression is observed separating GETM from OSKM reprogramming. **(B)** Boxplot of DNA methylation  
966 level across bulk samples during reprogramming towards both pluripotent and TSC states. OSKM  
967 reprogrammable cells exhibit an overall hypomethylation dynamics on CpG-enriched sites while GETM  
968 reprogrammable cells exhibit hypermethylation dynamics. **(C-D)** Heatmaps demonstrating the dynamics  
969 of DNA methylation alterations and patterns across bulk samples during reprogramming towards both  
970 pluripotent and TSC states, respectively. Each row represents one differentially methylated block for  
971 which there are at least one CpG with  $\geq 10\times$  coverage. While most of the patterns observed are specific  
972 to each reprogramming system, a few clusters show similar DNA demethylation initiated as early as day  
973 6 in both systems. **(E)** DNA methylation level of genomic loci containing genes that are shared between  
974 pluripotent cells and TSCs.

Figure 5

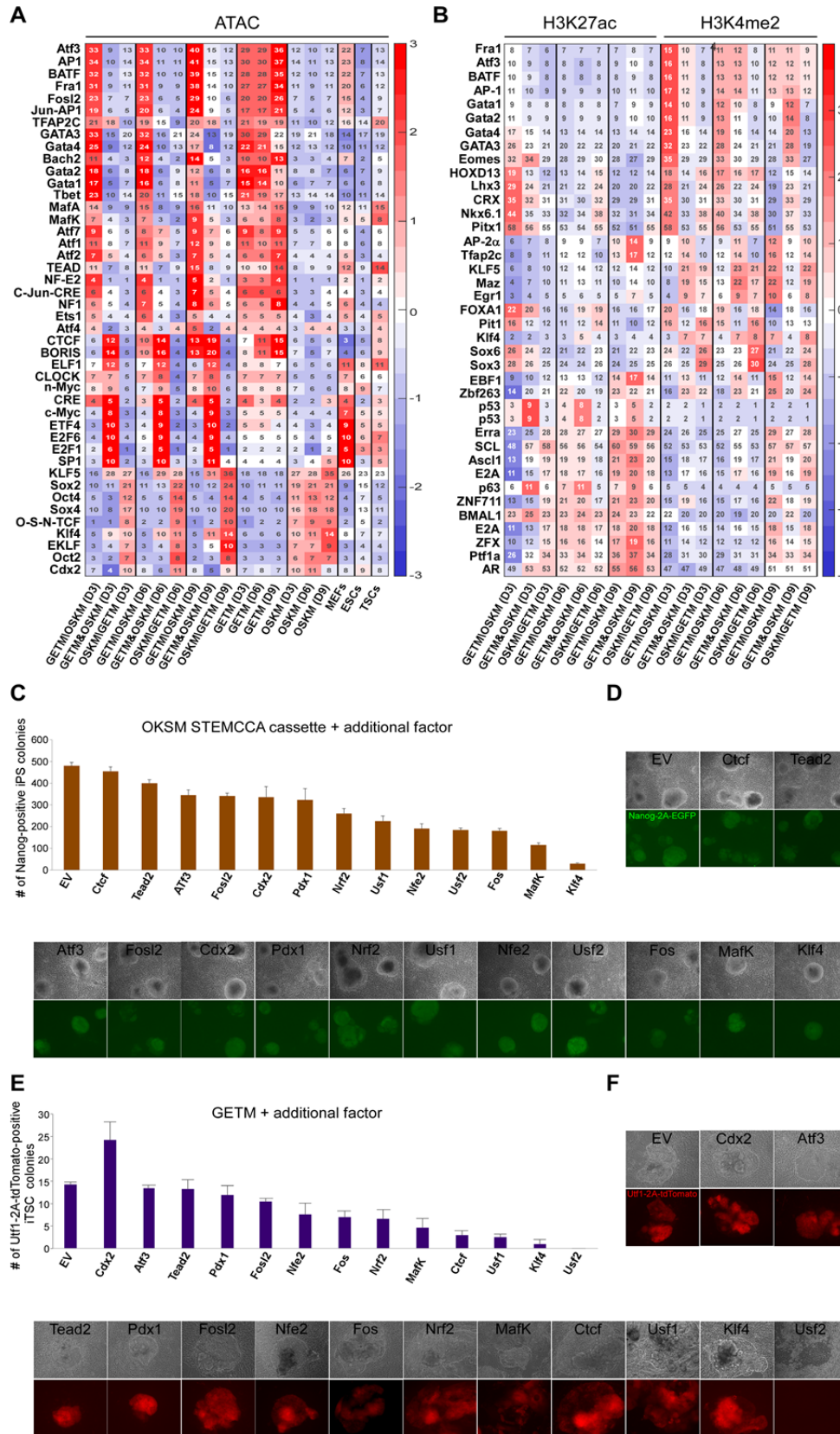


975

976 **Fig. 5. Chromatic accessibility and activity during GETM and OSKM reprogramming demonstrating a**  
977 **'V'-shaped behavior. (A-C)** Top 3 PCA components of the Z-scores of ATAC-seq (A), H3K27ac (B) and  
978 H3K4me2 peaks (C). Peaks were clipped to range [0, 500] and filtered by length ( $\geq 500$ bp). Replicates  
979 were merged by taking the mean peak height. **(D-F)** Genomic annotations of ATAC-seq peaks (D),  
980 H3K27ac peaks (E) and H3K4me2 peaks (F). Shown are the fraction of various genomic annotations  
981 (Promoter, Exons, Introns, etc) among peaks. Genomic regions accessible in both GETM and OSKM  
982 conditions (D, top three rows) are enriched for promoter regions, compared to GETM or OSKM regions  
983 (below). GETM and OSKM mark regions accessible in those conditions, excluding MEF peaks. Below are  
984 cell-type specific accessible regions such as "GETM\OSKM D3", which includes GETM Day 3 peaks not  
985 accessible in OSKM Day 3. In addition to Promoter regions (blue), most accessible regions fall within  
986 Intronic regions (purple) and Intergenic regions (red). **(G-J)** Mean ATAC-seq, H3K27ac and H3K4me2 at  
987  $\pm 5$ Kb surrounding OSKM or GETM H3K27ac peak locations, for Day 3 (blue), Day 6 (green), and Day 9  
988 (yellow). (G) OSKM-specific H3K27ac signal is strongest at Day 6, and is accompanied by matching  
989 H3K4me2 signal, but with no dynamic change in DNA accessibility. (H) Same for GETM ATAC-seq peaks.  
990 These regions are already marked by H3K4me2 in Day 3, and gain accessibility over time. These genome  
991 regions also show H3K27ac and H3K4me2 enrichments for later OSKM stages. (I) Same for OSKM ATAC-  
992 seq peaks. (J) Same of GETM H3K27ac peaks. These peaks show gradual increase in ChIP-seq signal even  
993 following OSKM induction (below).



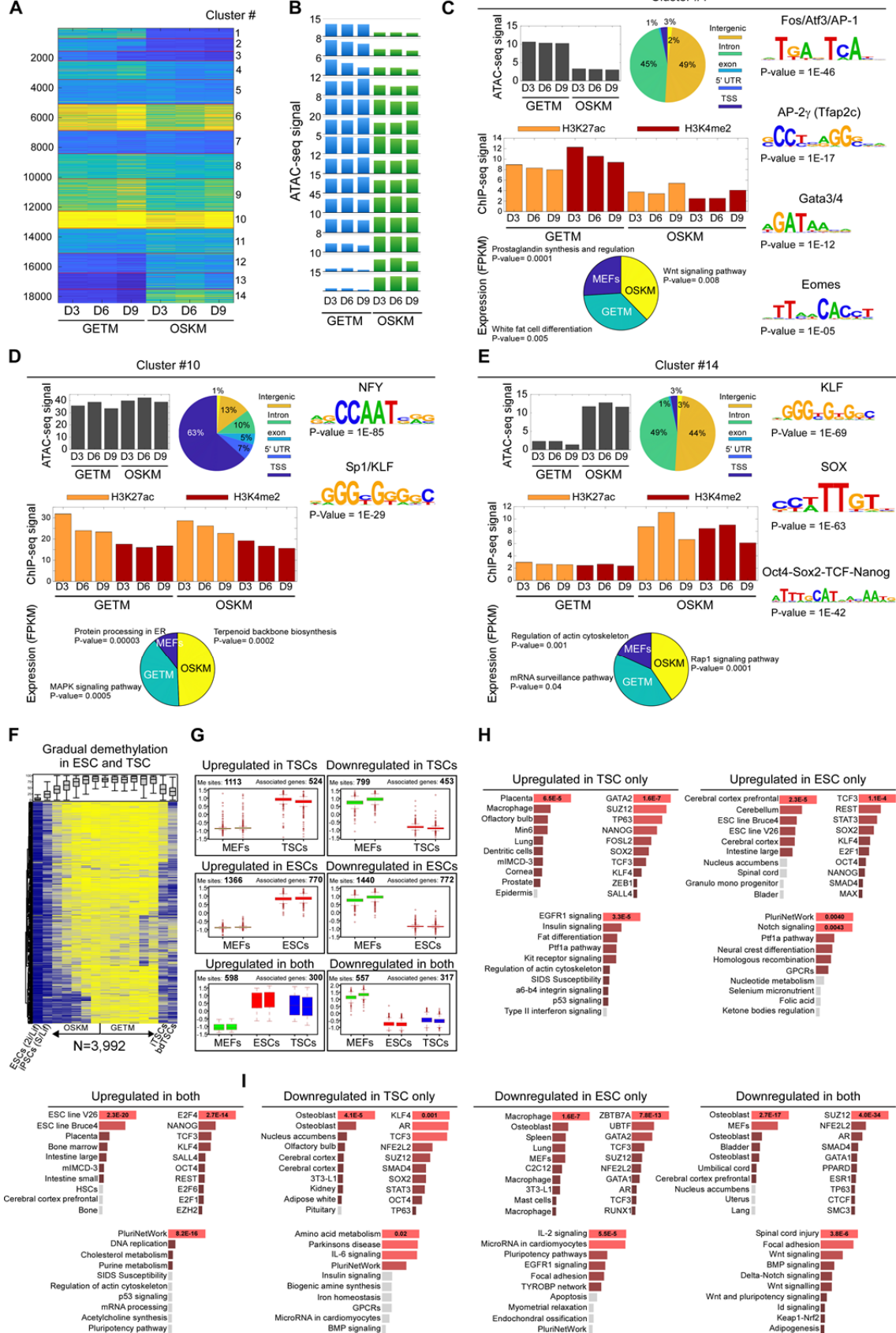
Figure 6



995 **Fig. 6. Motif enrichment and the effect of their corresponding transcription factor on OSKM and GETM**  
996 **reprogramming**

997 **(A)** Heatmap showing motif enrichment among ATAC-seq peaks. For each row (motif) and each column  
998 (condition-specific ATAC-seq peaks) we calculated the percent of peaks containing it (shown numbers).  
999 Subsets of peaks include GETM-only peaks (GETM\OSKM), joined peaks (GETM&OSKM), and OSKM-only  
1000 peaks (OSKM\GETM) for each time point (Days 3, 6, 9). Also shown are joined sets of GETM and OSKM  
1001 peaks for each day, as well as MEF, ESC and TSC peaks. Each motif/condition is color-coded based on  
1002 relative motif enrichment (Z-scores) compared to all conditions. Only motifs with enrichment greater  
1003 than 2.5 standard deviations ( $Z > 2.5$ ) are shown. **(B)** Heatmap showing motif enrichment among H3K27ac  
1004 and H3K4me2 peaks. For each row and each column, we calculated the percent of peaks containing it  
1005 (shown numbers). Subsets of peaks include GETM-only peaks (GETM\OSKM), joined peaks  
1006 (GETM&OSKM), and OSKM-only peaks (OSKM\GETM) for each time point (Days 3, 6, 9). **(C-D)** BYKE MEFs  
1007 were infected with dox-inducible OSKM STEMCCA cassette plus additional factor as depicted. The cells  
1008 were reprogrammed for 8 days and then weaned of dox for additional 5 days. Nanog-2A-EGFP-positive  
1009 colonies were counted (C) and imaged (D). EV refers to empty vector control. **(E-F)** BYKE MEFs were  
1010 infected with dox-inducible GETM factors plus additional factor as depicted. The cells were  
1011 reprogrammed for 21 days and then weaned of dox for additional 10 days. Utf1-2A-tdTomato-positive  
1012 colonies were counted (E) and imaged (F). EV refers to empty vector control.

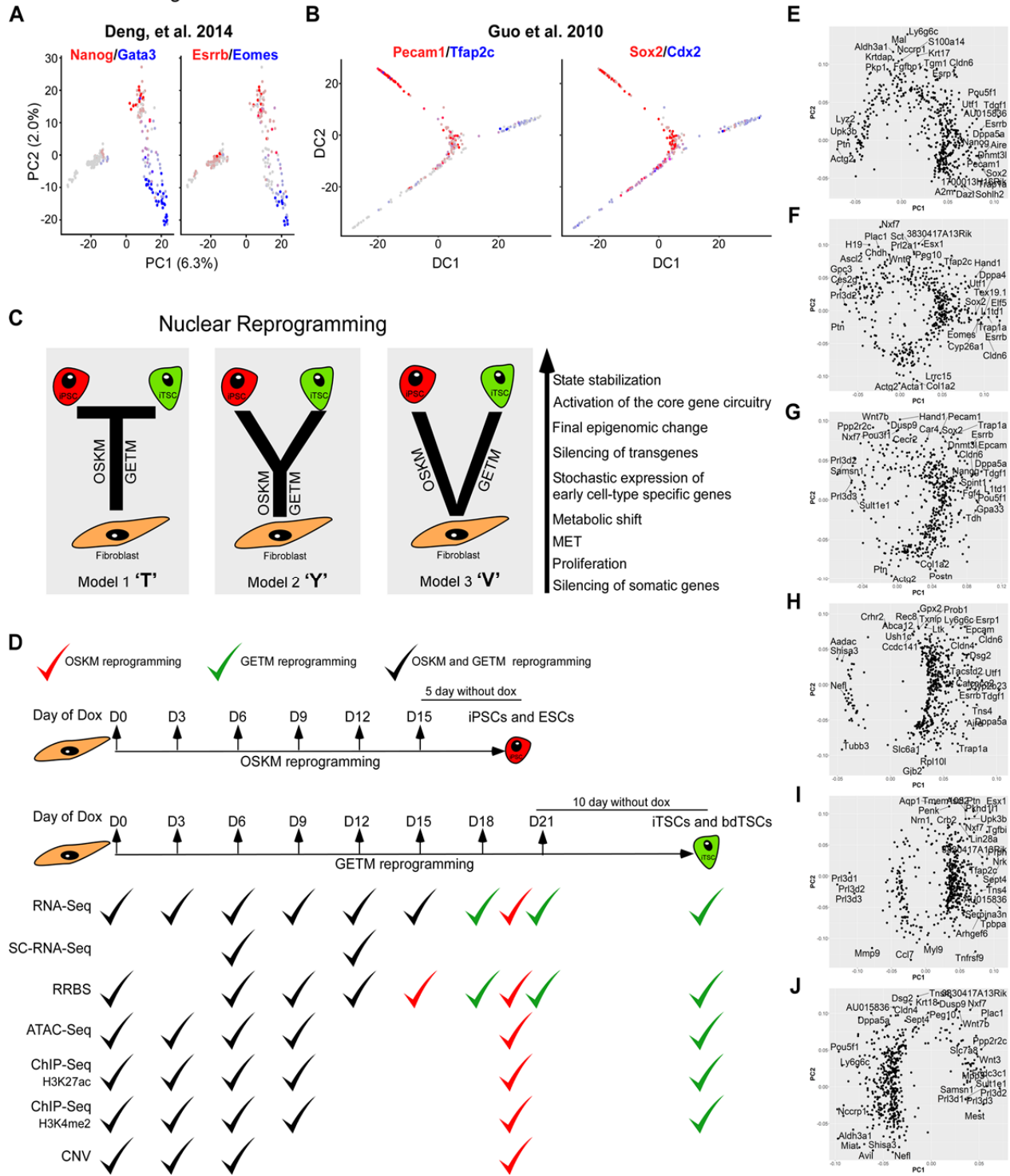
Figure 7



1014 **Fig. 7. Data integration of DNA methylation, gene expression and chromatin accessibility and activity**  
1015 **during the reprogramming process toward pluripotent and TSC states. (A-B)** Clustering of 18,420 GETM  
1016 and OSKM ATAC-seq peaks from days 3, 6, 9 into 14 clusters is shown as a heatmap (A) or barplot of  
1017 mean ATAC-seq signal per cluster (B). **(C)** Cluster #1 is mostly composed of distal (Intergenic and  
1018 Intronic) GETM-specific peaks, enriched for AP, GATA and Eomes motifs, and near GETM-expressed  
1019 genes. Shown are mean ATAC-seq signals (top left), analysis of their genomic annotations (pie chart,  
1020 center), enriched transcription factor motifs (right panel), average ChIP-seq signals of H3K27ac and  
1021 H3K4me2 following GETM and OSKM induction (middle panel), and a pie chart for RNA expression levels  
1022 and GO term for genes that are associated with each cluster ATAC-seq peaks and exhibit the highest  
1023 expression levels in MEFs (blue), or GETM (green) or OSKM (yellow, Bottom panel). **(D)** Same for cluster  
1024 10, enriched for highly accessible promoter peaks. **(E)** Same for cluster 14, with regions that are highly  
1025 accessible following OSKM, enriched for distal regions with KLF, SOX and Oct4 motifs, and are associated  
1026 with OSKM expressed genes. **(F)** A heatmap of differentially methylated blocks with DNA demethylation  
1027 during the final states of reprogramming to both pluripotent and TSC states. Each row represents one  
1028 out of 3992 blocks of DMBs. **(G)** Boxplots of relative expression of differentially expressed genes that are  
1029 associated with each individual block of DNA methylation. A significant negative correlation between  
1030 DNA methylation and expression is observed in both reprogramming systems in which DMBs are  
1031 associated with genes upregulated in both systems such as *Esrrb*, *Sox2*, *Sall4*, *Dppa4* and *Zfp42*. A  
1032 significant positive correlation is also observed in both reprogramming systems in which DMBs are  
1033 associated with MEFs genes downregulated in both systems such as *Acta2*, *Col5a1*, *Col5a2*, *Runx1* and  
1034 *Runx2*. **(H-I)** Enrichr Mouse gene atlas and KEGG pathways analysis of significantly over-represented  
1035 genes that are either upregulated (H) or downregulated (I) for each depicted group in (G).

1036 SUPPLEMENTARY

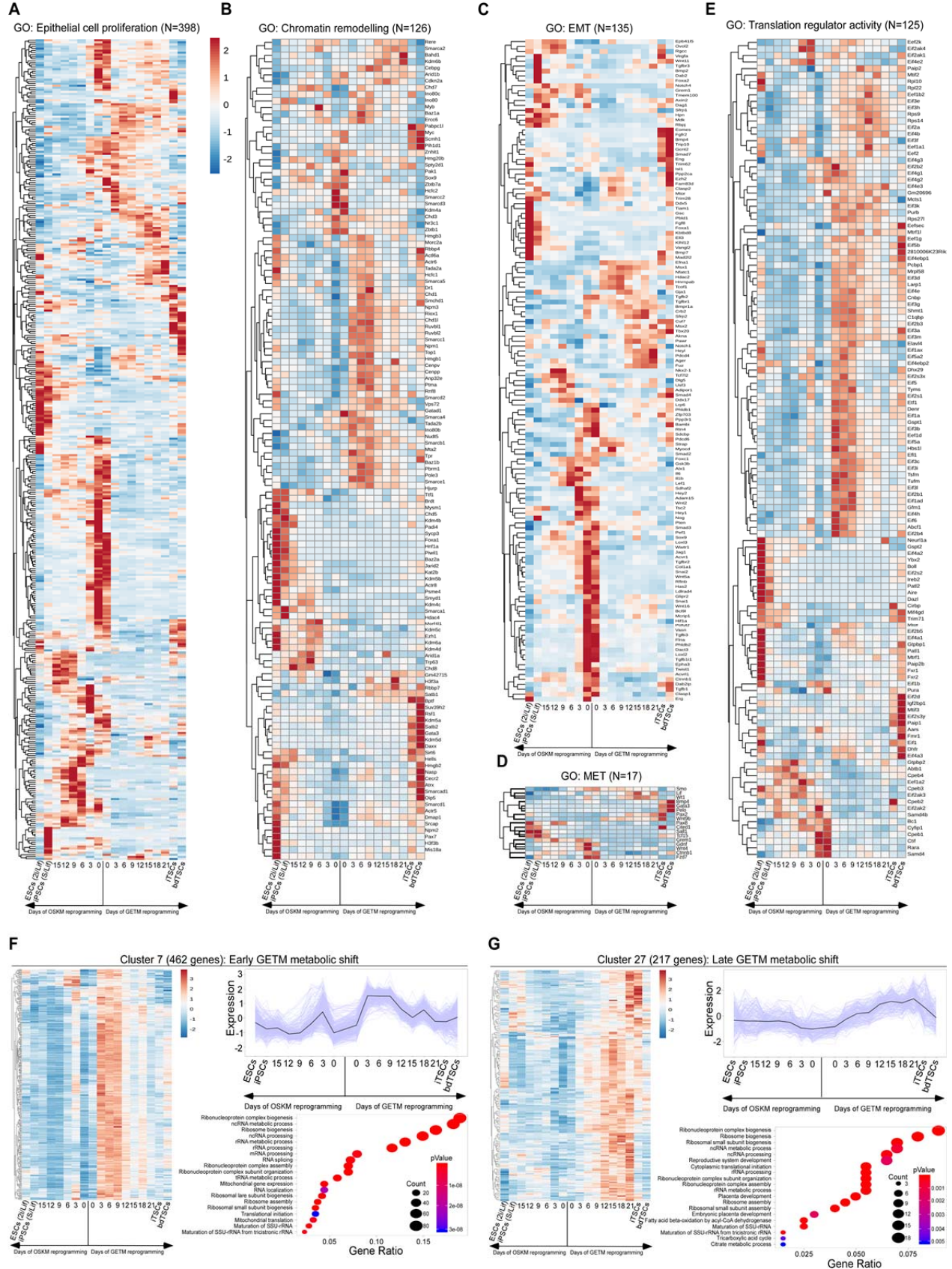
Extended Data Fig.1





1038 **Supplementary Fig. 1. Embryonic development trajectory and hypothetical nuclear reprogramming**  
1039 **progression models using comparative multi-omics analysis. (A)** Overplayed co-expression of  
1040 Nanog/Gata3 and Esrrb/Eomes among individual cells extracted from<sup>5</sup>. Red color indicates cells that are  
1041 enriched with key pluripotency genes while blue color indicates cells with enrichment of key TE genes.  
1042 **(B)** Overplayed co-expression of Pecam1/Tfap2c and Sox2/Cdx2 among individual cells extracted from<sup>7</sup>.  
1043 Red color indicates cells that are enriched with key pluripotency genes while blue color indicates cells  
1044 with enrichment of key TE genes. **(C)** Schematic illustration of three possible models, ‘T’, ‘Y’, ‘V’,  
1045 explaining the reprogramming progression of fibroblasts toward either iPSCs by OSKM or iTSCs by  
1046 GETM. **(D)** Schematic representation of the reprogramming process of fibroblasts to iPSCs (top, red) and  
1047 iTSCs (bottom, green) and the various high throughput experiments and time points that were analyzed.  
1048 Black ‘V’ represents a time point that was taken for both GETM and OSKM reprogramming while Green  
1049 ‘V’ represents GETM-only time point and red ‘V’ represents OSKM-only time point. **(E-G)** PCA loading  
1050 plots showing the contribution of individual genes out of top 500 most differentially expressed genes to  
1051 the first and second PCA components; distance from the origin along each axis corresponds to strength  
1052 of contribution to that component. Shown are PCA loading plots for the reprogramming process to  
1053 either iPSCs (E), iTSCs (F) or both (G) as assessed by gene expression profiles of bulk RNA-seq data  
1054 projected onto the first two principal components. 32 highest ranked genes are marked in each plot. **(H-**  
1055 **J)** same as in (E-G) but here only reprogrammable cells (cells on dox) are plotted.

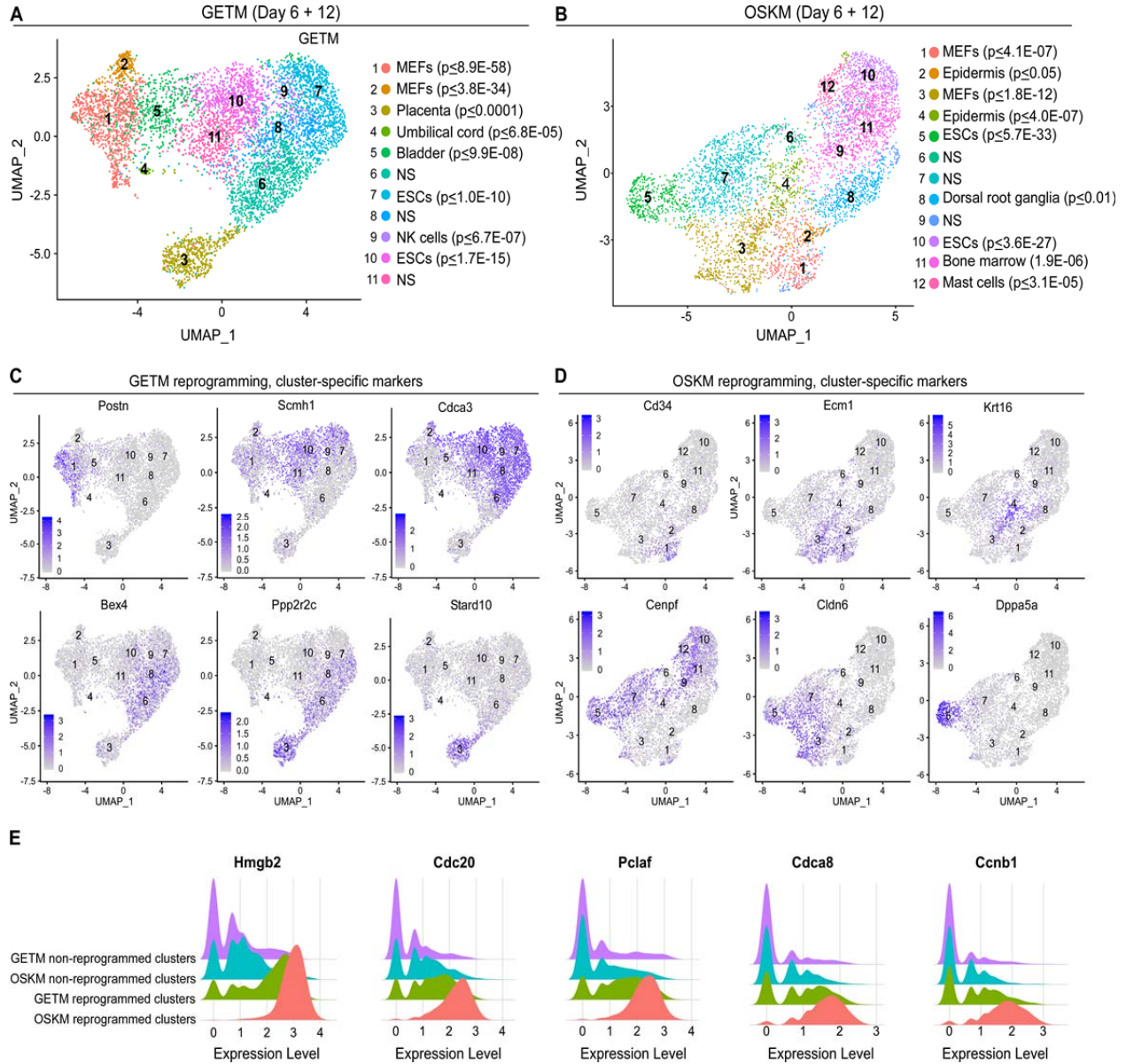
Extended Data Fig.2





1057 **Supplementary Fig. 2. GETM and OSKM reprogramming factors mostly exhibit mutually exclusive**  
1058 **transcriptional profiles during reprogramming. (A-E)** Heatmaps showing the expression levels of genes  
1059 involved in early and general reprogramming processes such as epithelial cell proliferation (A),  
1060 chromatin remodeling (B), EMT (C), MET (D) and translation regulator activity (E), during the conversion  
1061 of fibroblasts into iPSCs by OSKM and into iTSCs by GETM. **(F)** Heatmap, expression pattern plot and GO  
1062 terms of 462 genes of cluster #7 as detected by bulk RNA-seq during reprogramming toward iPSCs and  
1063 iTSCs. **(G)** Heatmap, expression pattern plot and GO terms of 217 genes of cluster #27 as detected by  
1064 bulk RNA-seq during reprogramming toward iPSCs and iTSCs.

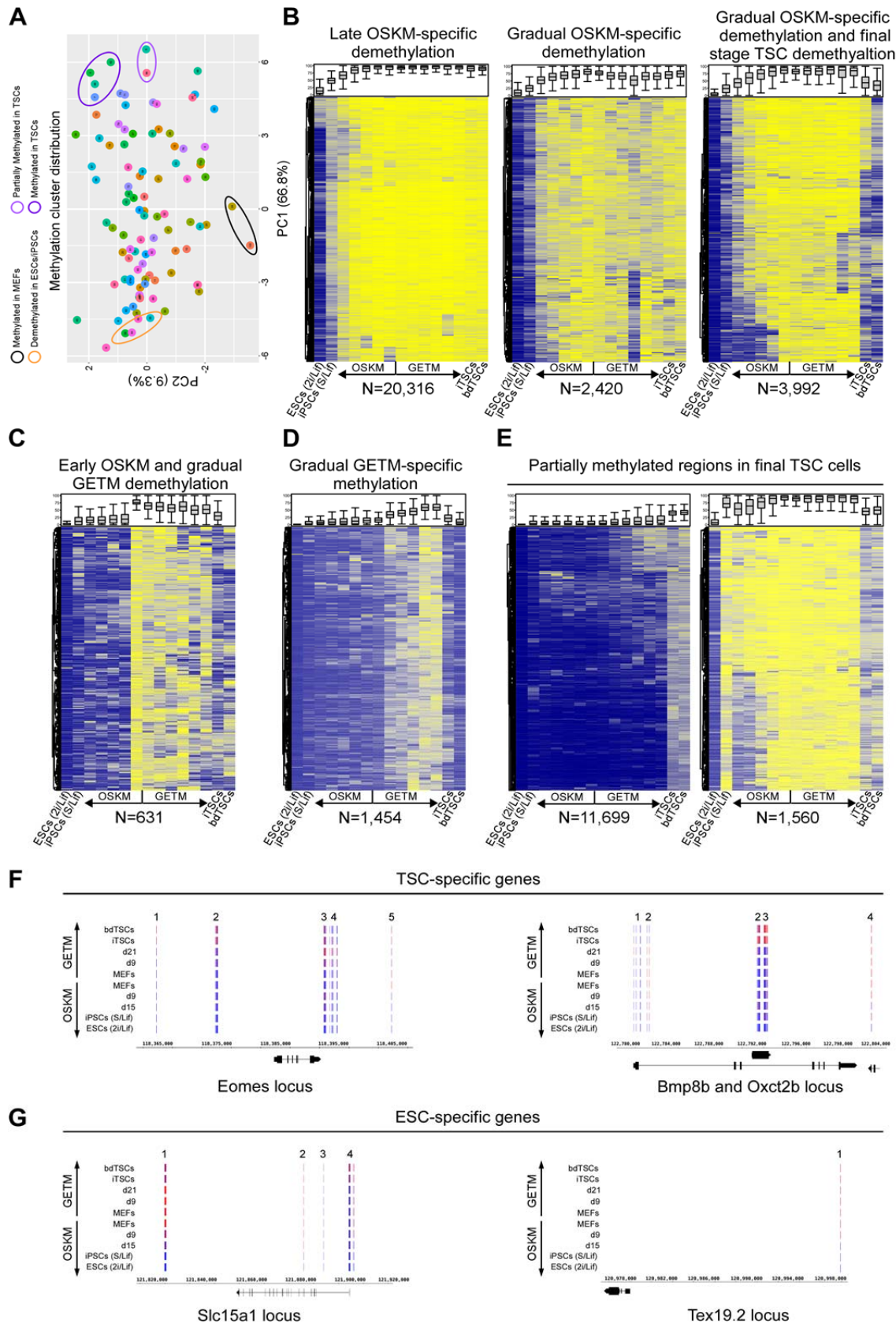
Extended Data Fig.3



1065

1066 **Supplementary Fig. 3. Single-cell RNA-seq analysis on OSKM and GETM reprogrammable cells**  
1067 **identifies reprogramming and stage-specific markers (A)** UMAP visualization analysis of both 3,288  
1068 single cells at day 6 and 3,029 single cells at day 12 of GETM reprogramming. Marker genes were used to  
1069 characterize all subpopulation and using the mouse gene atlas only the closest significant cell type was  
1070 assigned to each subpopulation. NS refers to non-significant. **(B)** UMAP visualization analysis of both  
1071 2,611 single cells at day 6 and 2,723 single cells at day 12 of OSKM reprogramming. Marker genes were  
1072 used to characterize all subpopulation and using the mouse gene atlas only the closest significant cell  
1073 type was assigned to each subpopulation. **(C-D)** Expression level of selected cluster-specific markers  
1074 during GETM reprogramming (days 6 and 12) and OSKM reprogramming (days 6 and 12), respectively.  
1075 The expression level of the specified markers is visualized by a range of intensities of a purple color. **(E)**  
1076 Redge plots showing the expression distributions of key proliferation genes where each community  
1077 represents potentially reprogrammed or non-reprogrammed cells for each reprogramming system.

Extended Data Fig.4



1079 **Supplementary Fig. 4. Methylation dynamics during OSKM and GETM reprogramming**

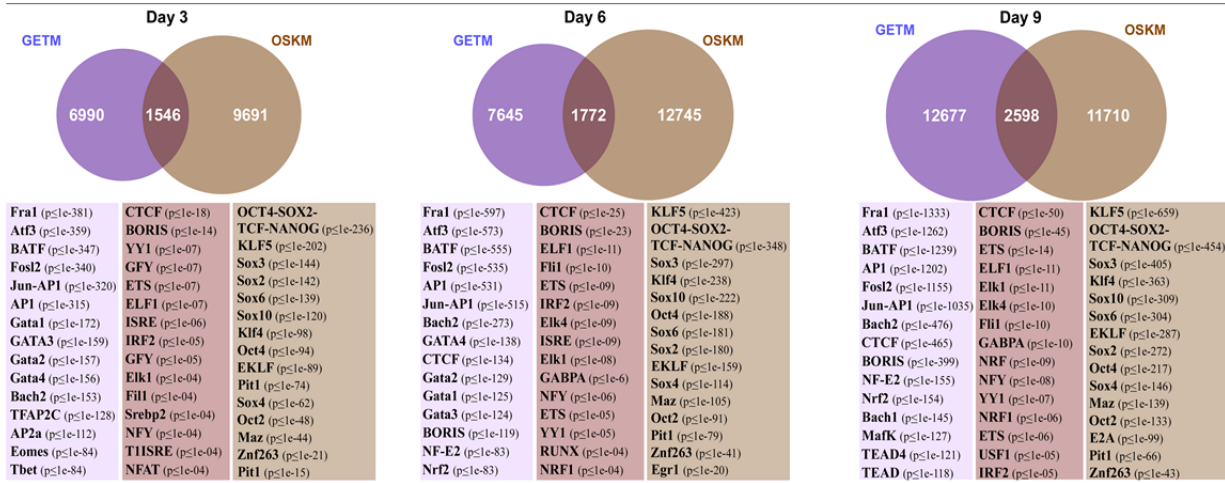
1080 **(A)** PCA plot projected by the first two principal components showing the average DNA methylation per  
1081 sample for 100 clusters obtained using K-Means clustering algorithm of 130,000 blocks. Clusters that  
1082 are near to each other show similar trend of methylation. **(B-E)** Heatmaps showing the dynamics of DNA  
1083 methylation alterations and specific cluster patterns across bulk samples during reprogramming towards  
1084 both pluripotent and TSC states. **(F)** DNA methylation level of genomic loci that contain gene specific to  
1085 TSCs (e.g. Eomes, Bmp8b and Oxct2b). **(G)** DNA methylation level of genomic loci that contain gene  
1086 specific to ESCs (e.g. Slc15a1 and Tex19.2).



Extended Data Fig.5

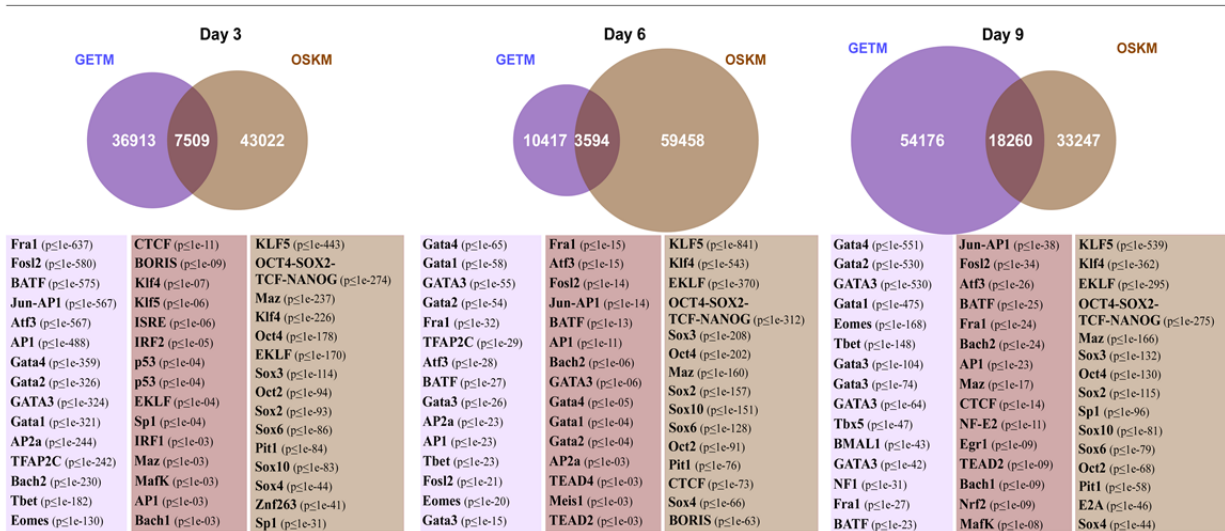
A

Open and remodeled regions during GETM and OSKM reprogramming (ATAC peaks)



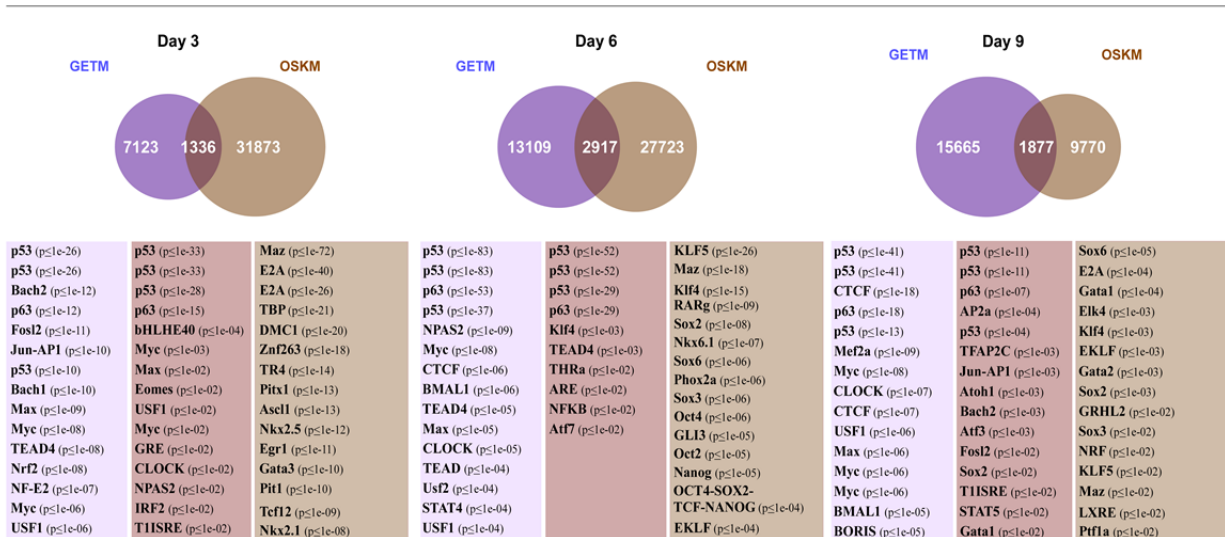
B

Open and remodeled regions during GETM and OSKM reprogramming (H3K4me2 peaks)



C

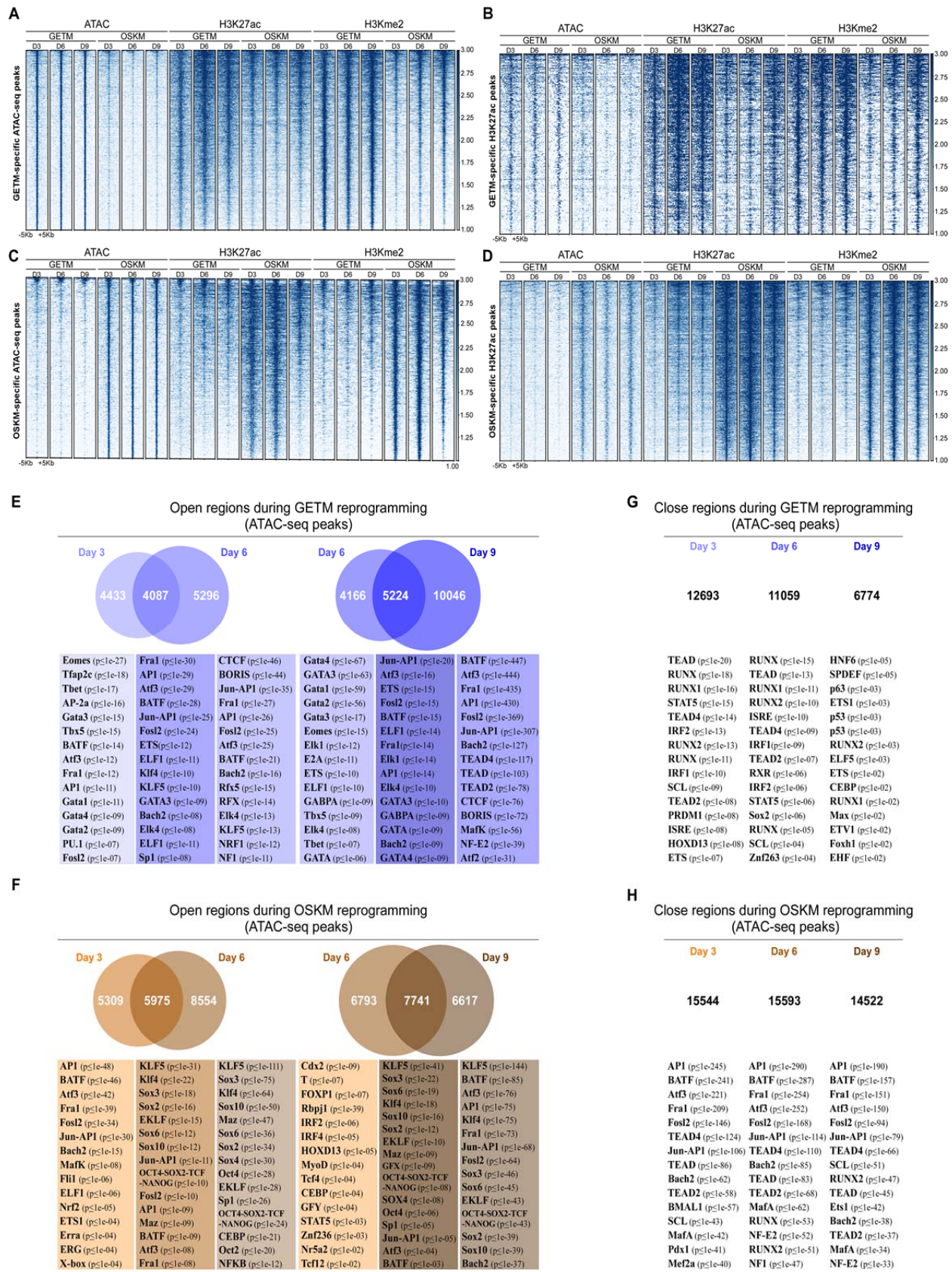
Active regions during GETM and OSKM reprogramming (H3K27ac peaks)



1088 **Supplementary Fig. 5. Motif enrichment of ATAC, H3K27ac and H3K4me2 peaks during GETM and**  
1089 **OSKM reprogramming. (A-C)** Venn diagrams and motif analysis for ATAC-seq peaks (A), H3K4me2 peaks  
1090 (B) and H3K27ac peaks (C). Comparison of GETM-only (left wedge, purple), GETM and OSKM  
1091 (interaction), and OSKM-only (right wedge, brown) peaks from day 3 to day 9. Below are motifs,  
1092 differentially enriched between each set of peaks vs. the rest.

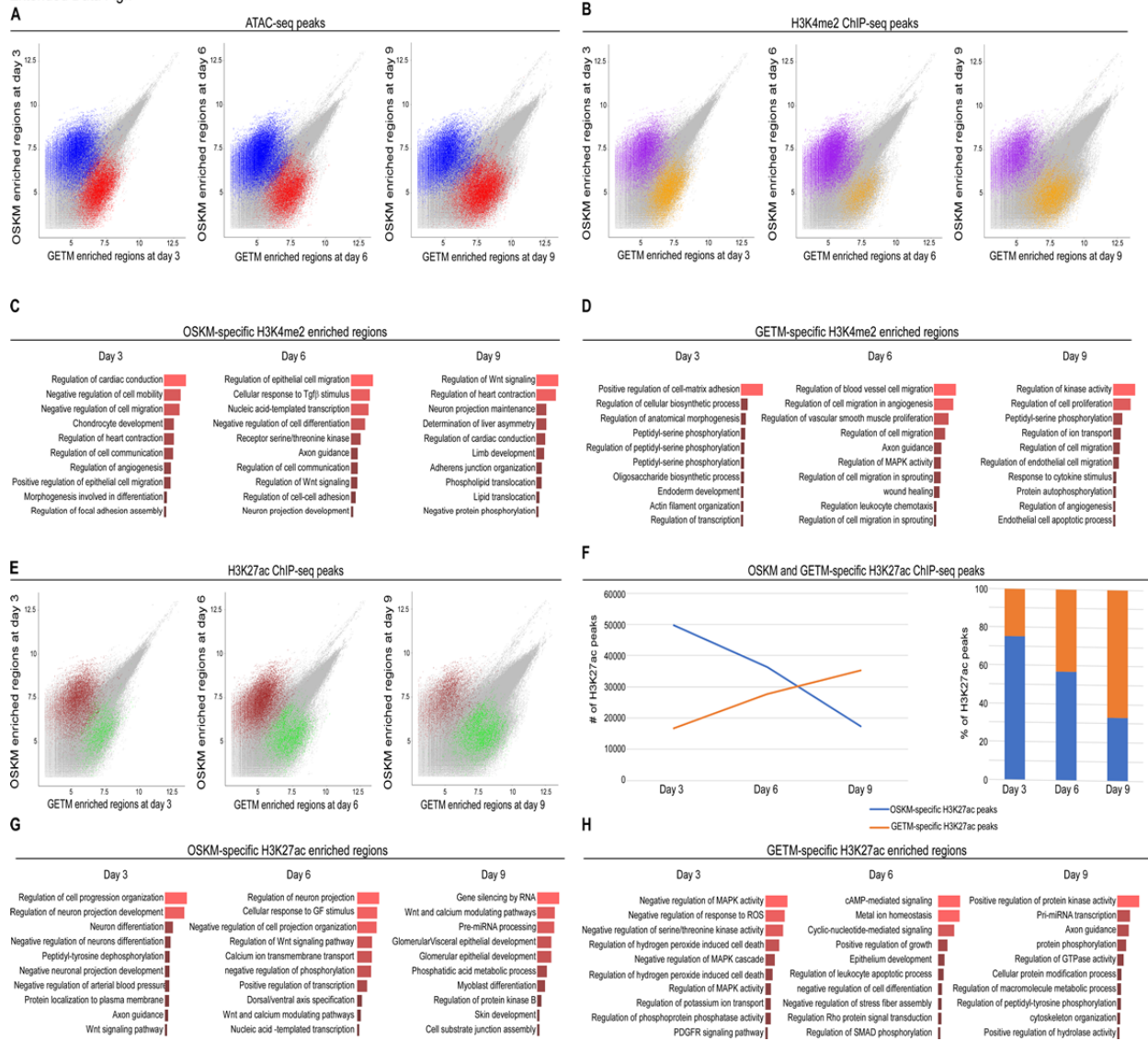


Extended Data Fig. 6



1094 **Supplementary Fig. 6. Unique behavior and motif enrichment of ATAC, H3K27ac and H3K4me2 peaks**  
1095 **during GETM and OSKM reprogramming. (A-D)** Heatmap showing ATAC-seq and ChIP-seq (H3K27ac and  
1096 H3K4me2) across 1716 differential ATAC-seq peaks ( $p < 1e-3$ ) in GETM (A and B) or 2848 OSKM (C and D).  
1097 Shown are genomic regions of peak locations  $\pm 5Kb$ . Differential peaks were called using DESeq2 analysis  
1098 on the number of reads in each of 18,421 ATAC-seq peaks, using a significance threshold of adjusted p-  
1099 value  $< 1e-3$ . **(E)** Comparison of GETM ATAC-seq peaks from days 3 and 6 (left) or 6 and 9 (right). Shown  
1100 below are enriched motifs for each binary set of peaks. From left to right: day 3-only (left wedge) vs all  
1101 day 6 peaks; day 3 and 6 peaks (intersection) vs day 3-only and day 6-only peaks; day 6-only (right  
1102 wedge) vs all day 3 peaks; day 6-only (left wedge) vs all day 9 peaks; day 6 and 9 peaks (intersection) vs  
1103 day 6-only and Day 9-only peaks; day 9-only (right wedge) vs all day 6 peaks. **(F)** Same as E but for OSKM  
1104 **(G)** Similar analysis for 12,693 genomic regions that are accessible in MEFs but close on GETM day 3; or  
1105 11,059 regions that are accessible in day 3 but close in day 6; or 6774 regions that are accessible in Day  
1106 6 but close in day 9 **(H)** Same as G but for OSKM.

Extended Data Fig.7



1107

1108 **Supplementary Fig. 7. OSKM and GETM enriched regions for ATAC-seq peaks, H3K4me2 peaks and**  
1109 **H3K27ac peaks. (A)** Scatter plot of enriched regions differentially accessible between different cellular  
1110 states during the reprogramming process using GETM and OSKM at days 3, 6, and 9. Differential  
1111 accessible regions are marked with blue for OSKM, and red for GETM with adjusted p-value < 0.001. **(B)**  
1112 Scatter plot of enriched regions that are both differentially accessible and enriched for H3K4me2 ChIP-  
1113 seq differential peaks on top of the ATAC-seq signal during the reprogramming process using GETM and  
1114 OSKM at days 3, 6 and 9. Enriched regions are marked with purple for OSKM and orange for GETM. **(C)**  
1115 Top 10 enriched gene ontology (GO) terms within OSKM-specific regions that are both differentially  
1116 accessible and enriched for H3K4me2 at days 3, 6 and 9 tested in the biological process ontology. **(D)**  
1117 Top 10 enriched gene GO terms within GETM-specific regions that are both differentially accessible and  
1118 enriched for H3K4me2 at days 3, 6 and 9 tested in the biological process ontology. **(E)** Scatter plot of  
1119 enriched regions that are both differentially accessible and enriched for H3K27ac ChIP-seq differential  
1120 peaks on top of the ATAC-seq signal for enriched regions during the reprogramming process using GETM  
1121 and OSKM at days 3, 6 and 9. Enriched regions are marked with maroon for OSKM and green for GETM.  
1122 **(F)** Line plot and stacked column chart showing differential dynamics of enrichment for H3K27ac during  
1123 both OSKM and GETM reprogramming at days 3, 6, 9. **(G)** Top 10 enriched GO terms within OSKM-  
1124 specific regions that are both differentially accessible and enriched for H3K27ac at days 3, 6 and 9 tested  
1125 in the biological process ontology. **(H)** Top 10 enriched GO terms within GETM-specific regions that are  
1126 both differentially accessible and enriched for H3K27ac at days 3, 6 and 9 tested in the biological process  
1127 ontology.



Extended Data Fig.8



1128

1129 **Supplementary Fig. 8. Data integration of chromatin accessibility and activity and gene expression and**  
1130 **CNV analysis. (A-K)** 18,420 GETM and OSKM ATAC peaks from days 3, 6, 9 were clustered to 14 clusters.  
1131 Shown for each cluster are: mean ATAC-seq signal (top left), analysis of their genomic annotations (pie  
1132 chart, center), enriched transcription factor motifs (right panel), average ChIP-seq signals of H3K27ac  
1133 and H3K4me2 following GETM and OSKM induction (middle panel), and a pie chart for RNA expression  
1134 levels and GO terms for genes that are associated with each cluster ATAC-seq peaks and exhibit the  
1135 highest expression levels in MEFs (blue), or GETM (green) or OSKM (yellow, Bottom panel). **(L)** A graph  
1136 summarizing the various copy number variations (CNVs) identified in OSK/M or GET/M reprogrammable  
1137 cells (days 0, 3 and 6) and in isolated iPSC clones. Final TSCs/iTSCs hold an intrinsic capacity to  
1138 accumulate genomic aberrations<sup>13</sup> and thus are not measured here. Parental ESC line and partially  
1139 reprogrammed iPSC colonies number 20 and 23<sup>23</sup> were used as negative and positive control,  
1140 respectively. All data were aligned to the parental MEFs.

1141 **References**

- 1142 1 Messerschmidt, D. M., Knowles, B. B. & Solter, D. DNA methylation dynamics during epigenetic  
1143 reprogramming in the germline and preimplantation embryos. *Genes & development* **28**, 812-  
1144 828, doi:10.1101/gad.234294.113 (2014).
- 1145 2 Chen, L., Wang, D., Wu, Z., Ma, L. & Daley, G. Q. Molecular basis of the first cell fate  
1146 determination in mouse embryogenesis. *Cell research* **20**, 982-993, doi:10.1038/cr.2010.106  
1147 (2010).
- 1148 3 Jaber, M., Sebban, S. & Buganim, Y. Acquisition of the pluripotent and trophectoderm states in  
1149 the embryo and during somatic nuclear reprogramming. *Current opinion in genetics &*  
1150 *development* **46**, 37-43, doi:10.1016/j.gde.2017.06.012 (2017).
- 1151 4 Wu, G. & Scholer, H. R. Lineage Segregation in the Totipotent Embryo. *Current topics in*  
1152 *developmental biology* **117**, 301-317, doi:10.1016/bs.ctdb.2015.10.014 (2016).
- 1153 5 Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random  
1154 monoallelic gene expression in mammalian cells. *Science* **343**, 193-196,  
1155 doi:10.1126/science.1245316 (2014).
- 1156 6 Goolam, M. *et al.* Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse  
1157 Embryos. *Cell* **165**, 61-74, doi:10.1016/j.cell.2016.01.047 (2016).
- 1158 7 Guo, G. *et al.* Resolution of cell fate decisions revealed by single-cell gene expression analysis  
1159 from zygote to blastocyst. *Dev Cell* **18**, 675-685, doi:10.1016/j.devcel.2010.02.012 (2010).
- 1160 8 Blakeley, P. *et al.* Defining the three cell lineages of the human blastocyst by single-cell RNA-seq.  
1161 *Development* **142**, 3613, doi:10.1242/dev.131235 (2015).
- 1162 9 Guo, F. *et al.* Single-cell multi-omics sequencing of mouse early embryos and embryonic stem  
1163 cells. *Cell research* **27**, 967-988, doi:10.1038/cr.2017.82 (2017).
- 1164 10 Li, L. *et al.* Single-cell multi-omics sequencing of human early embryos. *Nat Cell Biol* **20**, 847-858,  
1165 doi:10.1038/s41556-018-0123-2 (2018).
- 1166 11 Petropoulos, S. *et al.* Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in  
1167 Human Preimplantation Embryos. *Cell* **165**, 1012-1026, doi:10.1016/j.cell.2016.03.023 (2016).
- 1168 12 Gurdon, J. B., Elsdale, T. R. & Fischberg, M. Sexually mature individuals of *Xenopus laevis* from  
1169 the transplantation of single somatic nuclei. *Nature* **182**, 64-65, doi:10.1038/182064a0 (1958).
- 1170 13 Benchetrit, H. *et al.* Extensive Nuclear Reprogramming Underlies Lineage Conversion into  
1171 Functional Trophectoderm Stem-like Cells. *Cell stem cell* **17**, 543-556,  
1172 doi:10.1016/j.stem.2015.08.006 (2015).
- 1173 14 Buganim, Y., Faddah, D. A. & Jaenisch, R. Mechanisms and models of somatic cell  
1174 reprogramming. *Nat Rev Genet* **14**, 427-439, doi:10.1038/nrg3473 (2013).
- 1175 15 Buganim, Y. *et al.* The developmental potential of iPSCs is greatly influenced by reprogramming  
1176 factor selection. *Cell stem cell* **15**, 295-309, doi:10.1016/j.stem.2014.07.003 (2014).
- 1177 16 Kubaczka, C. *et al.* Direct Induction of Trophectoderm Stem Cells from Murine Fibroblasts. *Cell stem*  
1178 *cell* **17**, 557-568, doi:10.1016/j.stem.2015.08.005 (2015).
- 1179 17 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and  
1180 adult fibroblast cultures by defined factors. *Cell* **126**, 663-676, doi:10.1016/j.cell.2006.07.024  
1181 (2006).
- 1182 18 Theunissen, T. W. *et al.* Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell*  
1183 *stem cell* **19**, 502-515, doi:10.1016/j.stem.2016.06.011 (2016).
- 1184 19 Theunissen, T. W. & Jaenisch, R. Molecular control of induced pluripotency. *Cell stem cell* **14**,  
1185 720-734, doi:10.1016/j.stem.2014.05.002 (2014).
- 1186 20 Wilmut, I. *et al.* Somatic cell nuclear transfer. *Nature* **419**, 583-586, doi:10.1038/nature01079  
1187 (2002).



- 1188 21 Wakayama, T. *et al.* Differentiation of embryonic stem cell lines generated from adult somatic  
1189 cells by nuclear transfer. *Science* **292**, 740-743, doi:10.1126/science.1059399 (2001).
- 1190 22 Benevento, M. *et al.* Proteome adaptation in cell reprogramming proceeds via distinct  
1191 transcriptional networks. *Nat Commun* **5**, 5613, doi:10.1038/ncomms6613 (2014).
- 1192 23 Buganim, Y. *et al.* Single-cell expression analyses during cellular reprogramming reveal an early  
1193 stochastic and a late hierarchic phase. *Cell* **150**, 1209-1222, doi:10.1016/j.cell.2012.08.023  
1194 (2012).
- 1195 24 Hansson, J. *et al.* Highly coordinated proteome dynamics during reprogramming of somatic cells  
1196 to pluripotency. *Cell Rep* **2**, 1579-1592, doi:10.1016/j.celrep.2012.10.014 (2012).
- 1197 25 Hussein, S. M. *et al.* Copy number variation and selection during reprogramming to  
1198 pluripotency. *Nature* **471**, 58-62, doi:10.1038/nature09871 (2011).
- 1199 26 Ji, J. *et al.* Elevated coding mutation rate during the reprogramming of human somatic cells into  
1200 induced pluripotent stem cells. *Stem Cells* **30**, 435-440, doi:10.1002/stem.1011 (2012).
- 1201 27 Lee, D. S. *et al.* An epigenomic roadmap to induced pluripotency reveals DNA methylation as a  
1202 reprogramming modulator. *Nat Commun* **5**, 5619, doi:10.1038/ncomms6619 (2014).
- 1203 28 Li, Z. *et al.* Mouse SCNT ESCs have lower somatic mutation load than syngeneic iPSCs. *Stem Cell*  
1204 *Reports* **2**, 399-405, doi:10.1016/j.stemcr.2014.02.005 (2014).
- 1205 29 Lujan, E. *et al.* Early reprogramming regulators identified by prospective isolation and mass  
1206 cytometry. *Nature* **521**, 352-356, doi:10.1038/nature14274 (2015).
- 1207 30 Mathieu, J. & Ruohola-Baker, H. Metabolic remodeling during the loss and acquisition of  
1208 pluripotency. *Development* **144**, 541-551, doi:10.1242/dev.128389 (2017).
- 1209 31 Papp, B. & Plath, K. Reprogramming to pluripotency: stepwise resetting of the epigenetic  
1210 landscape. *Cell research* **21**, 486-501, doi:10.1038/cr.2011.28 (2011).
- 1211 32 Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPSCs. *Cell* **151**,  
1212 1617-1632, doi:10.1016/j.cell.2012.11.039 (2012).
- 1213 33 Schiebinger, G. *et al.* Optimal-Transport Analysis of Single-Cell Gene Expression Identifies  
1214 Developmental Trajectories in Reprogramming. *Cell* **176**, 1517, doi:10.1016/j.cell.2019.02.026  
1215 (2019).
- 1216 34 Tran, K. A. *et al.* Defining Reprogramming Checkpoints from Single-Cell Analyses of Induced  
1217 Pluripotency. *Cell Rep* **27**, 1726-1741 e1725, doi:10.1016/j.celrep.2019.04.056 (2019).
- 1218 35 Guo, L. *et al.* Resolving Cell Fate Decisions during Somatic Cell Reprogramming by Single-Cell  
1219 RNA-Seq. *Mol Cell* **73**, 815-829 e817, doi:10.1016/j.molcel.2019.01.042 (2019).
- 1220 36 Li, D. *et al.* Chromatin Accessibility Dynamics during iPSC Reprogramming. *Cell stem cell* **21**, 819-  
1221 833 e816, doi:10.1016/j.stem.2017.10.012 (2017).
- 1222 37 Krijger, P. H. *et al.* Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell  
1223 Reprogramming. *Cell stem cell* **18**, 597-610, doi:10.1016/j.stem.2016.01.007 (2016).
- 1224 38 Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-  
1225 epithelial transition in the initiation of somatic cell reprogramming. *Cell stem cell* **7**, 64-77,  
1226 doi:10.1016/j.stem.2010.04.015 (2010).
- 1227 39 Maekawa, M. *et al.* Direct reprogramming of somatic cells is promoted by maternal transcription  
1228 factor Glis1. *Nature* **474**, 225-229, doi:10.1038/nature10106 (2011).
- 1229 40 Cao, S. *et al.* Chromatin Accessibility Dynamics during Chemical Induction of Pluripotency. *Cell*  
1230 *stem cell* **22**, 529-542 e525, doi:10.1016/j.stem.2018.03.005 (2018).
- 1231 41 Liu, J. *et al.* The oncogene c-Jun impedes somatic cell reprogramming. *Nat Cell Biol* **17**, 856-867,  
1232 doi:10.1038/ncb3193 (2015).
- 1233 42 Benchetrit, H. *et al.* Direct Induction of the Three Pre-implantation Blastocyst Cell Types from  
1234 Fibroblasts. *Cell stem cell* **24**, 983-994 e987, doi:10.1016/j.stem.2019.03.018 (2019).

- 1235 43 Schiebinger, G. *et al.* Optimal-Transport Analysis of Single-Cell Gene Expression Identifies  
1236 Developmental Trajectories in Reprogramming. *Cell* **176**, 928-943 e922,  
1237 doi:10.1016/j.cell.2019.01.006 (2019).
- 1238 44 Francesconi, M. *et al.* Single cell RNA-seq identifies the origins of heterogeneity in efficient cell  
1239 transdifferentiation and reprogramming. *Elife* **8**, doi:10.7554/eLife.41627 (2019).
- 1240 45 Guo, L. *et al.* Resolution of Reprogramming Transition States by Single Cell RNA-Sequencing.  
1241 *bioRxiv*, 182535, doi:10.1101/182535 (2017).
- 1242 46 Pauklin, S., Pedersen, R. A. & Vallier, L. Mouse pluripotent stem cells at a glance. *J Cell Sci* **124**,  
1243 3727-3732, doi:10.1242/jcs.074120 (2011).
- 1244 47 Schroeder, D. I. *et al.* Early Developmental and Evolutionary Origins of Gene Body DNA  
1245 Methylation Patterns in Mammalian Placentas. *PLoS Genet* **11**, e1005442,  
1246 doi:10.1371/journal.pgen.1005442 (2015).
- 1247 48 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat*  
1248 *Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).
- 1249 49 Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency  
1250 reprogramming factors' initial engagement with the genome. *Cell* **151**, 994-1004,  
1251 doi:10.1016/j.cell.2012.09.045 (2012).
- 1252 50 Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate  
1253 reprogramming. *Cell* **161**, 555-568, doi:10.1016/j.cell.2015.03.017 (2015).
- 1254 51 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying  
1255 Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 21-21 29 29,  
1256 doi:10.1002/0471142727.mb2129s109 (2015).
- 1257 52 Creighton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts  
1258 developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:10.1073/pnas.1016071107  
1259 (2010).
- 1260 53 Pekowska, A., Benoukraf, T., Ferrier, P. & Spicuglia, S. A unique H3K4me2 profile marks tissue-  
1261 specific gene regulation. *Genome Res* **20**, 1493-1502, doi:10.1101/gr.109389.110 (2010).
- 1262 54 Tanigawa, S. *et al.* Jun dimerization protein 2 is a critical component of the Nrf2/MafK complex  
1263 regulating the response to ROS homeostasis. *Cell Death Dis* **4**, e921, doi:10.1038/cddis.2013.448  
1264 (2013).
- 1265 55 Dohi, Y. *et al.* Bach1 inhibits oxidative stress-induced cellular senescence by impeding p53  
1266 function on chromatin. *Nat Struct Mol Biol* **15**, 1246-1254, doi:10.1038/nsmb.1516 (2008).
- 1267 56 Fischer, B. *et al.* E-proteins orchestrate the progression of neural stem cell differentiation in the  
1268 postnatal forebrain. *Neural Dev* **9**, 23, doi:10.1186/1749-8104-9-23 (2014).
- 1269 57 Chanda, S. *et al.* Generation of induced neuronal cells by the single reprogramming factor  
1270 ASCL1. *Stem Cell Reports* **3**, 282-296, doi:10.1016/j.stemcr.2014.05.020 (2014).
- 1271 58 Ishihara, K., Oshimura, M. & Nakao, M. CTCF-dependent chromatin insulator is linked to  
1272 epigenetic remodeling. *Mol Cell* **23**, 733-742, doi:10.1016/j.molcel.2006.08.008 (2006).
- 1273 59 Debryne, D. N. *et al.* BORIS promotes chromatin regulatory interactions in treatment-resistant  
1274 cancer cells. *Nature* **572**, 676-680, doi:10.1038/s41586-019-1472-0 (2019).
- 1275 60 Qin, J. *et al.* The polycomb group protein L3mbtl2 assembles an atypical PRC1-family complex  
1276 that is essential in pluripotent stem cells and early development. *Cell stem cell* **11**, 319-332,  
1277 doi:10.1016/j.stem.2012.06.002 (2012).
- 1278 61 Ehara, H. *et al.* Structural insight into nucleosome transcription by RNA polymerase II with  
1279 elongation factors. *Science* **363**, 744-747, doi:10.1126/science.aav8912 (2019).
- 1280 62 Deng, C. *et al.* USF1 and hSET1A mediated epigenetic modifications regulate lineage  
1281 differentiation and HoxB4 transcription. *PLoS Genet* **9**, e1003524,  
1282 doi:10.1371/journal.pgen.1003524 (2013).

- 1283 63 Pentland, I. *et al.* Disruption of CTCF-YY1-dependent looping of the human papillomavirus  
1284 genome activates differentiation-induced viral oncogene transcription. *PLoS Biol* **16**, e2005752,  
1285 doi:10.1371/journal.pbio.2005752 (2018).
- 1286 64 Choe, C., Chen, N. & Sawadogo, M. Decreased tumorigenicity of c-Myc-transformed fibroblasts  
1287 expressing active USF2. *Exp Cell Res* **302**, 1-10, doi:10.1016/j.yexcr.2004.08.013 (2005).
- 1288 65 Nicolas, G. *et al.* Lack of hepcidin gene expression and severe tissue iron overload in upstream  
1289 stimulatory factor 2 (USF2) knockout mice. *Proc Natl Acad Sci U S A* **98**, 8780-8785,  
1290 doi:10.1073/pnas.151179498 (2001).
- 1291 66 Verrijt, C. E., Kroos, M. J., Huijskes-Heins, M. I., van Eijk, H. G. & van Dijk, J. P. Non-transferrin  
1292 iron uptake by trophoblast cells in culture. Significance of a NADH-dependent ferrireductase.  
1293 *Placenta* **19**, 525-530, doi:10.1016/s0143-4004(98)91046-3 (1998).
- 1294 67 Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome  
1295 Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922,  
1296 doi:10.1016/j.cell.2017.05.004 (2017).
- 1297 68 Lee, B. K. *et al.* Super-enhancer-guided mapping of regulatory networks controlling mouse  
1298 trophoblast stem cells. *Nat Commun* **10**, 4749, doi:10.1038/s41467-019-12720-6 (2019).
- 1299 69 Tunster, S. J., Watson, E. D., Fowden, A. L. & Burton, G. J. Placental glycogen stores and fetal  
1300 growth: insights from genetic mouse models. *Reproduction* **159**, R213-R235, doi:10.1530/REP-  
1301 20-0007 (2020).
- 1302 70 Wang, Z. *et al.* NOD1 and NOD2 control the invasiveness of trophoblast cells via the MAPK/p38  
1303 signaling pathway in human first-trimester pregnancy. *Placenta* **36**, 652-660,  
1304 doi:10.1016/j.placenta.2015.03.004 (2015).
- 1305 71 Pathmaperuma, A. N. *et al.* Fatty acids alter glycerolipid metabolism and induce lipid droplet  
1306 formation, syncytialisation and cytokine production in human trophoblasts with minimal glucose  
1307 effect or interaction. *Placenta* **31**, 230-239, doi:10.1016/j.placenta.2009.12.013 (2010).
- 1308 72 Wamaitha, S. E. *et al.* IGF1-mediated human embryonic stem cell self-renewal recapitulates the  
1309 embryonic niche. *Nat Commun* **11**, 764, doi:10.1038/s41467-020-14629-x (2020).
- 1310 73 Manso, A. M., Kang, S. M. & Ross, R. S. Integrins, focal adhesions, and cardiac fibroblasts. *J*  
1311 *Investig Med* **57**, 856-860, doi:10.2310/JIM.0b013e3181c5e61f (2009).
- 1312 74 Stratton, R. & Shiwen, X. Role of prostaglandins in fibroblast activation and fibrosis. *J Cell*  
1313 *Commun Signal* **4**, 75-77, doi:10.1007/s12079-010-0089-8 (2010).
- 1314 75 Kaczynski, J., Cook, T. & Urrutia, R. Sp1- and Kruppel-like transcription factors. *Genome Biol* **4**,  
1315 206, doi:10.1186/gb-2003-4-2-206 (2003).
- 1316 76 Oldfield, A. J. *et al.* NF-Y controls fidelity of transcription initiation at gene promoters through  
1317 maintenance of the nucleosome-depleted region. *Nat Commun* **10**, 3072, doi:10.1038/s41467-  
1318 019-10905-7 (2019).
- 1319 77 Zwilling, S., Annweiler, A. & Wirth, T. The POU domains of the Oct1 and Oct2 transcription  
1320 factors mediate specific interaction with TBP. *Nucleic Acids Res* **22**, 1655-1662,  
1321 doi:10.1093/nar/22.9.1655 (1994).
- 1322 78 Cornacchia, D. *et al.* Lipid Deprivation Induces a Stable, Naive-to-Primed Intermediate State of  
1323 Pluripotency in Human PSCs. *Cell stem cell* **25**, 120-136 e110, doi:10.1016/j.stem.2019.05.001  
1324 (2019).
- 1325 79 Kaneko, K. *et al.* Neuronal Rap1 Regulates Energy Balance, Glucose Homeostasis, and Leptin  
1326 Actions. *Cell Rep* **16**, 3003-3015, doi:10.1016/j.celrep.2016.08.039 (2016).
- 1327 80 Rosso, S. B. & Inestrosa, N. C. WNT signaling in neuronal maturation and synaptogenesis. *Front*  
1328 *Cell Neurosci* **7**, 103, doi:10.3389/fncel.2013.00103 (2013).

- 1329 81 Hsu, J. *et al.* E2F4 regulates transcriptional activation in mouse embryonic stem cells  
1330 independently of the RB family. *Nat Commun* **10**, 2939, doi:10.1038/s41467-019-10901-x  
1331 (2019).
- 1332 82 Rugg-Gunn, P. J. Epigenetic features of the mouse trophoblast. *Reprod Biomed Online* **25**, 21-30,  
1333 doi:10.1016/j.rbmo.2012.01.012 (2012).
- 1334 83 Levine, A. J., Ting, D. T. & Greenbaum, B. D. P53 and the defenses against genome instability  
1335 caused by transposons and repetitive elements. *Bioessays* **38**, 508-513,  
1336 doi:10.1002/bies.201600031 (2016).
- 1337 84 Padeken, J., Zeller, P. & Gasser, S. M. Repeat DNA in genome organization and stability. *Current*  
1338 *opinion in genetics & development* **31**, 12-19, doi:10.1016/j.gde.2015.03.009 (2015).
- 1339 85 Kuzyk, A. & Mai, S. c-MYC-induced genomic instability. *Cold Spring Harb Perspect Med* **4**,  
1340 a014373, doi:10.1101/cshperspect.a014373 (2014).
- 1341 86 Buganim, Y. & Jaenisch, R. Transdifferentiation by defined factors as a powerful research tool to  
1342 address basic biological questions. *Cell Cycle* **11**, 4485-4486, doi:10.4161/cc.22665 (2012).
- 1343 87 Wernig, M. *et al.* A drug-inducible transgenic system for direct reprogramming of multiple  
1344 somatic cell types. *Nat Biotechnol* **26**, 916-924, doi:10.1038/nbt1483 (2008).
- 1345 88 Buganim, Y. *et al.* Mutant p53 protects cells from 12-O-tetradecanoylphorbol-13-acetate-  
1346 induced death by attenuating activating transcription factor 3 induction. *Cancer Res* **66**, 10750-  
1347 10759, doi:10.1158/0008-5472.CAN-06-0916 (2006).
- 1348 89 Boyle, P. *et al.* Gel-free multiplexed reduced representation bisulfite sequencing for large-scale  
1349 DNA methylation profiling. *Genome Biol* **13**, R92, doi:10.1186/gb-2012-13-10-r92 (2012).
- 1350 90 Mendenhall, E. M. *et al.* Locus-specific editing of histone modifications at endogenous  
1351 enhancers. *Nat Biotechnol* **31**, 1133-1136, doi:10.1038/nbt.2701 (2013).
- 1352 91 Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-  
1353 committed cells. *Nature* **448**, 553-560, doi:10.1038/nature06008 (2007).
- 1354 92 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native  
1355 chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins  
1356 and nucleosome position. *Nat Methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).
- 1357 93 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory  
1358 requirements. *Nat Methods* **12**, 357-360, doi:10.1038/nmeth.3317 (2015).
- 1359 94 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for  
1360 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930,  
1361 doi:10.1093/bioinformatics/btt656 (2014).
- 1362 95 Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the  
1363 Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719-720,  
1364 doi:10.1093/bioinformatics/btm563 (2008).
- 1365 96 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016  
1366 update. *Nucleic Acids Res* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).
- 1367 97 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological  
1368 themes among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).
- 1369 98 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq.  
1370 *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).
- 1371 99 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments  
1372 with TopHat and Cufflinks. *Nat Protoc* **7**, 562-578, doi:10.1038/nprot.2012.016 (2012).
- 1373 100 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821,  
1374 doi:10.1016/j.cell.2019.05.031 (2019).
- 1375 101 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079,  
1376 doi:10.1093/bioinformatics/btp352 (2009).

1377 102 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis.  
1378 *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).

1379 103 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for  
1380 exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191, doi:10.1093/nar/gku365  
1381 (2014).

1382 104 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.  
1383 *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

1384 105 Klambauer, G. *et al.* cn.MOPS: mixture of Poissons for discovering copy number variations in  
1385 next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res* **40**, e69,  
1386 doi:10.1093/nar/gks003 (2012).

1387 106 Buganim, Y. *et al.* Direct reprogramming of fibroblasts into embryonic Sertoli-like cells by  
1388 defined factors. *Cell stem cell* **11**, 373-386, doi:10.1016/j.stem.2012.07.019 (2012).

1389 107 Buganim, Y. *et al.* Transcriptional activity of ATF3 in the stromal compartment of tumors  
1390 promotes cancer progression. *Carcinogenesis* **32**, 1749-1757, doi:10.1093/carcin/bgr203 (2011).

1391 108 Niakan, K. K., Schrode, N., Cho, L. T. & Hadjantonakis, A. K. Derivation of extraembryonic  
1392 endoderm stem (XEN) cells from mouse embryos and embryonic stem cells. *Nat Protoc* **8**, 1028-  
1393 1041, doi:10.1038/nprot.2013.049 (2013).

1394