Cooperative genetic networks drive a mammalian cell state transition

Andreas Lackner[#1], Robert Sehlke[#2], Marius Garmhausen[#2], Giuliano Giuseppe Stirparo[#3], Michelle Huth[$1], Fabian Titz-Teixeira[$2], Petra van der Lelij[1], Julia Ramesmayer[1], Henry Fabian Thomas[1], Meryem Ralser[3], Laura Santini[1], Mihail Sarov[4], A. Francis Stewart[4,5], Austin Smith[3], Andreas Beyer*[2], Martin Leeb*[1]

[1] Max Perutz Laboratories Vienna, University of Vienna, Vienna Biocenter, Dr Bohr Gasse 9, 1030 Vienna

[2] CECAD, Faculty of Mathematics and Natural Sciences, Medical Faculty Cologne, University of Cologne, Joseph-Stelzmann-Str. 26 50931 Köln, Germany

[3] Wellcome - MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge CB2 0AW, UK

[4] Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstraße 108
01307 Dresden, Germany

[5] Biotechnology Center, Center for Molecular and Cellular Bioengineering, Technische Universität Dresden, Tatzberg 47, 01307 Dresden, Germany

[6] Center for Molecular Medicine (CMMC), University of Cologne, Cologne, Germany.

# Equal contribution

$ Equal contribution

* Corresponding authors

Key words:

Haploid ES Cells, Systems Biology, Knock Out, Pluripotency, Naïve to Formative Transition, Epiblast, Signalling, Exit from Naïve Pluripotency, Gene Regulatory Networks

Abstract

In the mammalian embryo, epiblast cells must exit their naïve state and acquire formative pluripotency. This cell state transition is recapitulated by mouse embryonic stem cells (ESCs), which undergo pluripotency progression in defined conditions *in vitro*. However, our understanding of the molecular cascades and gene-networks involved in the exit from naïve pluripotency remains fragmented. Here we employed a combination of genetic screens in haploid ESCs, CRISPR/Cas9 gene disruption, large-scale transcriptomics and computational systems-biology to delineate the regulatory circuits governing naïve state exit. Transcriptome profiles for 73 knockout ESC lines predominantly manifest delays on the trajectory from naive to formative epiblast. We find that gene networks operative in ESCs are active during transition from pre- to post-implantation epiblast *in utero*. We identified 374 naïve-associated genes tightly connected to epiblast state and largely conserved in human ESCs and primate embryos. Integrated analysis of mutant transcriptomes revealed funneling of multiple gene activities into discrete regulatory modules. Finally, we delineate how intersections with signaling pathways direct this pivotal mammalian cell state transition.

Introduction

Mouse embryonic stem cells (ESCs) can self-renew in defined conditions in a ground state of naïve pluripotency [1]. The ESC exit from naïve pluripotency provides an amenable experimental system for dissection of a cell fate decision paradigm [2, 3]. Naïve pluripotency is under control of a gene regulatory network (GRN) containing the core pluripotency transcription factors (TFs) *Pou5f1*, *Sox2* and naïve-specific TFs like *Nanog*, *Esrrb*, *Klf4* and others [4-6]. In defined cell culture conditions that include inhibitors against Mek1/2 (PD0325901) and Gsk3 (CHIR990201, CH; collectively termed '2i'), ESCs can be homogenously maintained in the naïve state [7]. Within 24 to 36 hours after withdrawal of 2i, ESCs transit into formative pluripotency, entirely losing naïve identity [2]. During this transition, the naïve GRN is extinguished and expression of formative factors like *Otx2*, *Pou3f1*, *Dnmt3a/b* and *Fgf5* is initiated. A similar transition is evident during peri-implantation development, where the TF-network maintaining naïve pluripotency dissolves between embryonic day (E) 4.5 and E5.5 [8-10].

The speed of the naïve to formative GRN transition is notable because: i) the cell cycle is around 12 hours long; ii) all factors that are required to establish and maintain naïve pluripotency are expressed robustly in naïve cells; and (iii) the naïve pluripotency network is recursively self-reinforcing. The rapid dissolution of naïve pluripotency implies the existence of circuit-breaking mechanisms. In recent years, we and others have identified various factors promoting ESC differentiation using screens in haploid and diploid ES cells [11-15].

21 Robust assays employing ESCs expressing a Rex1-promoter driven destabilised GFP reporter
22 ESC line (Rex1::GFPd2) enable the dissection of the exit from naïve pluripotency in high
23 resolution [2, 16]. Rex1-GFP downregulation is initiated within 24h after 2i withdrawal (N24) and
24 completed after 48h (N48). Nevertheless, the exact nature, mechanistic underpinnings and
25 sequence of events during exit from naïve pluripotency remain only partially understood. In
26 particular, we lack insight into how the different molecular components of the system co-
27 operate to elicit proper cell fate transition.

28 Here, we have driven a Rex1-GFP reporter screen to saturation, thus providing an extensive
29 list of genes and pathways involved in the exit from naïve pluripotency. We utilized this
30 information in a systems biology approach to explore regulatory principles of the exit from
31 naïve pluripotency. To evaluate dependencies and causal relationships within the
32 pluripotency and differentiation circuitries, we probed the response of the differentiation
33 program to a comprehensive series of exit factor gene knockouts. Through computational
34 integration of molecular profiling data with regulatory networks and *in vivo* GRN-trajectories,
35 we expose the regulatory foundations of a cell fate choice paradigm at a pivotal junction in
36 early mammalian development.

Results

Haploid ES cell saturation screen

Haploid ES cells are an efficient platform for insertional mutagenesis-based screens [11, 17, 18]. We previously reported a medium-scale screen comprising approximately $5x10^4$ mutagenic events to identify factors regulating the exit from naïve pluripotency [12]. We have now driven this approach to saturation by assaying approximately 1.2 million mutations in receptive genomic regions that cause delays in Rex1 downregulation in two independent Rex1-reporter cell lines (Figure 1ab), utilizing three different mutagenic vectors in 35 independent screens (Figure 1cd).

Stringent filtering resulted in a candidate list comprising 489 genes (Supplementary Table 1). Reassuringly, the known exit from naïve pluripotency regulators *Tcf7l1, Fgfr2, Jarid2* and *Mapk1* (*Erk2*) [19] were among the highest ranked (Supplementary Figure 1a). The candidate hit genes were enriched for processes involved in transcription regulation, epigenetics and signalling-related functions (Supplementary Figures 1bc; Supplementary Table 1), as well as RNA binding functions in line with emerging evidence of RNA regulatory mechanisms in cell fate control [20].

These screens generated a candidate inventory of the machinery that mediates exit from naïve pluripotency. Many identified genes are not specific to pluripotency and have functions in common pathways and processes, implying that the exit from naïve pluripotency utilizes

58 widely expressed cellular machinery. Therefore, mechanisms mediating ESC transition might
59 also be utilised in other differentiation processes.

60 **Establishment of a mutant ESC library for systematic transcriptional profiling**

61 To characterise deficiencies in naïve exit, we generated KO ESC lines deficient for 73 selected
62 genes, comprising top ranked genes from the mutagenic screen together with components
63 from pathways and protein complexes for which multiple members were recovered, even if
64 just below the cut-off threshold (e.g. the Paf complex member *Leo1*, the mTORC1 regulator
65 *Tsc2* and the NMD component *Smg6*), and *Mbd3*, *Zfp281* and *L3mbtl3* as known players in
66 the exit from naïve pluripotency [14]. Three control genes were included that are either not
67 expressed in ES cells (Nestin), expected to be neutral (Hprt), or whose ablation was expected
68 to accelerate differentiation (cMyc). Paired gRNAs were used to disrupt target genes in a
69 diploid Rex1::GFPd2 reporter ES cell line carrying a Cas9-transgene (henceforth termed RC9
70 cells) (Figure 1a). Following an efficient parallelized approach, we established passage
71 matched and isogenic cell lines, thus maximising comparability (Supplementary Figure 1de).
72
73 Full protein deficiency was validated for 14 KOs (*Eed, Suz12, Jarid2, Kdm6a, Smg5, Smg6,*
74 *Smg7, Tsc2, Pten, Raf1, Tcf7l1, Leo1, Nmt1,* and *Csnk1a1*), while reduced protein expression
75 in concordance with a heterozygote phenotype was confirmed for *Mapk1* (Supplementary
76 Figure 1f). For five further genes (*Alg13, Dido1, Msi2, Etv5, Jmjd1c*) we confirmed absence of
77 the corresponding transcripts or specific out of frame deletion of an exon by RT-qPCR or
78 Sanger sequencing of RT-PCR products. Successful rescue experiments using 3xflag-tagged
79 transgenes for six genes (*Rbpj, Etv5, Fgfr1, Jarid2, Mbd3,* and *Tcf7l1*) established causality
80 between the observed genotype and phenotype (Supplementary Figures 1gh). Thus, all tested

81 knockouts showed the expected impact on RNA or protein expression. However, we cannot
82 exclude the possibility of hypomorphic phenotypes in some cases.
83
84 ESC differentiation behaviour is highly dependent on cell density and timing of medium
85 changes. To enable robust comparison of the differentiation of multiple KO ESC lines in
86 parallel, we performed differentiations batch wise in duplicates, always including WT ESCs
87 and negative controls across seven experiments. At N24 we assayed the differentiation status
88 by FACS analysis (Figure 1e) and extracted RNA for transcriptome analysis.

89 **The exit machinery is already poised in 2i**

90 Batch corrected RNA-seq data comprising 14 replicates of WT ES cells indicated 3058
91 differentially expressed genes (DEG) between 2i and N24 (H0: $|FC| < 1.5$, FDR ≤ 0.05; Figure
92 1f, Supplementary Figure 1i,j and Supplementary Table 1). Interestingly, most of the 489
93 genes identified in the haploid screens including the 73 genes selected for KO did not change
94 significantly in transcript expression between 2i and N24 and were not present in the list of
95 DEG (Figures 1f and S1K), with only *Jarid2*, *L3mbtl3*, *Pten* and *Zfp423* showing upregulation at
96 N24 (FC ≥ 2; adj.p. ≤ 0.05). This implies that the exit machinery is already embedded in the
97 ground state and ESCs are poised for rapid decommissioning of naïve identity and entry into
98 differentiation [19].
99
100 To facilitate interrogation of the KO gene expression datasets, we developed an interactive
101 online tool (GENEES - Genetic Network Explorer for Embryonic Stem Cells –
102 http://shiny.cecad.uni-koeln.de:3838/stemcell_network/; User: stemcells; Pw:
103 S3m7KmIfHr9).

104

105 Using t-Distributed Stochastic Neighbour Embedding (t-SNE), we visualized similarities
106 between KOs based on expression of the DEG in 2i and at N24. We observed clustering of
107 members of the same complex or pathway: *Eed-Suz12* (PRC2), *Ptpn11-Raf1-Fgfr1-Etv5*
108 (Fgf/ERK), *Smg5-Smg6-Smg7* (NMD; nonsense mediated decay), *Mta3-Mbd3* (NuRD), *Pten-*
109 *Tsc2* (mTorc1 signaling) (Figure 1gh). The transcription profiles of the KO ESCs clustered by
110 culture condition (2i and N24) and to a lesser extent by genotype (Supplementary Figure 1i).
111 This is consistent with the phenotypic observation that despite manifesting differentiation
112 delays at N24, all of the KO ESCs ultimately departed from the naïve state during longer
113 differentiation timecourses, as measured by loss of Rex1GFP. Furthermore, even KOs that
114 showed extensive Rex1GFP downregulation delays at N24 displayed transcription profiles
115 that were globally adjusted towards differentiation. Therefore, the knockout of a single gene
116 is not sufficient to permanently block exit from naïve pluripotency in culture, in accord with
117 the finding that ternary depletion of *Tcf7l1*, *Etv5* and *Rbpj* is required for sustained self-
118 renewal in the absence of 2i or LIF [21].

119

120 **The exceptional role of *Csnk1a1* and the involvement of compensatory mechanisms**

121 The *Csnk1a1* KO is an exception to this generalization, clustering with 2i samples at N24
122 (Supplementary Figure 1i). Csnk1a1 is a serine threonine kinase and a component of the beta-
123 catenin destruction complex. Although KO of another destruction complex member *Apc*, or
124 of the downstream repressor *Tcf7l1* resulted in the upregulation of similar gene-sets
125 (Supplementary Figure 1l), we observed stronger differentiation defects and larger amplitude
126 of gene deregulation in two independently derived *Csnk1a1* mutants. However, these
127 mutants also exhibited markedly reduced population doubling rates, which was not shown by

128 *Apc* or *Tcf7l1* mutants. Upon continuous culture in 2i medium (~5 passages), proliferation was
129 restored in *Csnk1a1* KO cells and differentiation potential was regained, suggesting up-
130 regulation of compensatory mechanisms and a likely effect of proliferation rate on
131 differentiation kinetics. Csnk1a1 can be chemically inactivated using Epiblastin A, [22]. We found
132 that Epiblastin A partially blocked exit from naïve pluripotency, without affecting proliferation
133 within the duration of the assay (Supplementary Figure 1m). A second case of phenotype
134 adaptation was observed with Pum1. *Pum1* KOs showed pronounced differentiation delays
135 during early passages (Supplementary Figure 1n), as also seen for acute Pum1 depletion by
136 siRNA and in previously generated CRISPR KO ESCs [12]. However, the phenotype was lost in
137 later passages and *Pum1* KO cells showed WT like expression-levels of naïve marker genes
138 (Figures 2a and S1a).
139
140 **Robust feedback wiring in the naïve TF network**
141 Interestingly, the transcriptome data revealed that exit factors do not, in general, reduce
142 naïve transcription factor (TF) expression in the ground state (Supplementary Figure 2a). *Rbpj*
143 and *Trim71* KO resulted in moderate increases in *Klf4*, while the aforementioned *Csnk1a1* KO
144 ESCs showed limited up-regulation of both *Klf4* and *Tbx3* in 2i. Other KOs had only minor
145 effects on the naïve network in 2i. These data are consistent with robust feedback wiring in
146 the naïve TF network [4, 5] and neutralisation of most 'differentiation drivers' in 2i in culture [23].
147 In contrast, we observed a more extensive impact on formative markers. Several KO cell lines
148 showed lower baseline expression in 2i of *Otx2, Fgf5, Dnmt3a/b* and *Pou3f1* (Oct6)
149 (Supplementary Figure 2b). In line with recent results, we noted that depletion of several
150 Fgf/ERK components resulted in reduced *Dnmt3a/b, Pou3f1* and *Fgf5* expression in 2i [21].
151 Although Fgf/ERK signalling is effectively inhibited in 2i [7], our data suggest that either residual

152 pathway activity or potential moonlighting functions of pathway components mediate poised
153 expression of the formative pluripotency programme in 2i.
154
155 Clustering based on the expression of ten naïve pluripotency marker genes showed that the
156 downregulation of the naïve pluripotency TFs during formative differentiation is defective
157 across multiple KOs (Figure 2a). Although expression of the naïve TFs was highly correlated,
158 *Klf2* appeared to be an exception. *Klf2* downregulation was notably impaired in *Tcf7l1* KO
159 ESCs, whereas it was unaffected by several KOs, including *Jarid2,* despite a comparable extent
160 of deregulation of most other naïve marker genes. This indicates that *Klf2* expression can be
161 uncoupled from the core naïve network. Forced *Klf2* expression stabilizes self-renewal [24, 25].
162 *Klf2* depletion destabilizes mouse ESC identity [26] and increases the speed of Rex1-GFP
163 downregulation upon 2i withdrawal (Supplementary Figure 2c). The differentiation delay in
164 *Tcf7l1* KO ES cells is partially dependent on *Klf2* (Figure 2b), consistent with direct regulation
165 [27]. Conversely, *Klf2* depletion in *Jarid2* KOs did not restore differentiation timing, indicating
166 that separable gene-networks contribute to dismantling naïve pluripotency.

167 **Effects of exit gene depletion on global gene regulation**

168 We used two measures to gauge differentiation delays in KOs compared to WT cells: first, the
169 average expression levels of a set of naïve marker genes (defined throughout this manuscript
170 as the average deregulation of the naïve marker genes *Esrrb, Nanog, Tfcp2l1, Tbx3, Prdm14*
171 and *Klf4* at N24); second, the extent of global transcriptome adjustments between 2i and N24.
172 Generally, the delayed extinction of the naïve TF network in KOs at N24 was accompanied by
173 reduced gene expression change between 2i and N24 (R = -0.91; Figure 2c). However, some
174 KOs deviated from this pattern. *Alg13, Jmjd1c* and *Jarid2* mutants showed a less pronounced

175 change in naïve marker gene down-regulation than expected based on the reduced level of
176 global gene expression adjustments towards an N24 profile. *Trim71* KO cells in contrast
177 displayed modest global transcription changes but a more evident effect on naïve TF
178 expression levels. Thus, the RNA binding protein Trim71 appears to be focused on the
179 regulation of naïve pluripotency TF genes, although the ensuing exit delay is modest.
180
181 A general correlation was observed between the numbers of genes deregulated in 2i and at
182 N24 (Figure 2d). Mutants with the strongest impact on the transcriptome (e.g. *Csnk1a1, Eed,*
183 *Suz12, Zfp281 and Smg5*) showed deregulation of several thousand genes both in 2i and at
184 N24. *Fgfr1* and *Tcf7l1* KOs were two exceptions to this correlation with relatively few genes
185 deregulated in 2i, but several hundred at N24. We surmise that the effect of these two genes
186 in 2i is largely masked because they are in the same pathways as the inhibitor targets Mek1/2
187 (PD0325901) and Gsk3 (Chiron 990201) [7, 23]. Defective differentiation could simply equate
188 with the extent of overall gene deregulation. We therefore mapped the average deregulation
189 of naïve marker genes onto Figure 2d. Most KOs showing the strongest deregulation of naïve
190 marker genes correlated with large-scale gene deregulation (both in 2i and at N24). However,
191 there were several exceptions: for example, KO of *Eed*, *Suz12*, *Trim71*, and *Ctbp2* affected a
192 substantial number of genes, but caused naïve marker deregulation that was weaker than
193 expected assuming a direct correlation between the number of deregulated genes and the
194 differentiation phenotype. Vice versa, KO of *Mbd3, Cdk8* or *Pten* showed marked naïve
195 marker downregulation delays, despite a relatively mild impact on overall transcription
196 (Figures 2d and Supplementary Figure 2d).
197
198

199 **Transcriptome analysis reveals a genetic interaction between *Jmjd1c* and *Tcf7l1***

200 The putative histone 3 lysine 9 (H3K9) demethylase *Jmjd1c (Kdm3c)* KO is a case of particular
201 note. Few genes were significantly deregulated at N24 (Figure 2d), with a mild global
202 neutralisation of differentiation-induced transcriptome changes (Figure 2c). Neither the role
203 of *Jmjd1c* in pluripotency regulation, nor its mode of action are known. We used multiple
204 regression analysis to determine the similarity of the *Jmjd1c* KO to all other KO RNA-seq
205 profiles. At N24, the *Jmjd1c* KO transcriptome was most similar to the *Tcf7l1* KO profiles
206 (Supplementary Figure 2e), suggesting a potential functional connection between *Jmjd1c* and
207 *Tcf7l1*. Indeed, Jmjd1c has recently been reported as a high-confidence protein-interactor of
208 Tcf7l1 [28]. The 2i component Chiron 99021 is a specific inhibitor of Gsk3 and phenocopies
209 deletion of *Tcf7l1*, which is the downstream repressor in ESCs [29, 30]. Accordingly, Chiron 99021
210 delays the differentiation of WT ESCs, but had little additional effect on *Tcf7l1* KOs (Figure 2e).
211 *Jmjd1c* deficient ESCs did not show a discernible phenotype at N72, in line with minimal
212 deregulation of the naïve TF network. Addition of Chiron 99021 had a stronger than expected
213 effect on Jmjd1c KOs and resulted in a synthetic enhanced delay phenotype (Figure 2e),
214 suggesting a cooperative activity of *Jmjd1c* and the Wnt/Tcf7l1 axis in the exit from naïve
215 pluripotency.

216 **Relative quantification of differentiation delays *in vitro***

217 To quantify differentiation delays, we obtained RNA-seq data from a 2h-resolved WT ESC
218 differentiation time course (Figures 3a and Supplementary Figure 3a). We then compared the
219 expression patterns of specific gene-sets in a given KO at N24 to the expression of the same
220 gene-set along the WT 2i to N32 trajectory. This enabled us to 'position' each KO along the
221 trajectory and thus quantify the differentiation delay with a precision of about 4h (Figure 3b,

222 Supplementary Table 2). In a complementary approach, we used all 3058 DEG between 2i and
223 N24 as an alternative reference set. This yielded similar, but on average slightly less
224 pronounced differentiation delays (Figures 3c and Supplementary Figure 3b), supporting the
225 hypothesis that the naïve TF network is regulated in concordance with but partially
226 independently from the rest of the transcriptome during exit from naïve pluripotency.

227 **The regulatory program for pre- to post-implantation epiblast transition is preserved in**
228 **most mutants**

229 To explore the extent to which our *in vitro* data captures *in vivo* regulation, we first compared
230 global transcriptomes of WT and KO ESCs to single cell transcriptome data from the *in vivo*
231 pre- (E4.5) and post-implantation (E5.5 and E6.5) mouse epiblast [10] (Figure 4a,B). As expected
232 [8], WT and KO cells in 2i showed greater similarity (fraction of identity - FOI) to the E4.5 than
233 to the E5.5 epiblast. Cells at N24, in contrast, showed higher FOI to the E5.5 epiblast. Cells in
234 2i and at N24 showed low FOI to E6.5 epiblast, underscoring that 24h 2i withdrawal models
235 the E4.5 to E5.5 epiblast transition [2]. Interestingly, some KO ESCs showed expression features
236 in 2i with increased similarity to the *in vivo* epiblast (Figures 4c,d and Supplementary Table
237 3). Among those were KOs showing strong *in vitro* differentiation defects, such as *Zfp281,*
238 *Tsc2* and *Etv5,* but also *Trim71* and *Smg7*, which showed only a modest differentiation delay
239 in culture.
240
241 At N24 several KOs, such as *Zfp281*, *Tsc2*, *Smg5* and *Smg6*, retained strong similarity to the
242 E4.5 epiblast (Figures 4c,e and Supplementary Figure 4a). Strikingly, *Zfp281* KO profiles at N24
243 showed similarity to the E4.5 epiblast on par with WT cells cultured in 2i (Figures 4c and
244 Supplementary Figure 4a), consistent with an overt differentiation delay phenotype [31].

245 Overall, there was a good correlation between the *in vitro* differentiation delay and the
246 similarity with pre-implantation epiblast cells: the N24 KO transcriptomes that were more
247 similar to the E4.5 epiblast exhibited stronger exit-delay phenotypes (Figure 4e,
248 Supplementary Figure 4b). Together, this indicates that similar transcriptional networks are
249 regulated by similar mechanisms during *in vivo* and *in vitro* transitions to formative
250 pluripotency. However, at N24 three KOs (*Jarid2*, *Tcf7l1*, and *Trp53*) showed a similarity to
251 the E4.5 epiblast which was smaller than expected based on their *strong in vitro*
252 differentiation defects (Figures 2a and 4e).
253
254 We then performed principal component analysis (PCA) for genes variably expressed in
255 embryo development, using averaged values for inner cell mass (ICM), primitive endoderm
256 (PrE) and E4.5, 5.5 and 6.5 epiblast [10]. The PCA separated the 2i and N24 transcriptomes into
257 two clusters with proximity to PrE/EPI and E5.5 respectively. Notably*, in vitro* differentiation
258 defects were reflected by closer proximity of mutant N24 profiles to the E4.5 epiblast (Figure
259 4f). Overlaying the PCA with a colour gradient indicating similarity to PrE shows that multiple
260 KOs in 2i gained similarity to PrE (Supplementary Figure 4d). Concordantly, the expression of
261 PrE-defining transcription factors *Gata4*, *Gata6*, *Sox7* and *Sox17,* was readily detectable in
262 *Eed, Suz12, Zfp423* and *Csnk1a1* KOs in 2i (Supplementary Figure 4e). A shift toward PrE-
263 identity was most pronounced for the PRC2 KOs *Eed* and *Suz12*. Notably, this group of KOs
264 (Csnk1a1, PRC2, Zfp423, which exhibited substantial *in vitro* differentiation defects while
265 showing reduced similarity to the E4.5 epiblast, also showed decreased similarity between
266 their N24 profiles and the E5.5 epiblast (Supplementary Figure 4d). We therefore suggest that
267 deregulation of the PrE-programme above a certain threshold has disengaged these KOs from
268 the normal embryonic developmental trajectory. In contrast KOs showing relatively increased

269 similarity to the E4.5 in 2i and decreased similarity to the E5.5 epiblast at N24 are deficient
270 for genes regulating the formative transition during development (e.g. *Zfp281*, *Etv5*, *Rbpj* and,
271 *Trim71*, Supplementary Figure 4d).
272

273 **Correspondence to primate embryogenesis**

274 Despite morphological and timing differences between rodent and primate peri-implantation
275 development, embryos of both orders appear to transit through similar pluripotency states [1,
276 32]. To examine this issue, we compared transcriptional profiles of our KO series and cells of
277 the macaque *in vivo* naive- and post-implantation epiblast [33]. In general, ESCs in 2i were more
278 similar than cells at N24 to the pre-implantation macaque epiblast (Figure 4g).
279 Correspondingly, N24 cells were closer to the macaque post implantation epiblast.
280 Interestingly, KOs showing higher identity with the mouse *in vivo* epiblast also displayed an
281 increased similarity with the macaque pre-implantation epiblast (Supplementary Figures
282 4f,h). We also compared our datasets to data obtained from *in vitro* capacitation during
283 human naïve to primed ESC differentiation [34]. KOs displaying strong differentiation defects in
284 mouse cell culture had higher correspondence to naïve human ESCs and, at N24, lower
285 correspondence to primed human ESCs (Figures 4h and Supplementary Figure 4g,I). These
286 observations suggest that the overall GRN-redeployment during the naïve to formative
287 transition is conserved between rodents and primates.

288 **Identification of an extended naïve pluripotency network**

289 ESC differentiation requires fine-tuned coordination between extinction of the naïve and
290 initiation of the formative transcription networks. To date, only incomplete inventories of the
291 genes that functionally define these two states have been made. These genes cannot be

292 defined based simply on differential expression between 2i and N24, because not all DEG will
293 be functionally linked to the naïve GRN. To identify those genes that show specific linkage to
294 the core naïve pluripotency network, we trained regression models to predict expression
295 changes across all KOs (in 2i and at N24) as a function of naïve marker log-fold-changes. The
296 374 genes whose expression was tightly correlated ($R^2 > 0.7$) with one or several of the seven
297 core pluripotency markers (*Nanog, Esrrb, Tbx3, Tfcp2l1, Klf4, Prdm14 and Zfp42)* were defined
298 as 'Naive-Associated Genes' (NAGs; Supplementary Table 3). The naïve pluripotency TFs not
299 represented in the defining TF-set were correctly identified as NAGs, including *Klf5, Nr0b1*
300 and *Nr5a2*. However, *Klf2* is not one of the NAGs being only weakly associated (R^2=0.53) with
301 the naïve network. This concords with our earlier observation that *Klf2* expression can be
302 uncoupled from the naïve TF network. Of further note, *Klf2* expression is barely detectable in
303 marmoset or human pre-implantation epiblast cells (Supplementary Figure 5b).
304
305 The NAG showing the highest correlation with the naïve core network is *Pdgfa*, which is
306 relatively highly expressed in ES cells (FPKM ~70). *Pdgfa* has no known role in ESC self-
307 renewal, but functions in segregation of the primitive endoderm [35]. To examine whether the
308 link between *Pdgfa* and the naïve transcription factor network is maintained *in vivo*, we
309 utilized GRAPPA, a tool to visualize single cell expression data from preimplantation
310 development [36]. Indeed, *Pdgfa* is uniformly expressed in the E4.5 epiblast (Supplementary
311 Figure 5c). Interestingly, its cognate receptor *Pdgfra* is neither expressed in ES cells nor the
312 naïve epiblast (FPKM in WT ESCs <0.5), but specifically marks the neighbouring primitive
313 endoderm at E4.5 [37].
314

315 We surveyed the expression of NAGs at the transition from naïve to post-implantation
316 pluripotency in the single cell RNA-seq datasets from E4.5, E5.5 and E6.5 epiblast cells [10]. We
317 detected a clear enrichment of NAGs within genes that separate the pre- from the post-
318 implantation epiblast in the differentiation-state resolving dimension of a principal
319 component analysis (PCA) (Figures 5a and Supplementary Figure 5d), highlighting that NAGs
320 are a set of indicators for the naïve epiblast state *in vitro* and *in vivo*. Strikingly, NAGs showed
321 a near identical fold change behaviour *in vitro* and *in vivo* during the epiblast transition from
322 E4.5 to E5.5 (Figure 5b,e,f). Notably, this effect was not observed for the top 250 differentially
323 expressed genes in ESC differentiation *in vitro*. Furthermore, NAG orthologues showed
324 significantly co-regulated expression dynamics during macaque pre- to post-implantation
325 epiblast differentiation *in vivo* (Figure 5c,e,f) and during capacitation of human ESCs *in vitro*
326 (Figures 5d,e,f), which further underscores their relevance. We thus propose that NAGs
327 constitute an integral component of the naïve transcriptional network with potential
328 functional relevance *in vivo* across mammalian species.

329 **Deregulation of signaling cascades is a hallmark of differentiation delay**

330 We next analysed the extent of deregulation of five key signalling pathways known to be
331 active in pluripotent cells: LIF/Stat3, mTOR, Wnt/β-catenin, Fgf/ERK and Notch [7, 38-42]. Changes
332 in the activity of a signalling pathway are not necessarily reflected in expression changes of
333 pathway member transcripts. Thus, in order to quantify pathway activities, we employed
334 'expression footprints'. We identified pathway specific marker gene-sets, each containing 50
335 genes, reporting pathway activity changes. These marker sets were determined using N24
336 derived KO transcription-profiles of ESC lines deficient for key signalling pathway
337 components. Thereby we defined key mTOR (affected by *Tsc2 KO*), Wnt/β-catenin signaling

338 (*Tcf7l1 KO*), Fgf/ERK (*Fgfr1* and *Ptpn11 KOs*), and Notch (*Rbpj KO*) pathway targets
339 (Supplementary Figure 6a, Supplementary Table 4). Overlaps with available Tcf7l1 and Rbpj
340 ChIP and activity profiles [27, 43] supported the reliability of this approach (Supplementary Figure
341 7E). For LIF-signaling, we compared RC9 ESCs grown in 2i in the presence and absence of LIF
342 for 24 hours. This resulted in a list of LIF-sensitive genes including known targets like *Socs3,*
343 *Gbx2, Junb, Tfcp2l1, Klf4* and *Klf5* [27, 44] (Supplementary Table 4). In summary, the expression
344 footprints present non-overlapping sets of marker genes whose expression state is indicative
345 of the activity of the respective signaling pathway.
346
347 Using the expression footprints, we then asked whether we could detect a preferential
348 deregulation of one or more of these signalling footprints in specific KOs either in 2i (for LIF)
349 or at the N24 time-point (for all other pathway-profiles) (Figure 6a, Supplementary Table 4).
350 Surprisingly, we detected mis-regulated LIF target genes in several KO ESC lines cultured in 2i
351 in the absence of LIF (Figure 6a). Activation of such a 'LIF-like'-profile in 2i was closely
352 correlated with the extent of differentiation delay observed at N24 (Supplementary Figure
353 6b). Notably, presence of LIF together with 2i before induction of differentiation slows down
354 naïve state exit (Supplementary Figure 6c) [4, 16]. *Ptpn11*, *Zfp281*, *Tsc2* and *Trim71* KO ESCs
355 showed the greatest similarity to the LIF profile in 2i; in contrast, *Tcf7l1* KO ESCs, despite
356 showing a pronounced differentiation defect, lack the 'LIF footprint' (Supplementary Figure
357 6b). Addition of a JAK inhibitor to several KOs showing a LIF-like profile resulted in no or only
358 minor amelioration of the differentiation defect. This suggests that these KOs do not directly
359 activate Jak/Stat signalling, but that a LIF-like expression profile reflects a consolidated naïve
360 network that is resilient to dismantling.
361

362 As expected, *Raf1, Fgfr2, Mapk1* and *Etv5* KO ESCs showed an Fgf-signalling footprint at N24.
363 *Mbd3*, *Mta3*, *Nsd1* and *Arid5b* KOs showed Fgf/ERK target deregulation similar to reference
364 KOs, indicating an involvement of chromatin regulators in ERK target-gene control (Figure 6a).
365 *Tsc2* deficiency lead to constitutive activity of the mTORC1 pathway, which has previously
366 been associated with an exit from naïve pluripotency phenotype [45]. Many KOs (including *Pten*)
367 exhibited deregulated Tsc2 responsive genes, suggesting that regulation of mTORC1 (or its
368 downstream effectors) is a common node to gate the exit from naïve pluripotency. *Tsc2*
369 depletion resulted in partial activation of a LIF footprint, suggesting that a component of the
370 LIF response is mediated through Akt [39, 46]. *Rbpj* and *Tcf7l1* KOs did not show a correlation
371 with each other or with Tsc*2* KO, consistent with independent mechanisms [21]. The beta-
372 catenin destruction complex member *Apc* and the previously discussed *Jmjd1c* showed the
373 expected similarities with the *Tcf7l1* profile. Interestingly, several KOs with strong phenotypes
374 showed footprints similar to *Tcf7l1* KOs at N24. This is in line with evidence that β-catenin
375 constitutes or regulates a major differentiation switch during the exit from naïve pluripotency
376 and is influenced by multiple exit KO genes [29, 47]. Genes downstream of the Notch pathway
377 were most strongly affected in a group of mutants deficient for mRNA homeostasis (*Smg5*
378 and *Smg6*) and chromatin regulators, including *Jarid2, L3mbtl3 and Mbd3*. The latter suggests
379 an interaction between the Notch pathway (*Rbpj*) with the Polycomb and Nurd complexes to
380 modulate network rewiring during the exit from naïve pluripotency. Cooperativity between
381 Rbpj and the Polycomb-associated protein L3mbtl3 has been reported in *Drosophila* and *C.*
382 *elegans* [48].
383
384 We asked to what extent the alteration of any of the five signalling pathways was predictive
385 for the strength of the differentiation defect. We found that whereas aberrant Fgf/ERK or

386 mTORC1 activity were unreliable predictors, transcriptional similarity to *Rbpj* or *Tcf7l1* KOs at
387 N24 or deregulation of LIF target genes in 2i correlated with differentiation delay
388 (Supplementary Figure 6d). These data indicate that many KOs with a delay phenotype have
389 altered activity of at least one of the five key signalling pathways. Strikingly, however, there
390 was no single signalling pathway perturbed downstream of all KOs. We observed that KOs
391 showing multiple pathway footprints are rare, but most KOs showing a delayed naïve marker
392 gene downregulation at N24 deregulate at least one specific pathway. *Trp53*, *Ncor1* and
393 *Usp9x* constitute exceptions, which show appreciable differentiation defects without
394 deregulating of any of the tested signalling cascades. We conclude that signalling pathways
395 are harnessed in parallel and that most KO genes contribute directly or indirectly to at least
396 one pathway activity that promotes the exit from naïve pluripotency.

397 **Identification of gene-networks downstream of multiple KO genes**

398 To explore common regulatory targets of multiple KO-factors, we classified deregulated genes
399 into two groups: (i)**,** deregulated by a KO in both 2i and at N24, termed 'constitutive KO
400 response'; (ii) not, or only weakly, deregulated by a KO in 2i, but significantly deregulated at
401 N24, termed 'N24 induced KO response'. Genes in both groups are affected by knockouts, but
402 genes in the second group deviate from wild-type expression only under conditions of
403 differentiation. We identified all genes that showed either type of behaviour in at least one
404 KO (Supplementary Figure 6e; Supplementary Table 5). Genes deregulated exclusively in PRC2
405 core-components (*Eed* or *Suz12* KOs) or *Csnk1a1* KOs were excluded from further analysis,
406 because the deregulation of thousands of genes in these KOs had a disproportionate impact
407 on the resulting gene-lists. In total, the constitutive and the N24 induced KO response groups
408 contained 1,837 and 720 genes, respectively. Genes belonging to the N24 induced KO

409 response group were strongly enriched for factors that are dynamically regulated during
410 normal differentiation, whereas genes of the constitutive response group were not
411 (Supplementary Figure 6f). Thus, naïve marker genes were absent from the constitutive
412 group. Together with NAGs, they were strongly enriched in the N24 induced group, because
413 deregulation of the naïve network only initiates upon 2i withdrawal.
414
415 To characterize the constitutive KO response genes, we grouped them based on their log-fold-
416 changes between KO and WT at N24. This resulted in twelve clusters, ranging in size from 20
417 to 597 genes (Supplementary Table 5) (Figure 6b). By Gene Ontology (GO) analysis, eight of
418 the twelve clusters showed a significant enrichment of at least one GO term containing more
419 than five genes (Figure 6c, Supplementary Table 5). Many of these GO terms relate to cell fate
420 specification and development. Notably, the constitutive response clusters often shared
421 identical or similar GO terms, suggesting functional convergence. NAGs were
422 underrepresented in constitutive clusters and only contained in clusters 1 (const c1) and 2,
423 for which expression profiles across all KOs correlated with naïve marker expression (Figure
424 6b). Therefore, deregulation of genes within const c1 and c2 in 2i is proposed to directly
425 translate to differentiation defects observed at N24. Cluster 1 genes were enriched for
426 general differentiation-related factors and response to Bmp signalling. Constitutive cluster 2
427 contained a significant number of Fgf/ERK pathway defining genes (Supplementary Figure 7e)
428 and was consistently enriched for Fgf/ERK signalling related terms (Supplementary Table 5).
429 Intriguingly, constitutive cluster 10 contained multiple *Zscan4* isoforms [49] and 14 out of 45
430 genes in this cluster are bound by the co-repressor Trim28, which is involved in silencing of
431 endogenous retroviruses [50-52]. Strong upregulation of this cluster was observed in *Csnk1a1*

432 and *Zfp281* KOs, but strong downregulation in *Rbpj* and *Jarid2* KOs. Overall, cluster 10
433 regulation showed no strict correlation with a differentiation delay phenotype.
434
435 Analysis of the N24 induced response genes yielded 12 clusters ranging from 8 to 103 genes
436 (Figure 6d and Supplementary Table 5). All naïve marker genes were detected in N24 induced
437 cluster 1 (N24i c1), with the notable exception of *Klf2*. *Klf2* was part of N24 induced cluster
438 10, within which more than half the genes are Tcf7l1 ChIP targets [27] (Supplementary Figure
439 7E). This provides further support for the notion that Wnt/Tcf7l1 controls *Klf2* expression
440 independently of the rest of the naïve network. Interestingly, twelve percent of genes in N24
441 induced cluster 10 were screen hit genes, suggesting a major impact of this cluster on the exit
442 from naïve pluripotency (Supplementary Figure 7e). NAGs were distributed in eight out of the
443 twelve N24 induced clusters. In marked contrast to the constitutive KO response, individual
444 N24-induced clusters displayed specific functional enrichments (Figure 6e). Therefore, the
445 N24 induced response clusters represent distinct molecular pathways and cellular functions
446 that were altered in multiple KOs. N24 induced cluster 7 was significantly regulated in *Etv5*,
447 *Trim71* and *Zfp281* KOs at N24 (Figure 6d). Cluster 7 contains pivotal cell fate switch genes
448 like *Fgf4*, *Tfe3* [14] and *Tcf7l1* itself, but NAGS were underrepresented in this cluster. Notably,
449 *Etv5, Trim71* and *Zfp281* are the three KO lines that show the strongest exit delay while
450 retaining E4.5 identity at N24 (Supplementary Figure 4d), suggesting that N24 induced cluster
451 7 contains crucial cell fate switch genes that are modulated during *in vitro* and *in vivo* exit
452 from naïve pluripotency. In summary, constitutive and N24 induced clusters define crucial
453 genetic modules that are co-regulated by multiple exit factors.
454

455 We asked whether the expression status of these genetic modules correlated with the activity
456 of differentiation regulating pathways. We utilized pathway activity levels and cluster specific
457 expression levels across all KOs to build a model predicting the impact of pathway activity on
458 cluster-gene expression (Figures 7A and S7A to F). Adding further evidence to the pathway to
459 cluster relationships, we found that in most cases activity changes of the regulatory pathways
460 preceded the expression changes of their putative target clusters over the 2h differentiation
461 time course. The Notch and mTor pathways showed high connectivity to N24-induced clusters
462 (8 and 6 out of 12 clusters, respectively), whereas Wnt/catenin signalling was specifically
463 correlated with clusters containing high levels of NAGs (N24i c1 and c10) and the only
464 constitutive cluster containing an appreciable number of NAGs (const c1) (Figure 7A).
465 Together, this shows that at least 24 distinct transcription modules can be defined, whose
466 expression is deregulated upon exit gene depletion and which are largely under the control
467 of five key signalling pathways gating differentiation.

468 # Discussion

469 Ordered progression through pluripotency requires shutdown of the naïve TF-network and
470 concomitant large-scale rewiring to establish the formative GRN [2, 3, 19]. Following Harald
471 Weintraub's pioneering work with muscle differentiation and Thomas Graf's work on
472 haematopoiesis, the ability of certain transcription factors to change cell identity without a
473 developmental context gained central prominence with somatic cell reprogramming to
474 pluripotency [53-55]. However, most of these studies involved forced expression of selected
475 transcription factors to achieve synthetic cell state transitions *in vitro*. Here we utilized the
476 remarkable properties of ESCs to recapitulate the pivotal embryonic transition from naïve to
477 formative pluripotency. Notably, we show that this developmental cell state transition is not
478 based on the action of one or two master transcription factors. Although crucial for defining
479 lineage trajectories, transcription factors do not initiate cell fate decisions. Instead, the robust
480 stability of naïve pluripotent stem cells is sustained by multiple crosstalking components and
481 is dismantled by at least four diverse signalling inputs.

482

483 Our dissection of an authentic cell state transition indicates that the exit from naïve
484 pluripotency is largely guided along a trajectory constrained by signalling pathways and
485 funnelled through a handful of genetic modules. These adaptations of cellular networks
486 establish the opportunity for TFs to initiate and consolidate cell identity. In the specific case
487 of the exit from naïve pluripotency, expression of Oct4 and Sox2, which are key TFs for both
488 the naïve and formative states, provide cornerstones for both the naïve and formative
489 regulatory networks to ensure lineage fidelity during differentiation (Figure 7B). Our data
490 support a cell fate transition model in which extinction of cell identity induced by the interplay

491 of signalling and the chromatin modification machinery precedes the initiation of a novel cell
492 fate.
493
494 Several signalling-cascades have been implicated in ESC self-renewal and naïve pluripotency
495 exit, including pathways controlled by LIF, Akt/mTOR, Wnt, Fgf/ERK and Notch. We derived
496 transcription footprints of these pathways and scrutinized the extent to which they were
497 deregulated by the different KOs to find that virtually all KOs showing a strong differentiation
498 defect also showed deregulation of at least one of these five pathways. Notably, however, we
499 observed segregation in terms of pathway deregulation. Thus, KOs that induce e.g. a *Tcf7l1*-
500 like profile were less likely to show a signature of another pathway. This suggests that ESC
501 differentiation requires coordinated activity changes of independent pathways under control
502 of separable genetic networks. The observation that single gene depletion is not sufficient to
503 prevent exit from the naïve state is consistent with this interpretation [21]. Our analyses
504 exposed discrete pathway and GRN features that mediate timely and robust mammalian cell
505 state transitions. We found that 12 genetic sub-networks (clusters) with non-overlapping
506 cellular functions emerge during the exit from naïve pluripotency. Interestingly, signalling
507 pathway activity was closely co-regulated with these clusters suggesting a functional linkage.
508 This implies that formative differentiation is driven by a limited number of parallel acting
509 genetic modules under control of, or controlling, key signalling pathway activities.
510
511 By examining co-regulation with the core naïve network across 146 perturbations (73 KOs in
512 2i and at N24) we identified the NAGs. NAGs obey strikingly similar expression dynamics *in*
513 *vitro* and *in vivo*. Moreover, in the macaque pre- to post-implantation transition and in human
514 naïve cell capacitation, orthologous NAGs show similar behaviour as during mESC exit from

naïve pluripotency. Therefore, this cohort of genes constitutes a layer of the pluripotency network that is tightly linked to, and likely acts in conjunction with, the core naïve TF-network in mammals [4]. We suggest that collective modulation of NAG expression will propel naïve cells into formative differentiation and conversely, that collective NAG deregulation will delay proper differentiation. The control of NAG expression appears to be interwoven with the signalling pathways involved in controlling the naïve to formative transition. The N24 induced clusters enriched for NAGs showed high connectivity to all tested pathways. The functional significance of NAGs likely extends beyond self-renewal and naïve identity. Indeed, most NAGs are not transcription factors. The growth factor Pdgfa is a case in point. *Pdgfa* is the most strongly associated NAG, but has no known activity on ESCs, which do not express the cognate receptor, Pdfgra. Within the ICM, Pdgfa expression is restricted to the naïve epiblast, whereas Pdgfra is present exclusively in the primitive endoderm. This reciprocal expression pattern is consistent with the known paracrine action of Pdgfa to promote primitive endoderm segregation [35]. Thus, conserved linkage of NAGs to the naïve TF network may mediate paracrine communication to other lineages within the blastocyst as well as consolidate naïve epiblast fate.

Whereas most naïve TFs were tightly co-regulated across the 73 KOs, *Klf2* expression was uncoupled from other members of the core network. Interestingly, while *Klf2* is highly expressed in mouse ESCs and naïve epiblast, it is very lowly expressed in primate blastocysts, and also in the porcine pluripotent compartment [56]. *Klf2* may therefore be a rodent specific addition to the naïve TF-network. A further significant species difference is the very low expression of *TCF7L1* in primate naïve cells, which underlies the differential responsiveness of mouse and human naïve cells to GSK3 or Wnt pathway inhibition [34, 57]. *Klf2* is a genomic

539 target of Tcf7l1 [27]. Interestingly, *Klf2* and *Tcf7l1* are found in the same N24 induced gene-
540 expression cluster (cluster 10), suggesting coregulation of both components of this mouse
541 specific naïve pluripotency feedback loop. For most KOs the degree of failure in naïve TF
542 downregulation and the neutralisation of global differentiation-related transcription changes
543 are comparable. However, the relatively weak *Trim71* phenotype despite specific
544 upregulation of some naïve TF-genes, suggests that up-regulation solely of core naïve factors
545 must reach a certain threshold to stably maintain naïve pluripotency, consistent with negative
546 feedback constraints within the network [4, 5, 58].
547
548 The remarkable advantages of ESCs as an experimental venue include the utilisation of well-
549 defined culture conditions to recapitulate key features of *in utero* developmental progression
550 [1, 2, 59, 60]. Our high-resolution transcriptome analysis confirms that differentiation over 24
551 hours *in vitro* mimics the peri-implantation transition from naïve to formative epiblast *in*
552 *utero.* These data provide rich resources for further studies. In particular, many strong
553 differentiation delay KO transcriptome profiles retained transcriptome resemblance to the
554 pre-implantation epiblast. This relationship held true in 2i and at N24 and indicates that the
555 KO of these genes perturb the operative *in vivo* cell state transition machinery. In general, we
556 surmise that stalling the activity of the exit machinery is critical to *in vitro* capture and self-
557 renewal of naïve stem cells, consistent with empirical requirements for signal inhibition in
558 both mouse and human [7, 57]. Interestingly, deficiency for some differentiation drivers
559 increases similarity of 2i profiles to the *in vivo* pre-implantation epiblast. Therefore, we
560 propose that some of the mechanisms driving exit from naïve pluripotency contribute to the
561 lack of complete identity between *in vivo* E4.5 epiblast and ES cells in 2i. This is in line with
562 ESCs representing an engineered interruption in developmental progression and suggests

563 that state-of-the-art self-renewal culture conditions may not yet capture the naïve state
564 perfectly [1].
565
566 Our analyses provide a comprehensive inventory of genetic factors and regulatory networks
567 governing the major cell fate transition from pre- to post-implantation epiblast (Figure 7B).
568 We present evidence that mammalian development, employs similar regulatory networks
569 operating with similar mechanisms [61, 62]. These programmed cell state transitions in
570 development do not rely upon instructions delivered by one or two master transcription
571 factors. Rather, a cloud of activity, involving multiple co-ordinated inputs, serves to destabilise
572 an existing, multiply stabilised GRN. Thereby a consolidated cell state can be rapidly
573 dismantled by a combination of signals, triggering transition to the next developmental
574 waystation.

575 **Acknowledgements**

576 We thank Johanna Stranner, Thomas Sauer and Andy Riddell for help with flow cytometry.
577 Meng Li and Kosuke Yusa for sharing RC9 cells, Andreas Dahl for NGS-sequencing support and
578 Christa Bücker for critical comments on the manuscript. ML is funded by a WWTF-VRG grant
579 (VRG14-006). This study was supported by an FWF/DFG DACH grant to AB and ML (FWF grant
580 number: I 3786; DGF grant number 398882498). AB and MG were supported by the BMBF
581 (Sybacol). RS received support by the Cologne Graduate School of Ageing Research. Wellcome
582 and the Medical Research Council provide core support for the Wellcome-MRC Cambridge
583 Stem Cell Institute. AS is an MRC Professor. This project was initiated within the EU FP7
584 integrated project SyBoSS (Grant agreement: 242129) co-ordinated by AFS and AB including
585 partners ML and AS.

586

587 **Author contributions:**

588 AB, ML, AS and AFS conceptualized the study. AB, ML designed the study. AL, ML, AB, RS, MG
589 and FTT designed experiments. AL, MH, JR, PvdL, HFT, LS and MS carried out wet-lab
590 experiments. RS, MG, GGS, FTT and AL performed bioinformatic analyses. ML, AB, AS, AFS
591 wrote the paper with input from AL, RS, MG, MH, FTT and GG. Funding was acquired by ML,
592 AB, AFS and AS.

593

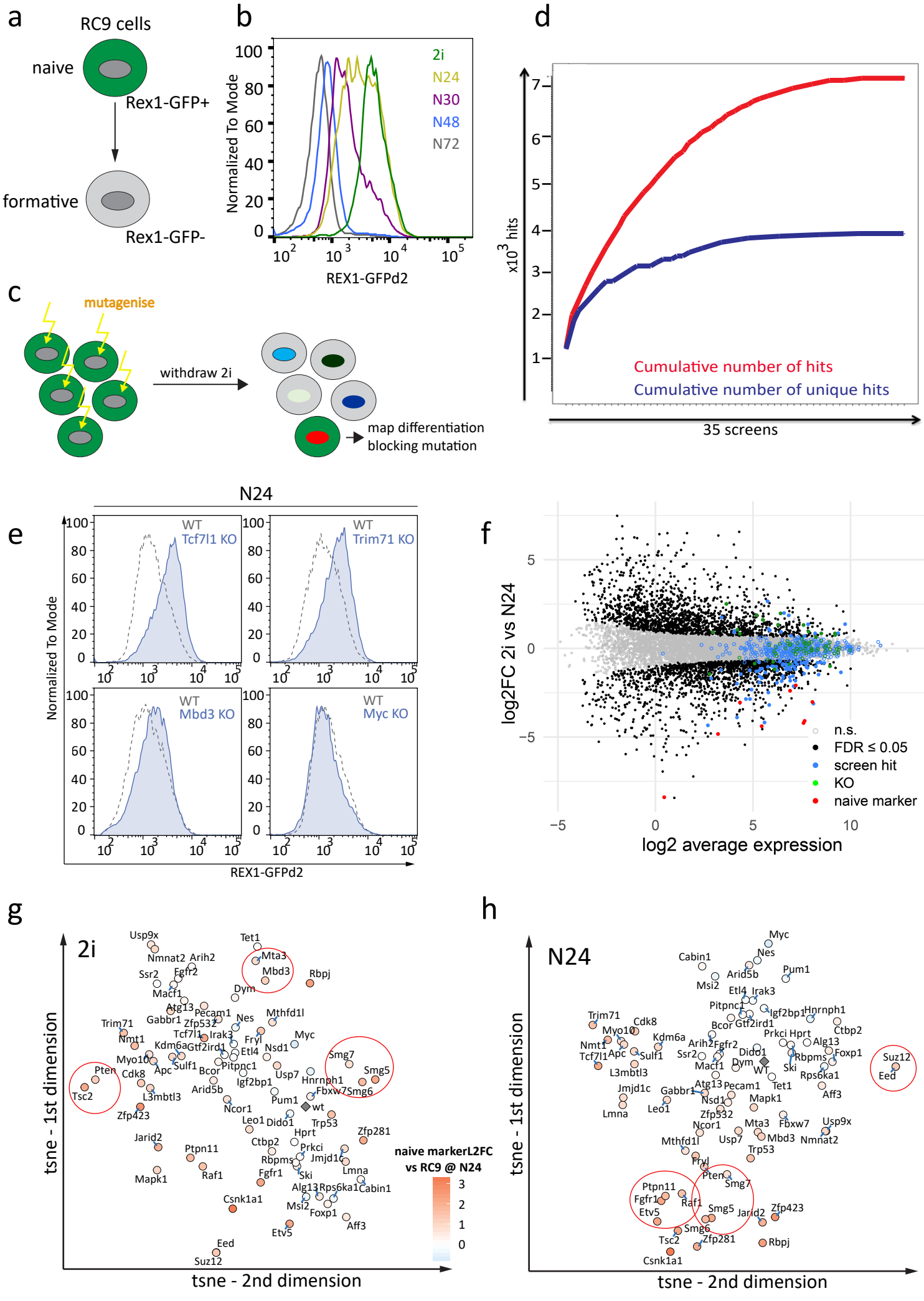
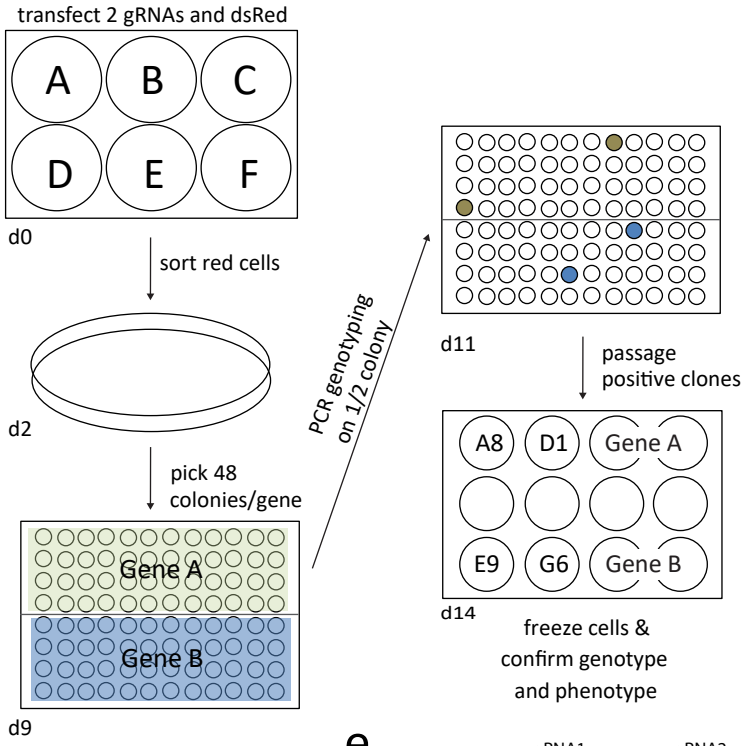

a
RC9 cells
naive
Rex1-GFP+
formative
Rex1-GFP-
b
Normalized To Mode
2i
N24
N30
N48
N72
REX1-GFPd2
c
mutagenise
withdraw 2i
map differentiation
blocking mutation
d
x10^3 hits
Cumulative number of hits
Cumulative number of unique hits
35 screens
e
N24
WT
Tcf7l1 KO
Trim71 KO
Mbd3 KO
Myc KO
Normalized To Mode
REX1-GFPd2
f
log2FC 2i vs N24
log2 average expression
n.s.
FDR ≤ 0.05
screen hit
KO
naive marker
g
2i
tsne - 1st dimension
tsne - 2nd dimension
naive markerL2FC
vs RC9 @ N24
h
N24
tsne - 1st dimension
tsne - 2nd dimension

Figure 1

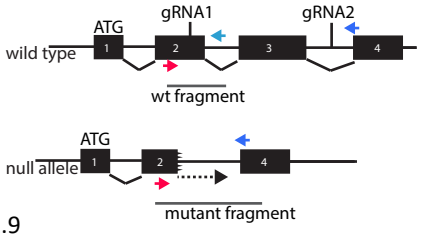
a

Gene	times hit in 35 screens	integration sites	fdr
Tcf7l1	34	166	1E-232
Fgfr2	25	78	6.1E-21
Alg13	20	35	9.3E-39
Ctbp2	17	43	5.6E-45
Arfip1	17	29	4.5E-34
Jarid2	16	35	7E-31
Rbpj	16	25	7E-19
Nsd1	14	29	1.1E-26
Arid5b	14	17	1.9E-11
Fbxw7	13	25	1.9E-16
Msi2	13	24	1.2E-12
Dym	13	16	6.6E-07
Etv5	12	26	8E-30
Mapk1	12	19	1.1E-16
Smg7	12	19	3.5E-15

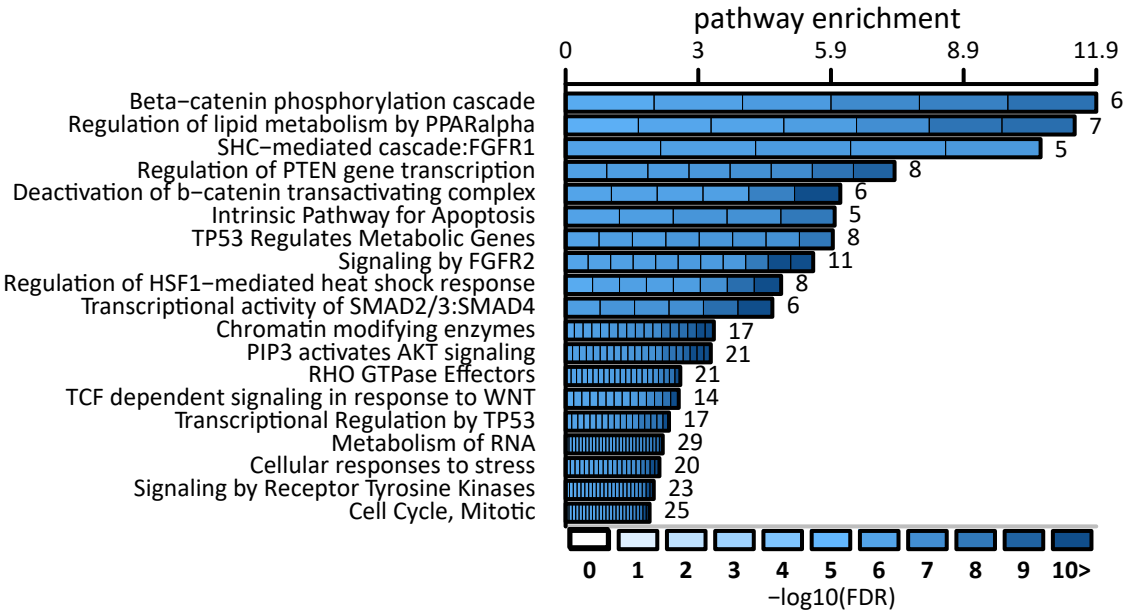
d



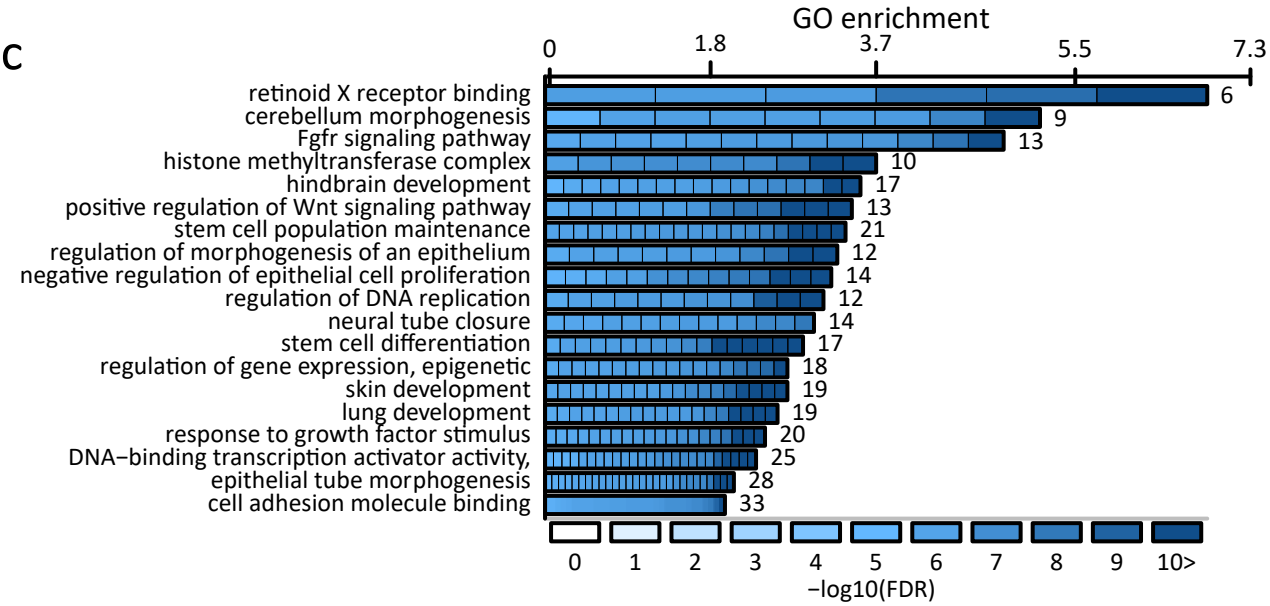
e



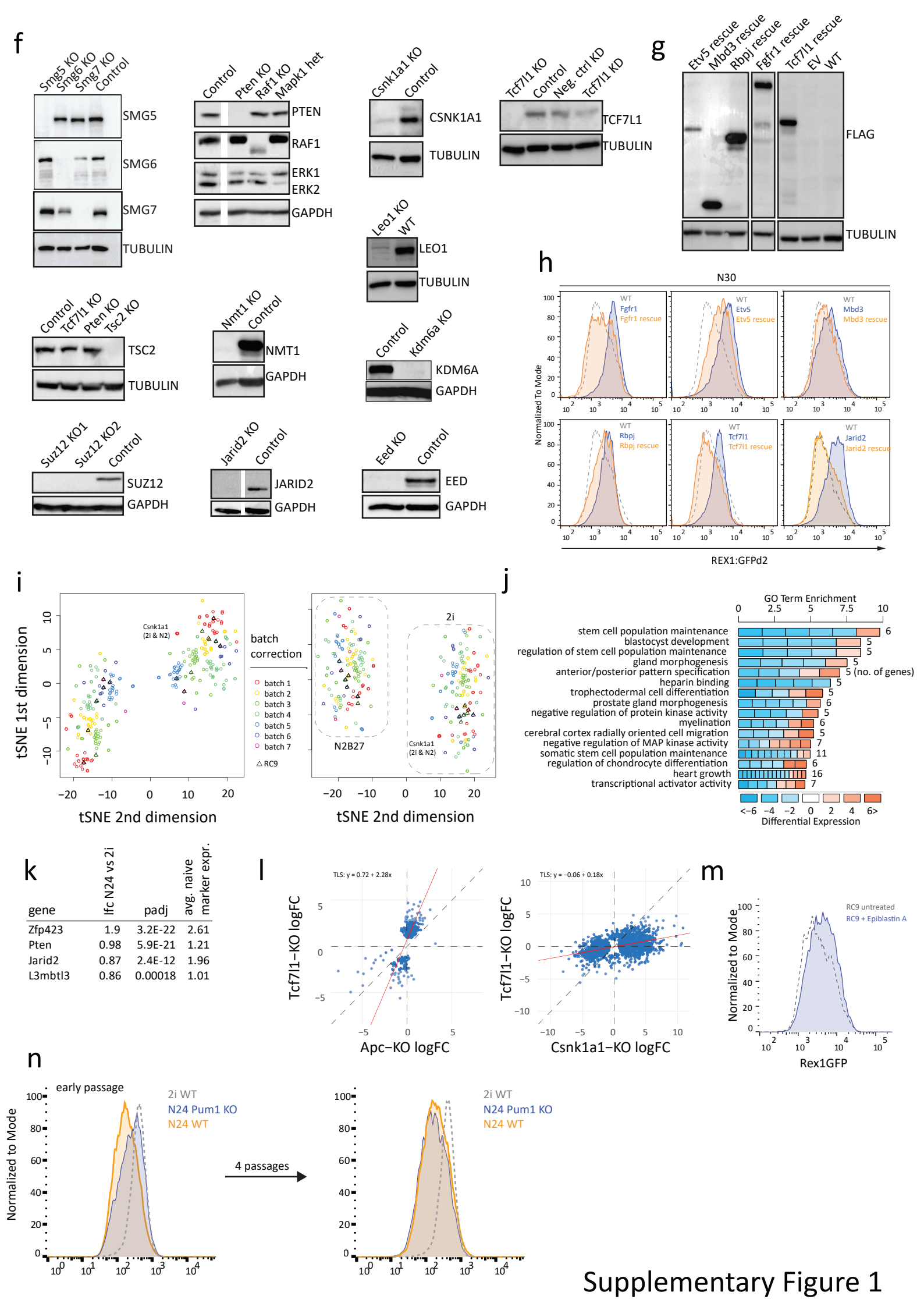
b



c



Supplementary Figure 1



gene	lfc N24 vs 2i	padj	avg. naive marker expr.
Zfp423	1.9	3.2E-22	2.61
Pten	0.98	5.9E-21	1.21
Jarid2	0.87	2.4E-12	1.96
L3mbtl3	0.86	0.00018	1.01

Supplementary Figure 1

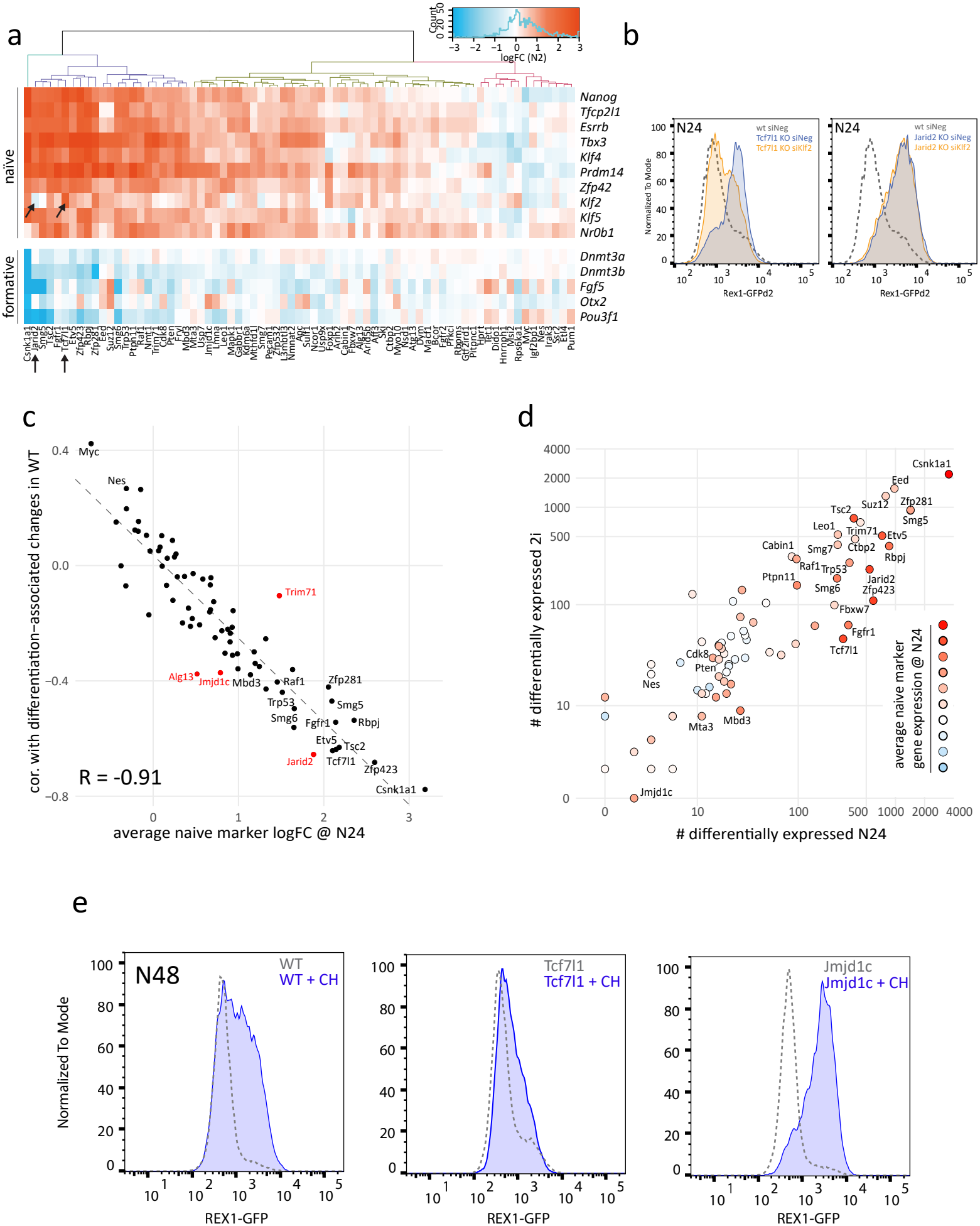
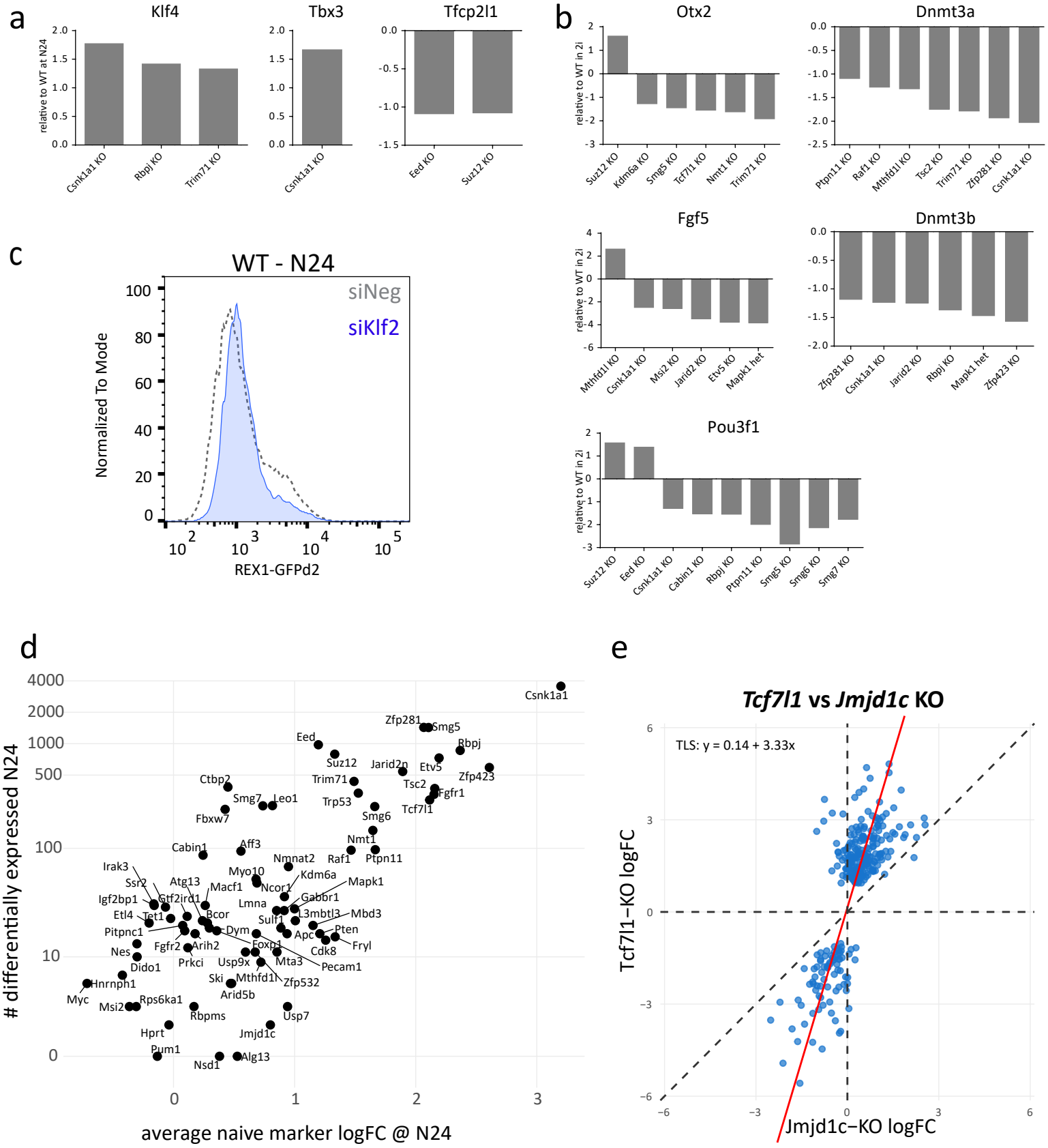
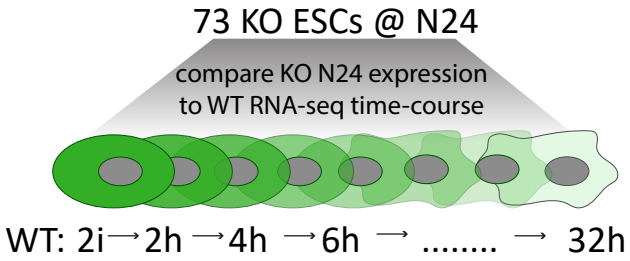

a
Count
logFC (N2)
naïve
formative
Nanog
Tfcp2l1
Esrrb
Tbx3
Klf4
Prdm14
Zfp42
Klf2
Klf5
Nr0b1
Dnmt3a
Dnmt3b
Fgf5
Otx2
Pou3f1
b
N24
wt siNeg
Tcf7l1 KO siNeg
Tcf7l1 KO siKlf2
Jarid2 KO siNeg
Jarid2 KO siKlf2
Normalized To Mode
Rex1-GFPd2
c
cor. with differentiation-associated changes in WT
average naive marker logFC @ N24
R = -0.91
Myc
Nes
Trim71
Alg13
Jmjd1c
Mbd3
Raf1
Zfp281
Trp53
Smg5
Smg6
Fgfr1
Rbpj
Etv5
Tsc2
Jarid2
Tcf7l1
Zfp423
Csnk1a1
d
differentially expressed 2i
differentially expressed N24
average naive marker gene expression @ N24
Csnk1a1
Eed
Suz12
Zfp281
Tsc2
Smg5
Leo1
Trim71
Etv5
Cabin1
Smg7
Ctbp2
Rbpj
Raf1
Trp53
Ptpn11
Jarid2
Smg6
Zfp423
Fbxw7
Fgfr1
Tcf7l1
Cdk8
Pten
Nes
Mta3
Mbd3
Jmjd1c
e
N48
WT
WT + CH
Tcf7l1
Tcf7l1 + CH
Jmjd1c
Jmjd1c + CH
Normalized To Mode
REX1-GFP

Figure 2

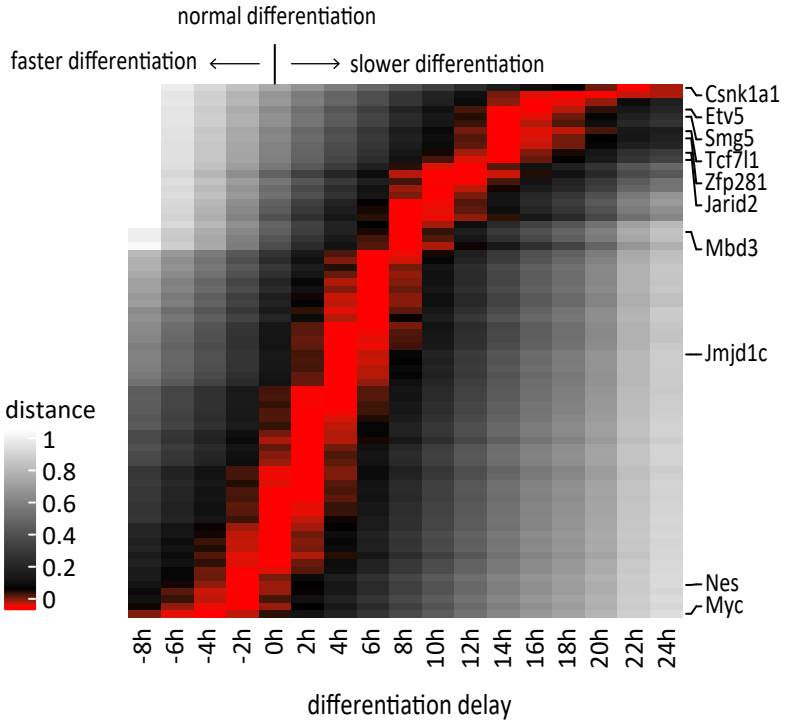


Supplementary Figure 2

a


73 KO ESCs @ N24
compare KO N24 expression
to WT RNA-seq time-course
WT: 2i→2h →4h →6h → ..... → 32h

b


normal differentiation
faster differentiation
slower differentiation
Csnk1a1
Etv5
Smg5
Tcf7l1
Zfp281
Jarid2
Mbd3
Jmjd1c
Nes
Myc
distance
1
0.8
0.6
0.4
0.2
0
-8h
-6h
-4h
-2h
0h
2h
4h
6h
8h
10h
12h
14h
16h
18h
20h
22h
24h
differentiation delay

c

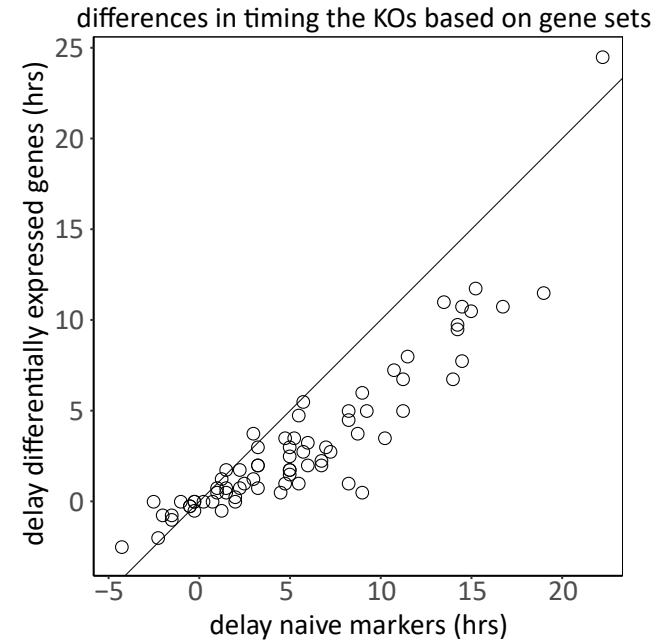
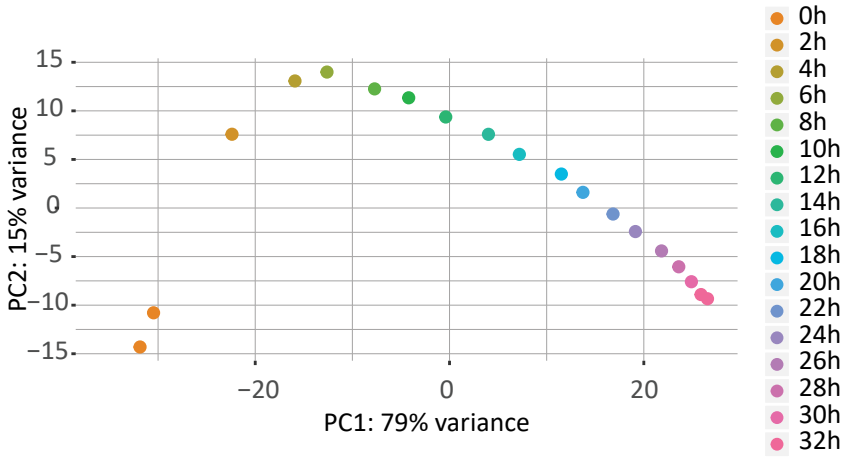
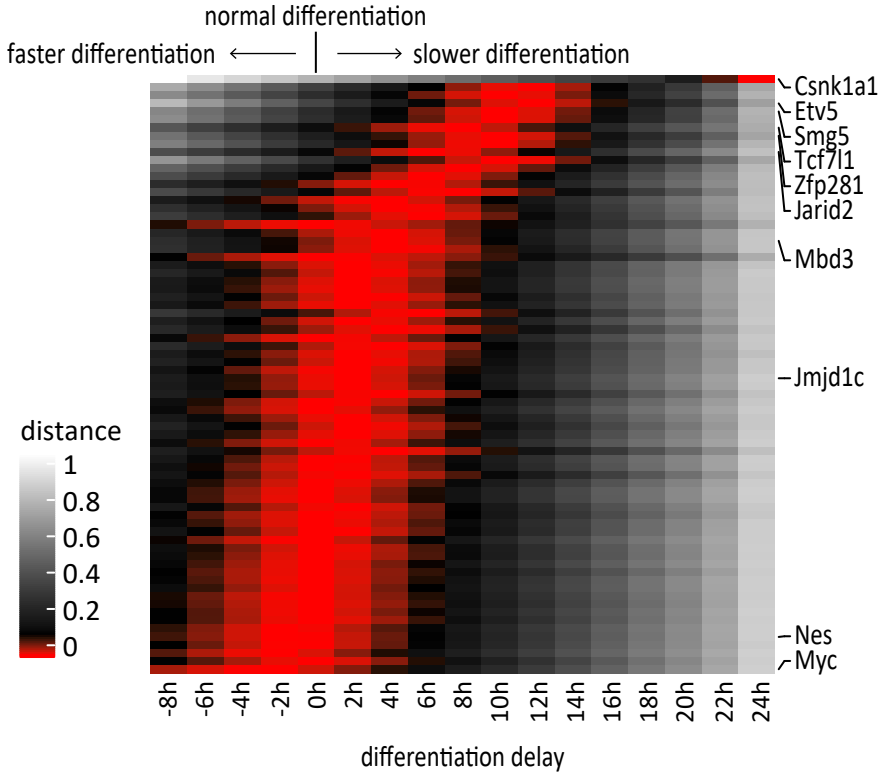

differences in timing the KOs based on gene sets
delay differentially expressed genes (hrs)
delay naive markers (hrs)
25
20
15
10
5
0
−5
0
5
10
15
20

Figure 3

a



b



Supplementary Figure 3

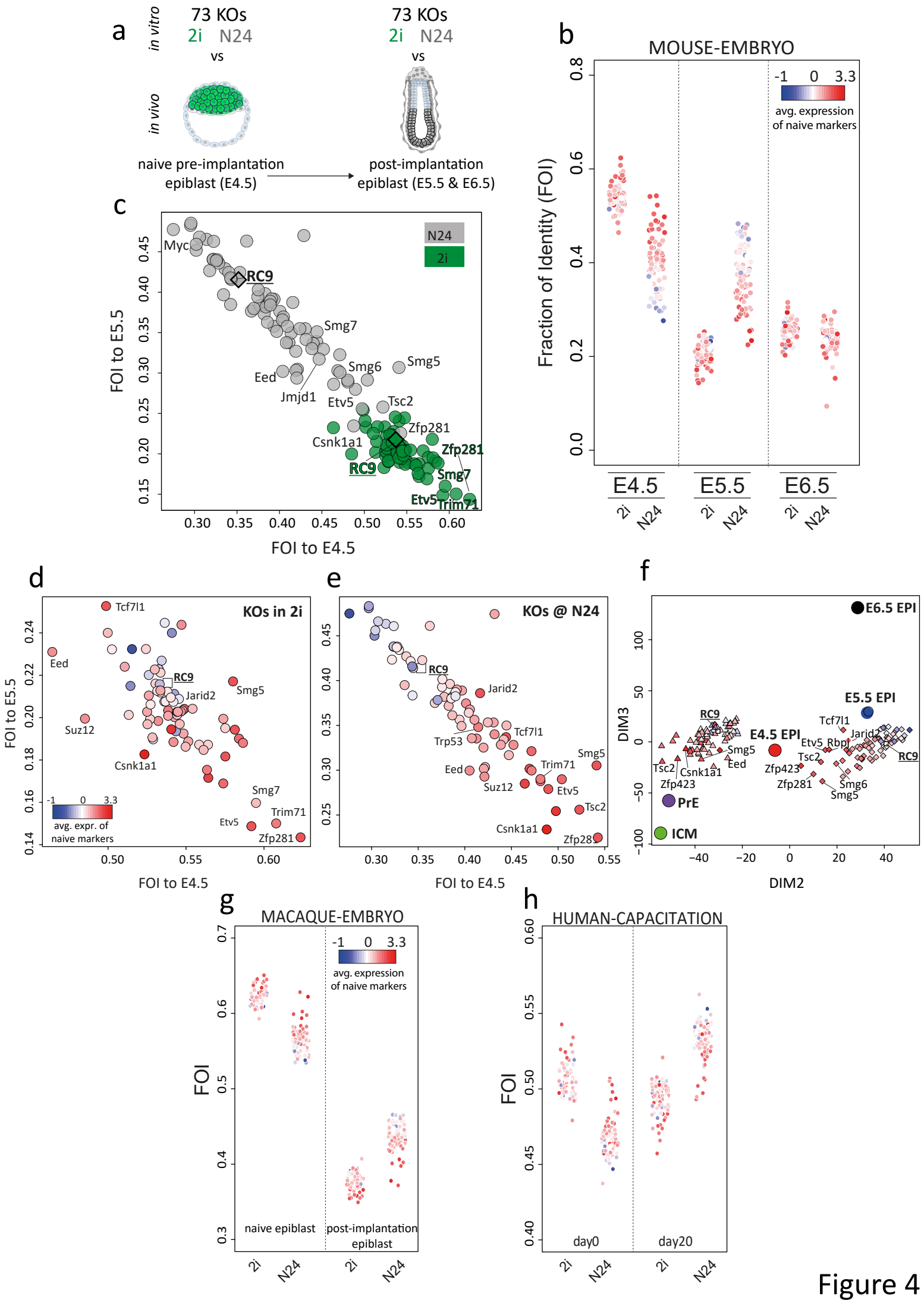
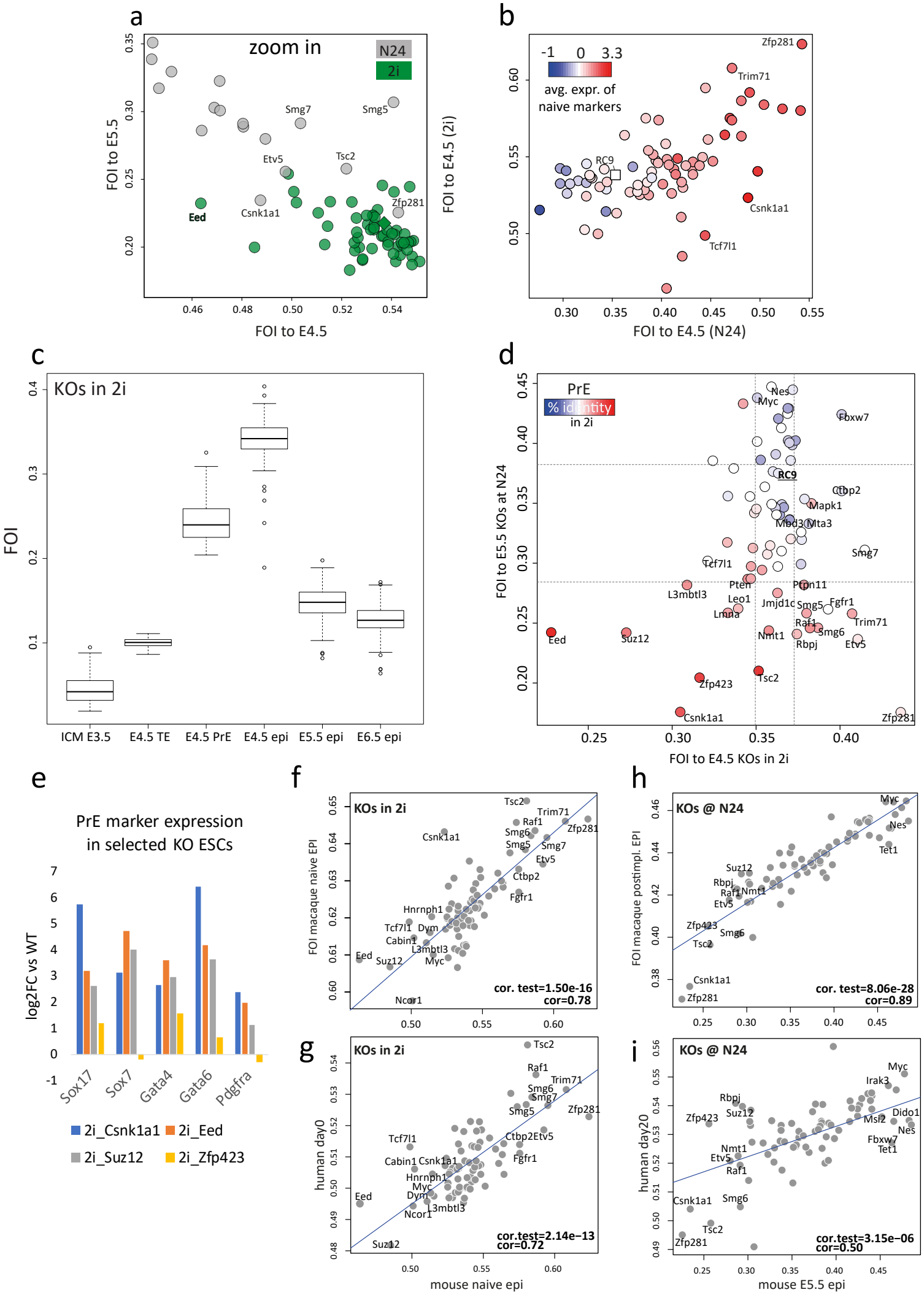

a
in vitro
in vivo
73 KOs
2i N24
vs
73 KOs
2i N24
vs
naive pre-implantation epiblast (E4.5)
post-implantation epiblast (E5.5 & E6.5)
b
MOUSE-EMBRYO
Fraction of Identity (FOI)
avg. expression of naive markers
E4.5
E5.5
E6.5
2i
N24
c
N24
2i
FOI to E5.5
FOI to E4.5
Myc
RC9
Smg7
Smg6
Smg5
Eed
Jmjd1
Etv5
Tsc2
Zfp281
Csnk1a1
Trim71
d
KOs in 2i
Tcf7l1
Eed
RC9
Jarid2
Smg5
Suz12
Csnk1a1
Smg7
Trim71
Etv5
Zfp281
avg. expr. of naive markers
e
KOs @ N24
RC9
Jarid2
Tcf7l1
Trp53
Smg5
Trim71
Eed
Suz12
Etv5
Tsc2
Csnk1a1
f
E6.5 EPI
E5.5 EPI
E4.5 EPI
RC9
Tcf7l1
Jarid2
Etv5
Rbpj
Smg5
Eed
Tsc2
Csnk1a1
Zfp423
Zfp281
Smg6
PrE
ICM
DIM3
DIM2
g
MACAQUE-EMBRYO
FOI
avg. expression of naive markers
naive epiblast
post-implantation epiblast
h
HUMAN-CAPACITATION
FOI
day0
day20

Figure 4



Supplementary Figure 4

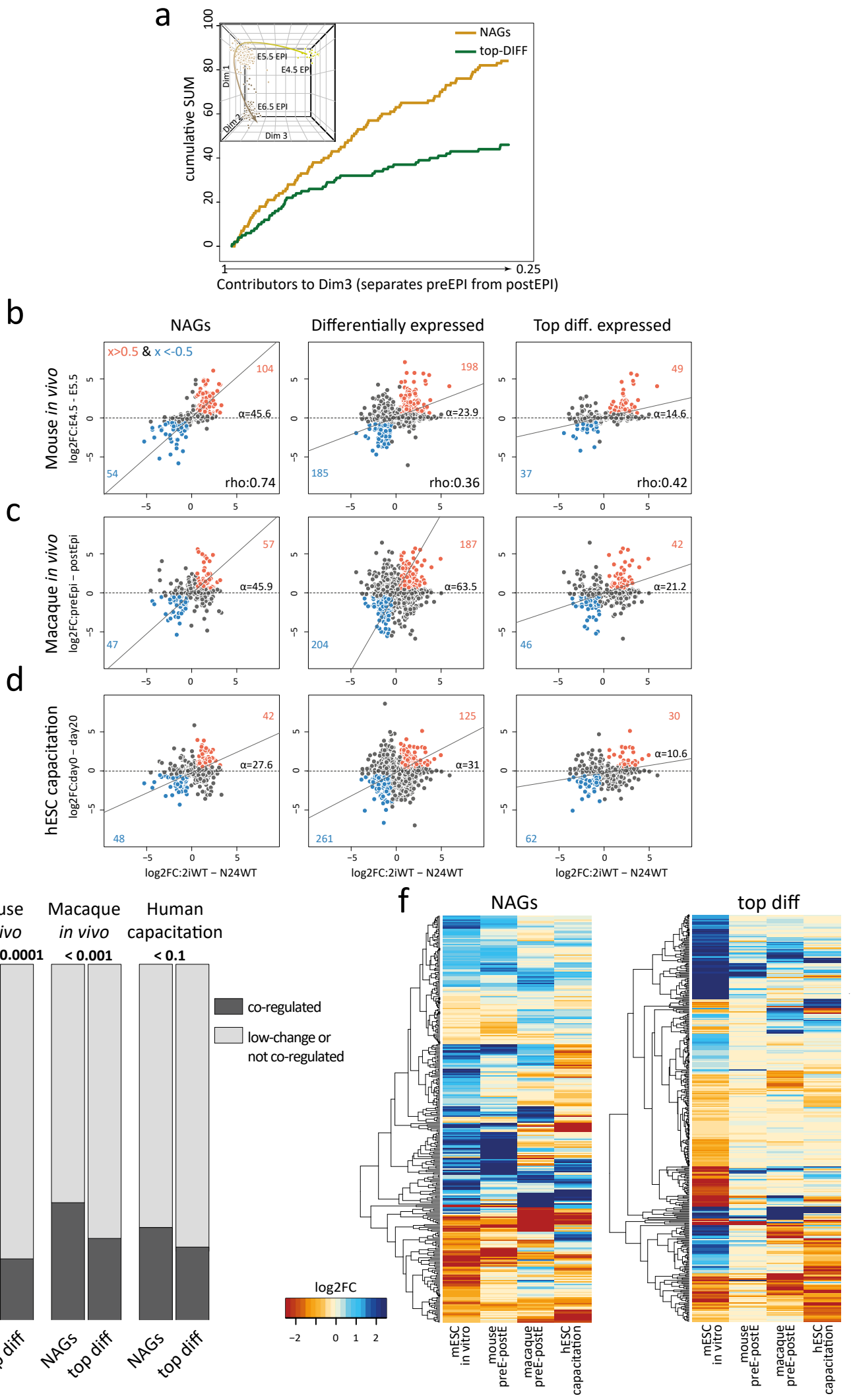
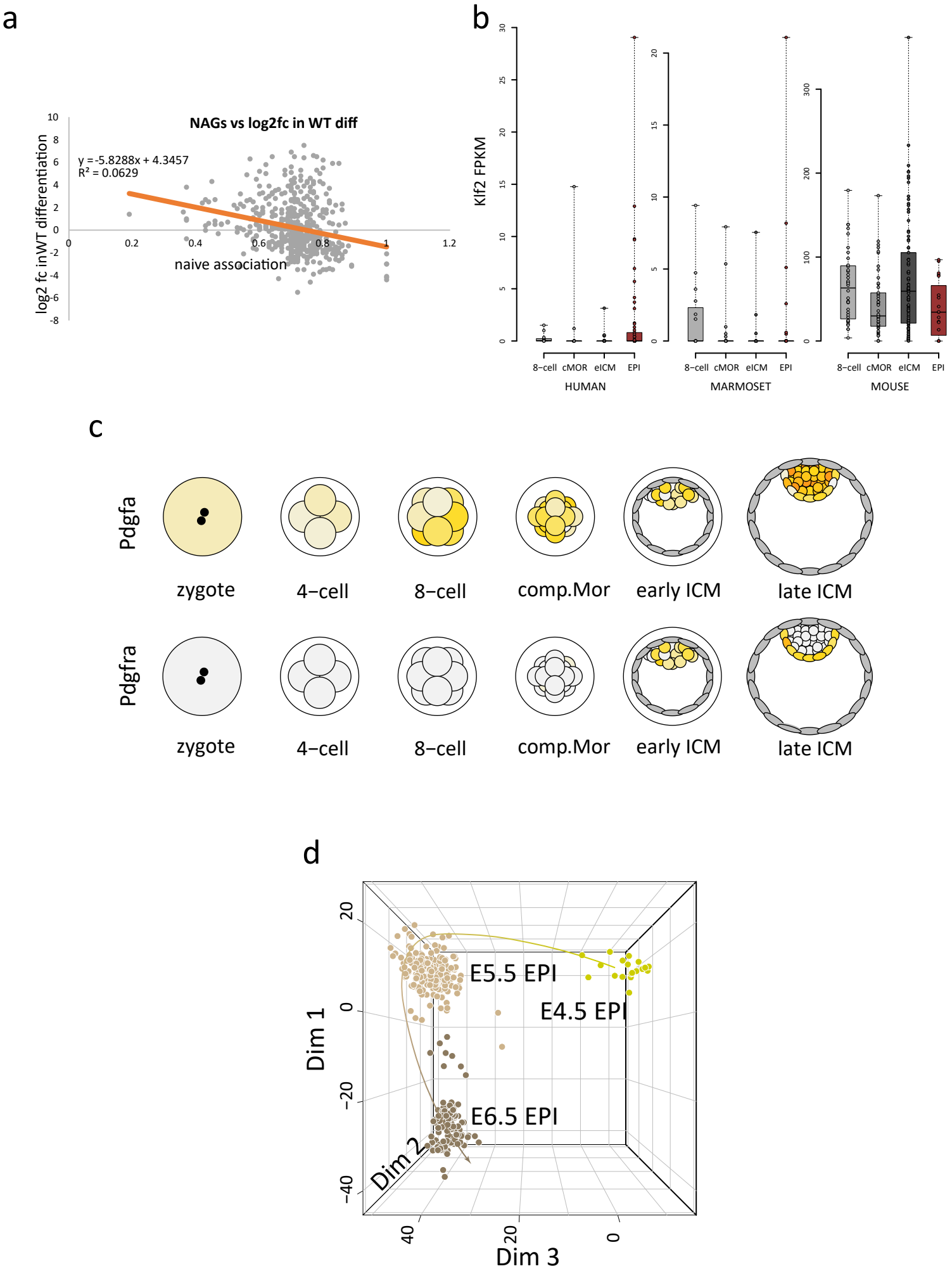

a
cumulative SUM
NAGs
top-DIFF
E5.5 EPI
E4.5 EPI
E6.5 EPI
Dim 1
Dim 2
Dim 3
Contributors to Dim3 (separates preEPI from postEPI)
b
NAGs
Differentially expressed
Top diff. expressed
Mouse in vivo
log2FC:E4.5 - E5.5
x>0.5 & x <-0.5
rho:0.74
rho:0.36
rho:0.42
c
Macaque in vivo
log2FC:preEpi – postEpi
d
hESC capacitation
log2FC:day0 – day20
log2FC:2iWT – N24WT
e
Mouse in vivo
Macaque in vivo
Human capacitation
adj. p < 0.0001
< 0.001
< 0.1
Frequency of co-regulation
co-regulated
low-change or not co-regulated
NAGs
top diff
f
NAGs
top diff
log2FC
mESC in vitro
mouse preE-postE
macaque preE-postE
hESC capacitation

Figure 5



Supplementary Figure 5

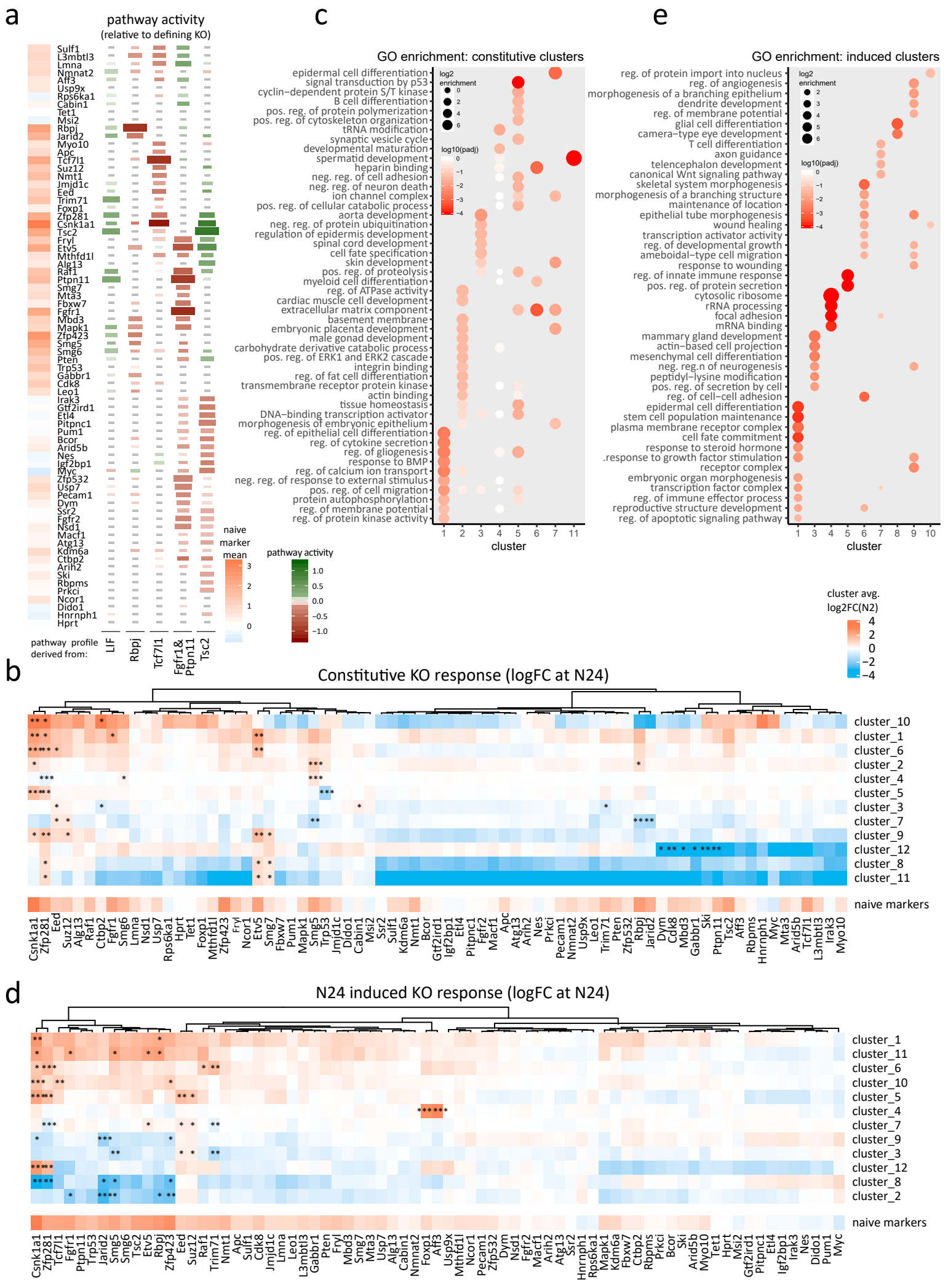
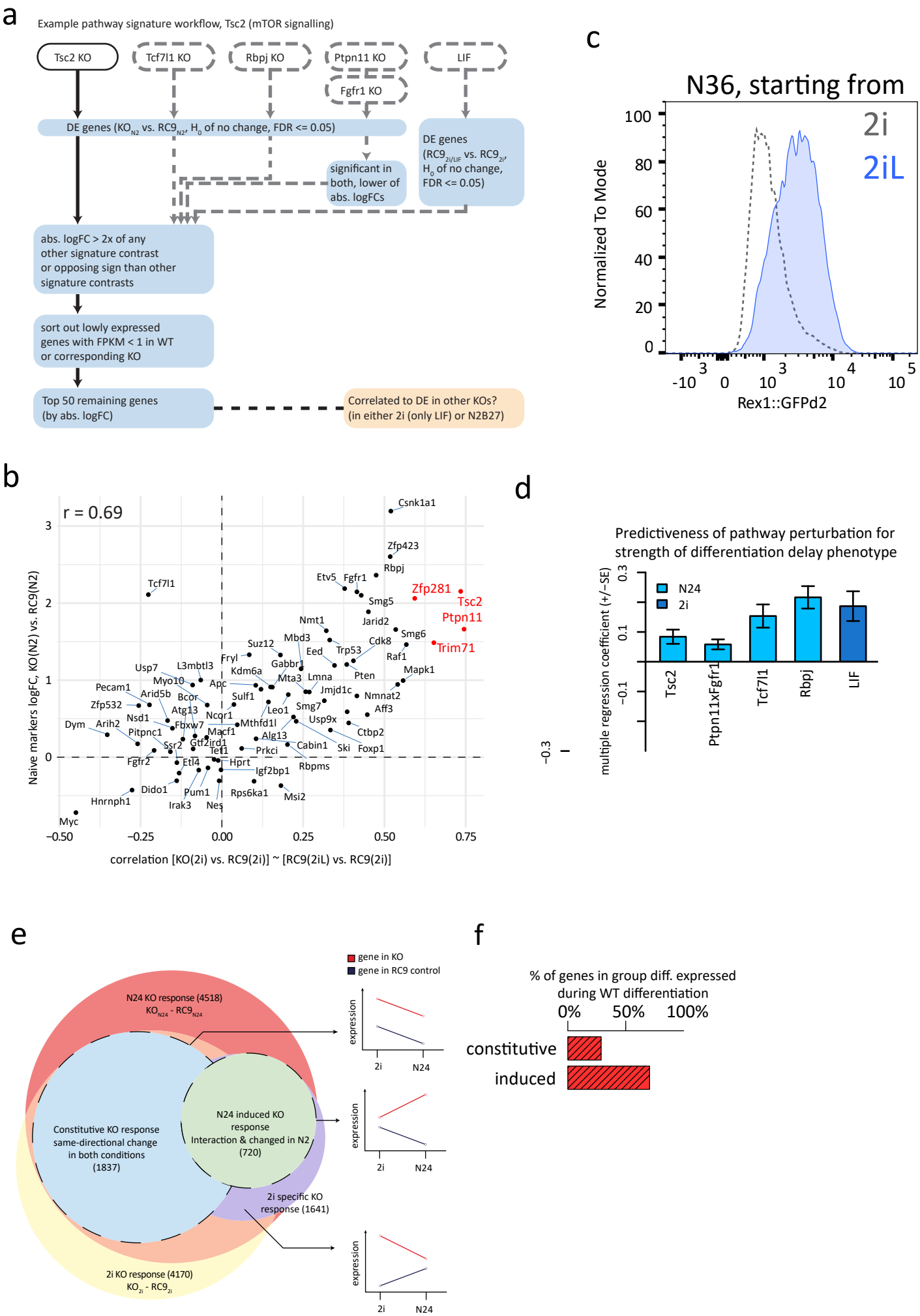
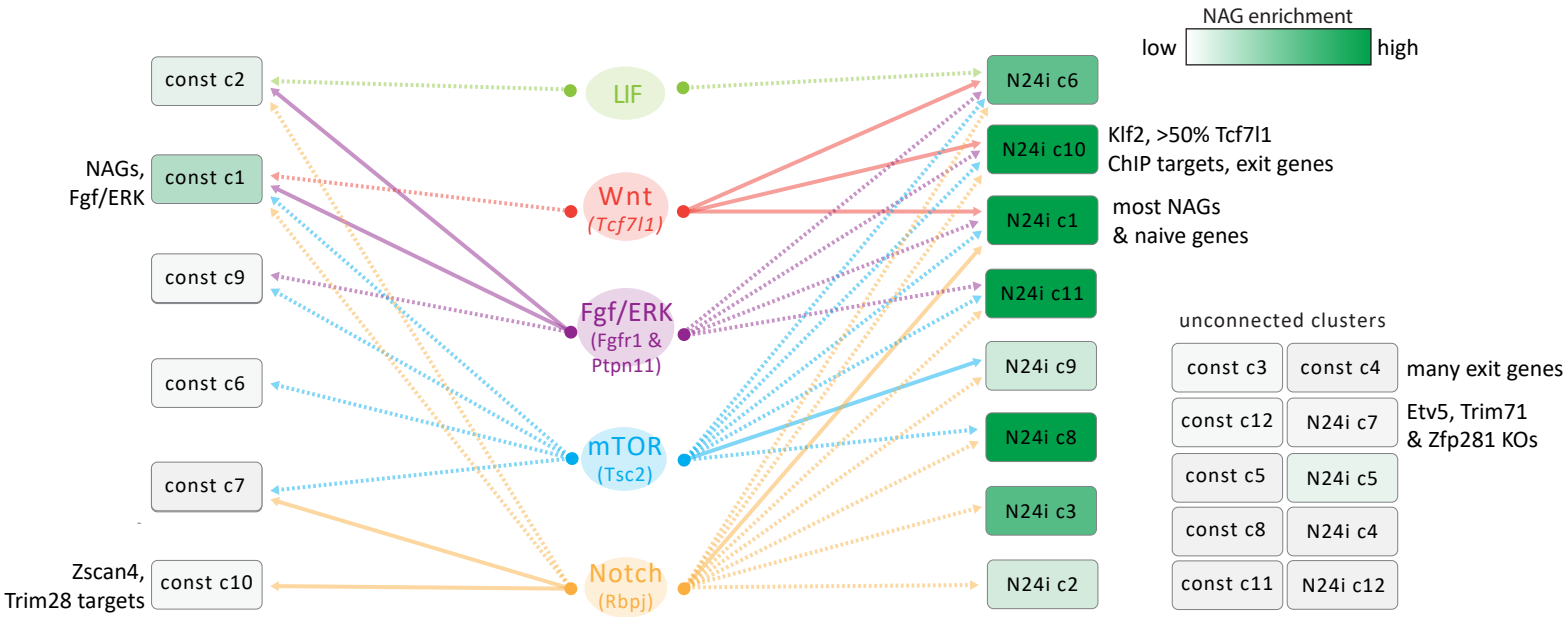

a
pathway activity
(relative to defining KO)
pathway profile derived from:
LIF
Rbpj
Tcf7l1
Fgfr1& Ptpn11
Tsc2
naive marker mean
pathway activity
c
GO enrichment: constitutive clusters
log2 enrichment
log10(padj)
cluster
e
GO enrichment: induced clusters
cluster avg. log2FC(N2)
b
Constitutive KO response (logFC at N24)
naive markers
d
N24 induced KO response (logFC at N24)
naive markers

Figure 6



Supplementary Figure 6

a


NAG enrichment
low
high
const c2
NAGs, Fgf/ERK
const c1
const c9
const c6
const c7
Zscan4, Trim28 targets
const c10
LIF
Wnt (Tcf7l1)
Fgf/ERK (Fgfr1 & Ptpn11)
mTOR (Tsc2)
Notch (Rbpj)
N24i c6
N24i c10
N24i c1
N24i c11
N24i c9
N24i c8
N24i c3
N24i c2
Klf2, >50% Tcf7l1 ChIP targets, exit genes
most NAGs & naive genes
unconnected clusters
const c3
const c4
many exit genes
const c12
N24i c7
Etv5, Trim71 & Zfp281 KOs
const c5
N24i c5
const c8
N24i c4
const c11
N24i c12

b

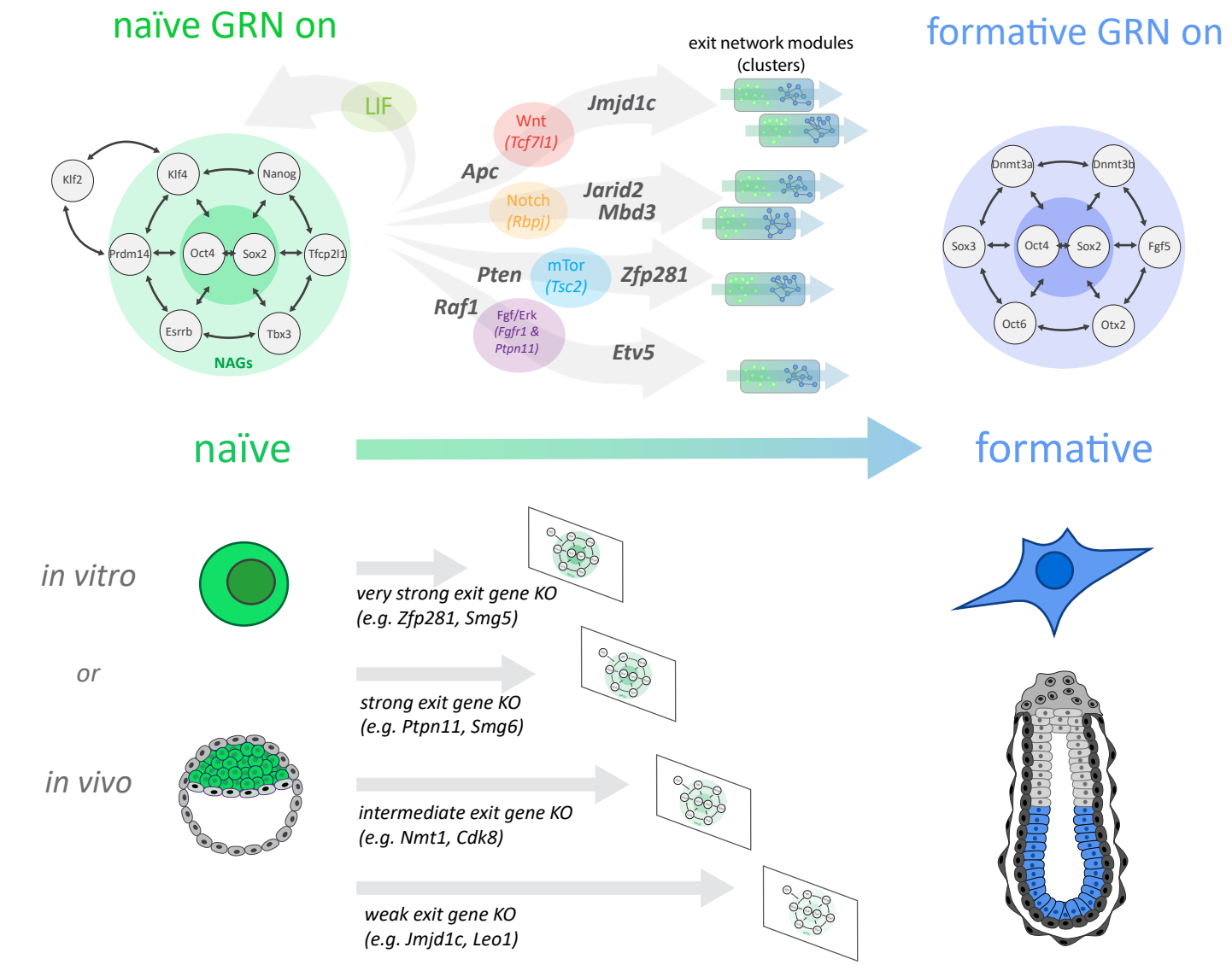
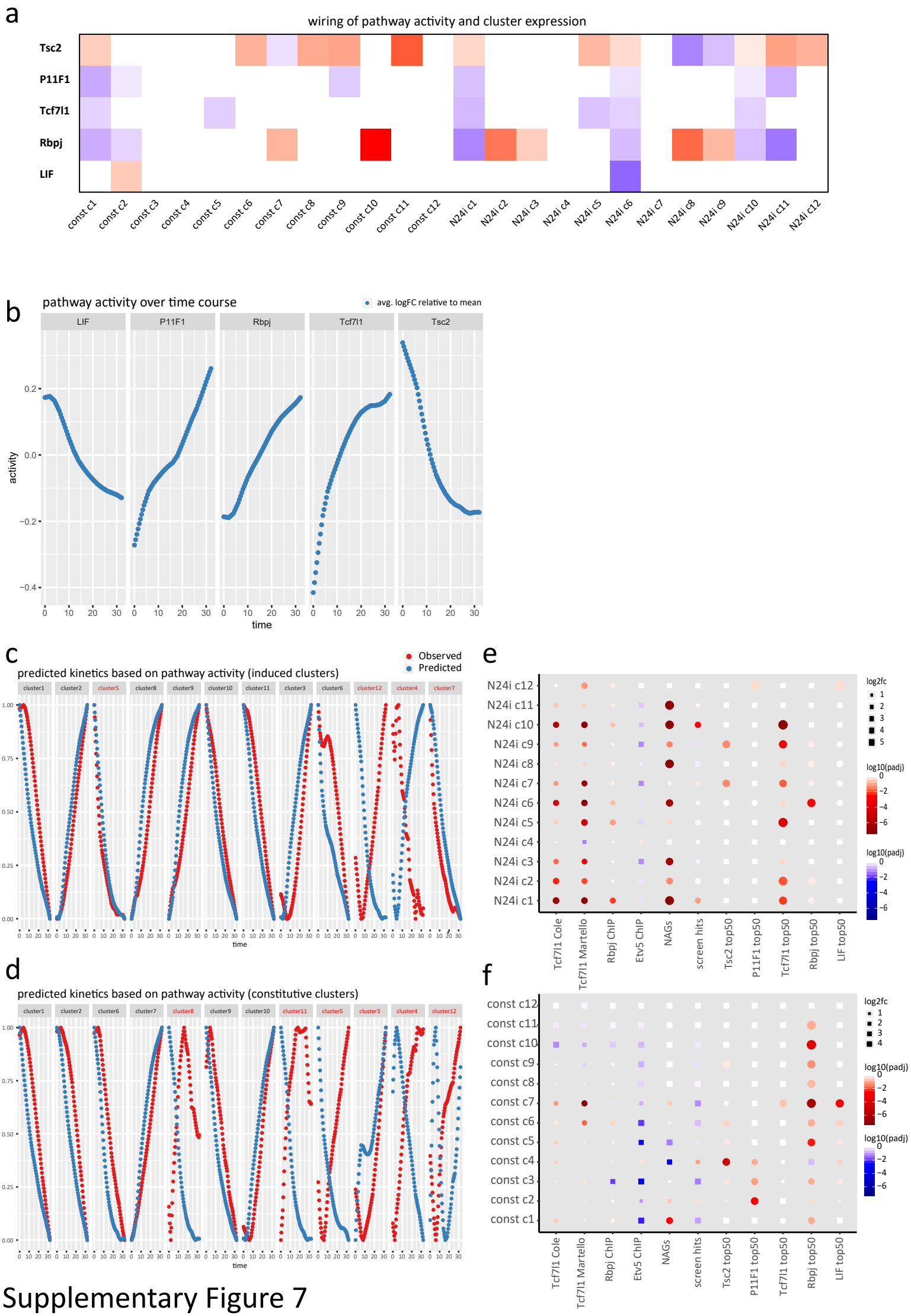

naïve GRN on
formative GRN on
exit network modules (clusters)
LIF
Klf2
Klf4
Nanog
Prdm14
Oct4
Sox2
Tfcp2l1
Esrrb
Tbx3
NAGs
Wnt (Tcf7l1)
Jmjd1c
Apc
Notch (Rbpj)
Jarid2
Mbd3
Pten
mTor (Tsc2)
Zfp281
Raf1
Fgf/Erk (Fgfr1 & Ptpn11)
Etv5
Dnmt3a
Dnmt3b
Sox3
Oct4
Sox2
Fgf5
Oct6
Otx2
naïve
formative
in vitro
or
in vivo
very strong exit gene KO (e.g. Zfp281, Smg5)
strong exit gene KO (e.g. Ptpn11, Smg6)
intermediate exit gene KO (e.g. Nmt1, Cdk8)
weak exit gene KO (e.g. Jmjd1c, Leo1)

Figure 7



Supplementary Figure 7

Figure Legends

Figure 1

a, Illustration of the Rex1::GFPd2 reporter cell line and its' exit from naïve pluripotency. Rex1-GFPd2 (in short Rex1-GFP) expression is tightly linked to naïve pluripotency. Shutdown of GFP expression indicates commitment to differentiation.

b, FACS analysis of Rex1-GFPd2 reporter levels throughout a 72h differentiation time-course after 2i withdrawal.

c, Scheme of the screening strategy to identify candidate genes involved in exit from naïve pluripotency. After random insertional mutagenesis using piggyBac transposon gene-traps, haploid Rex1::GFPd2 ESCs were released into differentiation. Cells that maintain GFP expression after exposure to differentiation conditions were isolated and the gene-trap insertion sites mapped.

d, Plot showing cumulative number of hits and the cumulative number of novel hits over 35 independent insertional mutagenesis screens in haploid ESCs.

e, Representative Rex1-GFP FACS plots showing the differentiation delays 30h after 2i withdrawal of *Tcf7l1*, *Trim71* and *Mbd3* KOs. A Myc KO served as a negative control. Blue indicates the KO FACS profiles and dashed lines indicates WT Rex-GFP.

f, Differential expression of genes at N24 vs. 2i in WT RC9 cells. Black dots show significance (FDR ≤ 0.05, H0: |FC| > 1.5). Pluripotency genes are red dots, haploid screen hits are blue dots, the 73 KO-genes are green dots.

g, t-SNE projection of the 73 KOs in 2i, based on expression of 3058 differentiation-associated genes ($RC9^{N24}$ vs. $RC9^{2i}$, FDR ≤ 0.05, H0: |FC| > 1.5). Differentiation defects observed at N24 in the respective KOs are indicated by a colour gradient, measured as average naive marker

617 log2fold change (based on expression levels of *Esrrb, Nanog, Tfcp2l1, Tbx3, Prdm14* and *Klf4*)
618 in the respective KO at N24. Red: delayed differentiation; blue: accelerated differentiation -
619 see key in **h,**.
620 **h,** Similar to **g,** for KOs at N24
621
622 **Supplementary Figure 1**
623 **a,** The top 15 candidate genes from the haploid screen ordered by the number of times the
624 hit was found in the 35 screens, including the number of independent integration sites and
625 the calculated FDR.
626 **b,** Significantly enriched pathways among screen hits ranked according to fold enrichment;
627 colours represent FDR in screen analysis. Numbers next to bars indicate number of hit-genes
628 within category,
629 **c,** Significantly enriched GO terms among screen hits as in **b,**
630 **d,** Workflow to generate Cas9 knockouts in RC9 ESCs.
631 **e,** Idealized strategy illustrating gRNAs and genotyping primers.
632 **f,** Western analysis using indicated KOs and indicated antibodies. Tubulin or Gapdh were used
633 as loading controls, as indicated.
634 **g,** Anti-flag specific Westerns in indicated KO^{rescue} ESCs upon stable forced expression of 3xflag
635 rescue cDNAs driven from a CAG promoter.
636 **h,** Rex1GFP levels measured by FACS showing restoration of differentiation behaviour in the
637 indicated rescue cell lines at N30.
638 **i,** t-SNE visualization of RNA-seq profiles in 2i and at N24 before and after batch correction.
639 **j,** GO enrichment analysis of the top 10% of genes (by absolute logFC) that were differentially
640 expressed between 2i and N24 in WT ESCs (FDR ≤ 0.05, H0: |FC| < 1.5).

641 **k,** Exit from pluripotency factors showing significant expression changes between 2i and N24
642 (lfc≥0.8; adj. p-value≤0.05). Log2FC, adj. p value and quantification of the molecular
643 phenotype using the average upregulation of a set of naïve marker genes (*Esrrb, Nanog,*
644 *Tfcp2l1, Tbx3, Prdm14, Klf4*) are shown.
645 **l,** Dot-plot showing regression analysis of KO-induced changes at N24, comparing *Tcf7l1* KO
646 to Apc and Csnk1a1 KOs. Total least squares (TLS) analysis results are indicated.
647 **m,** Rex1-GFP levels measured by FACS at N24 showing the effect of Epiblastin A addition
648 during differentiation.
649 **n,** Rex1GFP analysis at N24 of *Pum1* KO cells, showing a relatively strong differentiation defect
650 at early passages.
651
652 **Figure 2**
653 **a,** Naïve (top) and formative (bottom) marker gene expression changes at N24 compared to
654 WT in all 73 KO ESCs. Clustering based on naïve marker gene expression shows a wide
655 distribution of differentiation delay phenotypes.
656 **b,** Rex1GFP FACS analysis of *Jarid2* and *Tcf7l1* KO ESCs after transfection with negative control
657 or *Klf2* specific siRNAs.
658 **c,** For each of the 73 KOs, the change of mean naïve marker gene expression at N24 (KO vs
659 WT) was plotted against the change of the global transcriptome at N24 state (correlation of
660 KO-induced change at N24 with differentiation-induced changes in WT differentiation, N24 vs
661 2i).
662 **d,** Comparison of number of genes deregulated in 2i and at N24 in all KOs (FDR ≤ 0.05, H0:
663 |FC| < 1.5). Differentiation phenotypes are colour coded according to mean naïve marker
664 gene expression change in KOs at N24.

665 **e,** Rex1GFP FACS analysis of WT, *Jmjd1c* and *Tcf7l1* KO cells at N48 cultured with and without
666 the Gsk3 inhibitor, CH.
667
668 **Supplementary Figure 2**
669 a, RNA-seq based fold changes (FC, relative to WT) of indicated naïve marker genes in
670 indicated KOs in 2i (only FCs with adj. $p \leq 0.05$ are shown).
671 b, RNA-seq derived data showing FCs of indicated formative marker genes in indicated KOs in
672 2i (only FCs with adj. $p \leq 0.05$ are shown).
673 c, Rex1-GFP analysis of WT cells at N24, transfected with negative control or *Klf2* specific
674 siRNAs. *Klf2* depletion increases the speed of Rex1GFP downregulation.
675 **d,** Comparison of the number of differentially expressed genes at N24 (FDR ≤ 0.05, H0: |FC|
676 < 1.5) to the mean naïve marker gene expression at N24 (phenotype strength). Stronger
677 differentiation phenotypes correlate with more differentially expressed genes.
678 **e,** Expression log2 fold changes (log2FC) at N24 between knockout and RC9 control,
679 comparing *Tcf7l1* KOs to *Jmjd1c* KOs. Each dot corresponds to one gene. Only genes showing
680 significance (FDR ≤ 0.05, H0: |FC| < 1.5) in either one of the KOs are plotted. Red line: total
681 least square regression; regression coefficients are shown. Convergence on similar target
682 genes is shown by largely identical directionality of regulation of significantly regulated genes
683 in both KOs.
684
685 **Figure 3**
686 **a,** Scheme illustrating comparisons of the 73 KO gene expression profiles at N24 with a 2h-
687 resolution *in vitro* differentiation time course of WT RC9 ESCs.

688 **b,** Heatmap showing differentiation delays of the 73 KO lines quantified by average naïve
689 marker gene expression. Red bars indicate closest correlation to that specific time-point.
690 Positive values indicate delayed differentiation. Negative values more rapid differentiation.
691 Each line corresponds to one KO. Selected KOs are indicated. See Supplementary Table 2 for
692 the full hierarchy of the 73 KOs.
693 **c,** Plot comparing the differentiation delays calculated using naïve markers as in **b,** or using all
694 3058 genes changed in WT differentiation from 2i to N24.
695
696 **Supplementary Figure 3**
697 **a,** Principle component analysis (PCA) of the 2h resolved WT differentiation time course
698 sampled by RNA-seq. 0h corresponds to cells in 2i.
699 **b,** Heatmap showing differentiation delays of the 73 KO lines quantified using expression of
700 all 3058 genes differentially expressed in WT between 2i and N24. Red bars indicate closest
701 correlation to that specific time-point. Positive values indicate delayed differentiation.
702 Negative values more rapid differentiation compared to WT. Each line corresponds to one KO.
703 Selected KOs are indicated.
704
705 **Figure 4**
706 **a,** Scheme illustrating the comparison of gene expression profiles of 73 KOs in 2i and at N24
707 with *in vivo* derived averaged single cell expression data comprising peri-implantation epiblast
708 tissues. Specifically, transcription profiles of each KO in 2i and at N24 were compared with in
709 expression datasets obtained from the E4.5, E5.5 and E6.5 epiblast.
710 **b,** Fraction of identities (FOI) between 2i and N24 expression data and mouse E4.5, E5.5 and
711 E6.5 epiblast, computed using all expressed genes (log2FPKM > 0) for all KOs. Each dot

712 represents one KO. Average naïve marker expression in KOs at N24 is indicated by a colour
713 gradient.
714 **c,** Plot showing a comparison of FOIs of WT (RC9) and KO cells in 2i (green) and at N24 (grey)
715 to E4.5 and E5.5 mouse epiblast, computed using all expressed genes (log2FPKM > 0).
716 **(d,e)** Data shown in Figure 4c separated by culture condition. Comparison of FOIs of WT and
717 KO cells in 2i **d,** and at N24 **e,** to E4.5 and E5.5 mouse epiblast; average naïve marker
718 expression in KOs at N24 is indicated by a colour gradient.
719 **f,** PCA showing all KOs in 2i (triangles) and at N24 (diamonds) and embryo derived datasets
720 (ICM, PrE, E4.5Epi, E5.5Epi, E6.5Epi). As in **d,**, blue to red colour gradient indicates naïve
721 marker expression at N24 of the respective KO.
722 **g,** Similar to Figure 4b, showing FOIs between 2i and N24 datasets and macaque pre- and
723 post-implantation embryos.
724 **h,** Similar to Figure 4b, showing FOIs between 2i and N24 datasets and hESC capacitation
725 datasets comprising hESCs in naïve conditions (day0) and hESCs 20 days after release into
726 primed conditions (day20).
727
728 **Supplementary Figure 4**
729 **a,** Zoom in of the plot in Figure 4c, focusing on the overlap of several KOs in 2i (green) and in
730 N24 (grey). N24 samples of *Etv5, Smg7, Tsc2, Smg5* and *Zfp281* KOs show FOIs to the E4.5 EPI
731 similar to KOs cultured in 2i.
732 **b,** Comparison of FOIs to the E4.5 epiblast of all KOs in 2i compared to FOI of all KOs at N24;
733 average naïve marker expression in KOs at N24 is indicated as colour gradient. With some
734 exceptions, stronger *in vitro* phenotypes translate to increased similarity to the E4.5 epiblast
735 in 2i and at N24.

736 **c,** FOIs of KOs in 2i to ICM, E4.5 trophectoderm (TE), E4.5 primitive endoderm (PrE), E4.5
737 epiblast (Epi), E5.5 Epi and E 6.5 Epi. Analysis extended but similar to Figure 4b.
738 **d,** Plot comparing FOIs of 2i samples to E4.5 epiblast (x-axis) to FOIs of N24 samples to E5.5
739 epiblast(y-axis). Dotted grey lines indicate 95% confidence intervals. Colour gradient shows
740 similarity of 2i expression profiles to PrE.
741 **e,** Correlation plot of FOIs between 2i profiles and mouse naïve (E4.5) epi and FOIs between
742 2i profiles and macaque naive epi (correlation significance is indicated).
743 **f,** Correlation plot of FOIs between 2i profiles and mouse naive epi and FOIs between 2i
744 profiles and human naïve ESCs (day0) (correlation significance is indicated).
745 **g,** Correlation plot of FOIs between N24 profiles and mouse E5.5 epi and FOIs between N24
746 profiles and macaque post-implantation epi (correlation significance is indicated).
747 **h,** Correlation plot of FOIs between N24 profiles and mouse E5.5 epi and FOIs between N24
748 profiles and human ESCs at day 20 after release into primed culture conditions (correlation
749 significance is indicated).
750
751 **Figure 5**
752 **a,** x-axis displays all 3rd dimension (Dim3) genes, which separates mouse naïve from post-
753 implantation epiblast (the inset shows the full PCA plot, see also Supplementary Figure 5d) in
754 order of extent of contribution to Dim3. The contribution to Dim3 of NAGs (orange) and the
755 top 250 differentially regulated genes (green) are plotted in a cumulative manner.
756 **b,** Plots comparing relative log2FC in mouse ESC differentiation (2i vs N24) to relative log2FC
757 in mouse *in vivo* transition from pre- to post-implantation epiblast (E4.5 vs E5.5). Selected
758 gene groups are plotted: i) NAGs, ii) all 3058 differentially expressed genes in mESC
759 differentiation and iii) the top 250 up and downregulated genes in mESC differentiation. Rho

760 values indicate the level of correlation between *in vivo* and *in vitro* differentiation of given
761 gene groups. Alpha indicates the angle between the x-axes and the orthogonal regression.
762 Genes with log2FC > 0.5 and < -0.5 in both the x- and y-axes are highlighted in orange and
763 blue respectively.
764 **c,** As for **b,** except plots comparing log2FCs in mouse ESC differentiation (2i vs N24) to
765 macaque *in vivo* transition from pre- to post-implantation epiblast cells.
766 **d,** As for **b,** except plots compare log2FCs in mouse ESC differentiation (2i vs N24) to human
767 naïve ESC capacitation (d0 to d20).
768 **e,** Graphs showing the frequency of co-regulation of NAGs or the top 250 differentially
769 expressed genes (2i vs N24) with differentially expressed genes in mouse and macaque *in vivo*
770 pre- to post-implantation epiblast differentiation, and human naïve ESC capacitation. Two-
771 tailed chi-square test (CI 95%) was used to compute significance levels between the expected
772 and the observed number of modulated genes.
773 **f,** One-way hierarchical clustering showing expression changes of NAGs (left) and top
774 differentially expressed genes (right) in the indicated differentiation models.
775
776 **Supplementary Figure 5**
777 **a,** Expression change of NAGs during naïve to formative differentiation in WT ESCs (y-axis)
778 versus the strength of link to the naïve network (naïve association, x-axis). The orange line
779 indicates the linear regression. The R^2 value is indicated in the plot.
780 **b,** *Klf2* expression at the indicated stages. Compacted morula (cMOR), expanded ICM (eICM),
781 E4.5 epiblast (EPI) derived from single cell RNA-seq datasets from human, marmoset and
782 mouse preimplantation development.

783 **c,** Plots derived from the GRAPPA visualisation-app showing expression in yellow of *Pdgfa* and
784 its cognate receptor *Pdgfra* during mouse preimplantation embryo development.
785 **d,** Principal component analysis of published single cell RNA-seq datasets showing that
786 Dimension 3 (Dim 3) separates pre- from post-implantation epiblast *in vivo* (as shown in Figure
787 5a).
788
789 **Figure 6**
790 **a,** Relationship between the KO expression profiles at N24 (or in 2i for the LIF pathway) to the
791 expression footprints of the five signalling pathways as indicated (see also Supplementary
792 Figure 6a). Relative pathway activity is -1 for Fgf/ERK (defined by Fgfr1 & Ptpn11 KOs), Wnt
793 (Tcf7l1 KO) and Notch (Rbpj KO) and +1 for mTOR (Tsc2) and LIF (defined by presence of LIF
794 in 2i medium) for effects equally strong as in the representative KO. Pathway activity is
795 indicated by colour code (green for activity, red for inactivity). All tiles with an absolute
796 pathway activity smaller 0.3 are grey. The size of the tiles corresponds to the spearman
797 correlation of pathway footprints between each KO and the corresponding representative
798 KOs. The colour gradient on the sidebar at the left indicates mean change of naïve marker
799 gene expression (as a measure of phenotypic strength) for the corresponding KO at N24.
800 **b,** Heatmap showing average log2FCs (KO^{N24} vs. $RC9^{N24}$) of constitutive response gene
801 clusters. Asterisks indicate that the absolute row-wise z-value of the respective mean
802 expression value was above 1.96 ($p \leq 0.05$), indicating extreme values within the context of
803 that row. For reference, average naive marker log2FCs at N24 vs WT are shown in a separate
804 row below (similar to Figure 2a).
805 **c,** GO enrichment analysis per 'constitutive KO response' cluster. Dot-size scales with log2
806 enrichment. Adj. p vales are clour coded in shades of red.

807 **d,** As for **b,** except heatmap showing average log2FCs (KO^{N24} vs. $RC9^{N24}$) of N24 induced KO
808 response gene-clusters.
809 **e,** As for **c,** except GO enrichment analysis per 'N24 induced KO response' cluster.
810
811 **Supplementary Figure 6**
812 **a,** Schematised computational strategy to derive non-overlapping gene-expression footprints
813 for the five signalling pathways.
814 **b,** Average change in naive marker expression induced by KOs at N24 (y-axis) and their 'LIF-
815 likeness' in 2i (x-axis; correlation of RC9 LIF response with knockout induced response in 2i).
816 Overall correlation across these values is r=0.69. The KOs showing the strongest correlations
817 (*Tsc2*, *Ptpn11* and *Trim71* KOs) are indicated in red.
818 **c,** Flow analysis showing delayed Rex1-GFP downregulation in cells starting from cultures in
819 2i/LIF compared to cells starting from 2i.
820 **d,** Multiple regression coefficients of pathway activity profiles across all KOs when predicting
821 the differentiation delay phenotype (defined by mean naïve marker gene expression change
822 at N24). Error bars indicate the standard error of the coefficient.
823 **e,** Euler diagram showing the overlap of defined gene groups. Panels on the right show
824 exemplary gene expression behaviours. The constitutive and the N24 induced KO response
825 group of genes were used for further analysis.
826 **f,** Percentages of genes in the N24 induced KO response (N24 induced) or constitutive KO
827 response (constitutive) clusters that significantly change during WT differentiation (FDR ≤
828 0.05, H0: |FC| < 1.5).
829
830

831 **Figure 7**

832 **a,** Visualization of connections between pathways (central ellipses) and gene clusters
833 (rectangles) based on multiple regression models. The colour gradient of the clusters indicates
834 enrichment for NAGs. Dashed connections link pathways to clusters that fit time-course
835 validated regression models (see Supplementary Figure 7A and C). Continuous connections
836 are further supported by ChIP or pathway defining gene enrichments (see Supplementary
837 Figure 7E,F). Connections are coloured by source pathway. Unconnected clusters are listed at
838 the right.

839 **b,** Summary model of the genetic and signaling pathway control of naïve to formative
840 transition. Between the naïve and formative gene regulatory networks, the five signaling
841 pathways largely operate through partially overlapping separable gene clusters.

842

843 **Supplementary Figure 7**

844 **a,** Heatmap illustrating multiple regression models of pathway activities predicting cluster
845 expression. Only significant (adj. p. ≤ 0.01) interactions are shown. Colour intensity represents
846 the strength of the interaction. Red tiles indicate that higher activity of a certain pathway
847 results in higher expression of the cluster, while blue tiles indicate the opposite.

848 **b,** Pathway activities (average expression of the 50 specific pathway-defining genes) across a
849 WT differentiation time-course for all 5 pathways. y-axis shows average log2FC of pathway
850 defining genes relative to the mean expression throughout the time-course.

851 **c,** Observed cluster expression change (red) and cluster expression change predicted by
852 regression models (blue) throughout the 32h WT differentiation time course for all induced
853 clusters. Predicted cluster expression is calculated using pathway activities over the WT time
854 course and significant (adj. p ≤ 0.01) interactions from the multiple regression models. Graphs

855 show log2FC of cluster expression relative to the mean cluster expression across the time
856 course. Red cluster names indicate that cluster-pathway connections were not predicted in
857 Supplementary Figure 7A or validation through 2h time-course expression failed (predicted
858 expression did not precede observed expression).
859 **d,** Similar to Supplementary Figure 7C for constitutive clusters.
860 **e,** Enrichment from Fisher test of ChIP-targets, NAGs and pathway defining gene sets in
861 induced clusters. Enrichment is indicated by red circles, depletion by blue squares. Size
862 indicates strength of over- or underrepresentation and colour gradient encodes adj. p values.
863 **f,** Similar to Supplementary Figure 7E for constitutive clusters.

Methods

Cloning

Guide RNAs (gRNAs) were designed using the computational pipeline described below, or by http://crisprscan.org. To clone the gRNAs for ESC KO of the 154 selected genes, we used RecET recombineering to insert the guide RNA protospacer encoding sequence between the U6 promoter and the universal sgRNA scaffold in pBR322-U6-cm-ccdB-sgRNA-amp, as described previously (Baker et al, 2016). We generated 331 sgRNA expressing plasmids for 149 genes (2 sgRNAs for 131 genes, 3 sgRNAs for 5 genes, 4 sgRNAs for 12 gens and 6 sgRNAs for 1 gene). In brief, the plasmid was linearized at the point of insertion by BstZ17I digestion, purified using the Invitrogen Charge Switch PCR Purification Kit and dissolved in water; 80-mer oligonucleotides containing the protospacer sequence, flanked by 30bp of homology to the target plasmid were dissolved in water. *E. coli* GB05 transformed with the pSC101-Prha-ETgA-tet plasmid [63], was cultured to OD600 ~0.5 in deep-well 96 well plates and then induced for 1 hour for expression of the ETgA operon with L-rhamnose. Electroporation with 200 ng (0.1 pmol) of the linearised vector and 50 pmol of oligonucleotide, using BTX 96 well electroporation plates (MOS96, 2mm gap,) and the BTX ECM630 electroporator with HT-200 adapter, was described previously [64]. After one hour of recovery time, 100 microliters (1/10th) of the culture was transferred to a new plate with 900 microliters of LB media plus 100 micrograms/ml ampicillin. After overnight selection the saturated cultures were plated for single colonies on L-agar + 100 micrograms/ml ampicillin. A single colony for each construct was picked and the clones were grown under selection in 96 well plates, plasmid DNA was isolated, and Sanger sequenced from both directions using the M13F and M13R primers (flanking the U6-protospacer-sgRNA scaffold cassette, Supplementary Table 6). Sequence analysis confirmed the correct engineering of 316 out 331 plasmids. After sequencing of a

888 second colony or repeat cloning for the remaining 15 clones, all the desired constructs were
889 obtained. Alternatively, annealed oligonucleotides (Supplementary Table 6) were cloned into
890 a BsaI site of a gRNA expression vector (Addgene Plasmid number 41824) [65] and correct
891 insertion was determined by Sanger sequencing with the SP6 primer (Supplementary Table
892 6). For generating rescue cell lines, indicated coding sequences were cloned into a pCAG-
893 3xFLAG-empty-pgk-hph vector [14] after PCR amplification and correct insertion was detected
894 by restriction digest and Sanger sequencing with the 3xFlag_seq primer (Supplementary Table
895 6).
896
897 **Cell culture**
898 Diploid (biparental) ES cells were routinely cultured in gelatin coated 25cm^2 flasks in DMEM
899 supplemented with 15% FCS (batch tested, Biowest), 1x Penicillin-Streptomycin (Sigma),
900 0.1mM NEEA, 1mM sodium pyruvate, 1mM L-glutamine, 0.05mM 2-mercatpoethanol
901 (Gibco), 10ng/mL LIF (batch tested, in-house), 1.5µM CHIRON and 0.5µM PD0325901 (termed
902 ESDMEM-2i) on gelatin coated plates. The basal medium during all differentiation assays and
903 for haploid ES cell culture (N2B27) consisted of a 1:1 mixture of DMEM/F12 and Neurobasal
904 medium supplemented with 1x B27 (Gibco), 0.5x N2 (homemade), 0.1mM NEAA, 1mM L-
905 glutamine, 1x Penicillin-Streptomycin and 0.05mM 2-mercaptoethanol. Haploid ES cells were
906 routinely cultured in N2B27 supplemented with 3µM CHIR99201, 1µM PD0325901 and
907 10ng/ml LIF and regularly sorted based on FSC/SSC parameters by FACS [66]. Haploid
908 Rex1GFPd2-IRES-BSD (hRex1GFPd2, [67] and haploid Rex1::mKO2-IRES-BSD/pOct4-GFP-IRES-
909 Puro ESCs (generated in this study from a double reporter mouse) were derived from
910 activated oocytes in 2i/LIF medium and used for screens. Biparental diploid ES cells with a GFP
911 driven by the endogenous Rex1 promoter and EF1a driven Cas9 targeted to the Rosa26 locus

912 (RC9 cells, [13] served as parental cell line for the KO lines generated in this study. A list of KO
913 cell lines can be found in Supplementary Table 6.
914
915 **Transposon-based saturation screen for key players of the exit from pluripotency**
916 The genetrap vectors 5'-PTK-3', TNN, TNP (Horie et al. 2011, Allan Bradley Lab) together with
917 haploid Rex1::GFPd2 ES cells (hRex1GFPd2, Leeb et al. 2014) and haploid Rex1::mKO2-IRES-
918 BSD/pOct4-GFP-IRES-Puro ES cells (hOR, this study) were used in 35 independent experiments
919 to drive the screen to saturation. Mutant pools were generated by electroporation of 10^7
920 haploid ESCs using a GenePulser Xcell (270 V, 500 µF, ∞ Ω, Biorad) with 0.5 µg genetrap
921 plasmid and 10 µg hyperactive transposase (hyPBase) [68]. 24h after electroporation, selection
922 was started for 4 days using 1µg/ml Puromycin. Thereafter cells were plated at a density of
923 10^4 cells/cm^2 in N2B27 medium in the absence of LIF or inhibitors to allow differentiation.
924 After 7–10 days in differentiation conditions, GFP-positive cells were sorted and replated at a
925 density of 10^4 cells/cm^2. After culture in N2B27 medium for a further 7–10 days, GFP-positive
926 ESCs were resorted, expanded for 48h in 2i/LIF medium and DNA was isolated using the
927 PureGene kit (Qiagen) according to the manufacturer instructions. Sequencing libraries were
928 prepared using an optimized Splinkerette PCR protocol (Leeb et al. 2014) with a modified and
929 improved set of adapters and primers (see Supplementary Table 6). Adapters were annealed
930 at 50µM each in T4 DNA ligase buffer by incubating at 97.5°C for 150 sec followed by a
931 temperature decrease of 0.1°C/5sec for 775 cycles. Ready-to-use adapters were stored at -
932 20°C. Genomic DNA was isolated using the PureGene kit (Qiagen) according to the
933 manufacturers recommendations and quantified on the Qubit 2 fluorometer (Life
934 Technologies) using the Qubit dsDNA Broad Range Assay (Life Technologies, Q32853). 2µg of
935 genomic DNA was diluted to a total volume of 120µl with Low TE buffer (10mM Tris-HCl, pH

936 8.0, 0.1mM EDTA) and sheared to 250bp on a Covaris S-series Sample Preparation System
937 with the following settings: Duty Cycle – 20%, Intensity – 5, Cycles per Burst – 200, Time – 60
938 seconds, Temperature – 4° to 7°C. After shearing DNA was isolated with the QIAquick PCR
939 purification kit (Qiagen) and quality assessed using the Agilent Bioanalyzer with the DNA high
940 sensitivity kit (Agilent). End-repair was performed using the NEBNext End Repair Module
941 (E6050S) by mixing 100µl sheared DNAl, 11.6µl buffer and 6µl enzyme followed by incubation
942 at 20°C for 60min. After cleanup with the PureLink PCR Purification Kit (K3100-01), 42µl of the
943 end repaired DNA was A-tailed with the NEBNext dA-Tailing Module (E6053S) with 5µl buffer
944 and 3µl enzyme at 37°C for 60 minutes. Following PureLink purification of to the A-tailed DNA,
945 the annealed adapters were ligated using theNEBNext Quick Ligation (E6056S) kit at 20°C for
946 30min. Products were cleaned up using Ampure XP beads at a 1:1 ratio. Libraries were PCR
947 amplified using Kapa HIFI (Roche) with 10µM PB5_pr1 and 10µM SplAP1 during 18 cycles of
948 20sec 98°C, 20sec 63°C and 40sec 72°C with an initial denaturation at 95°C for 2minutes and
949 a final extension at 72°C for 5min. Samples were cleaned up using Ampure XP beads at a ratio
950 of 0.8. A second round of PCR was conducted using xxul from PCR1 using the Kapa HIFI
951 polymerase with 10µM SplAP2 and 10µM index primer for 13 cycles at an annealing
952 temperature of 60°C with all other settings as in PCR1. Final libraries were quantified using
953 the KAPA SYBR Fast qPCR Mix with primers SYB7 and SYB5 at 10µM each and a final quality
954 control was conducted with the Agilent Bioanalyzer DNA High Sensitivity Assay.I ntegration
955 sites were mapped as in [12] and described below.
956
957 **Generation of KO ESCs**
958 Direct comparison between KOs necessitates a similar culture and passaging history. We
959 therefore established an experimental pipeline to parallelise the generation of multiple KO

960 ESC lines per week. $2x10^5$ RC9 cells were transfected in a 6-well format with a set of two gRNA
961 containing vectors (Supplementary Table 6; (1µg each), together with 0.5µg of pCAG-dsRed
962 by Lipofectamine2000 (3µl per reaction). 12-16h after transfection, medium was replaced. To
963 enrich for transfectants, cultures were sorted for GFP/DsRed double positive cells on a BD
964 FACS Aria III 48h after transfection and plated at clonal density in 6cm dishes (100 to 800 cells
965 per dish) in ESDMEM2i. Approximately one week after sorting, 48 colonies per gene KO were
966 picked into 96-well plates, trypsinized and split into one expansion-plate and one plate for
967 PCR genotyping after boiling lysis and proteinase K treatment (see PCR genotyping).
968 Identified KO clones were expanded to 12-well or 6-well plates after 3 days of growth and the
969 remaining plate was frozen in 50%FCS/10% DMSO as a backup. Clones were frozen from 12-
970 or 6-well plates in freezing medium containing 50% FCS and 10% DMSO. Of the attempted
971 154 gene disruptions, we achieved 115 homozygous knockout cell lines. In 16 cases, only
972 heterozygous clones were obtained (indicating essentiality of these genes), and in 23 cases
973 we were unable to recover gene disruptions, despite multiple attempts (Supplementary Table
974 6).
975
976 **PCR genotyping**
977 Crude DNA lysates were generated by pelleting half of a picked colony in a PCR plate at 500g
978 at 4°C, followed by two washes with PBS. 25µl of water were added and pellets were boiled
979 for 5min at 95°C. After cooling, Proteinase K was added at a final concentration of 3µg/µl and
980 incubated at 65°C for 1h followed by inactivation at 95°C for 10min. Successful KO generation
981 was confirmed by PCR, employing a three-primer strategy. Specific reverse primers for the
982 deletion event and a possible wildtype allele were used in combination with a common
983 forward primer (Supplementary Table 6). Cell lines with indels caused by a NHEJ event around

984 the exonic gRNA1 target site, were detected by Sanger sequencing of the “wildtype”-band.
985 Verification of the KO-events was performed by the same PCR strategy on DNA, isolated using
986 the PureGene DNA isolation kit (Qiagen) according to the manufacturer recommendations.
987 For genotyping PCR, we used OneTaq (NEB) or JumpStart RedTaq (Sigma) PCR Master Mix
988 according to the manufacturers recommendations for cycling and a standard annealing
989 temperature of 55°C and 35-40 cycles.
990
991 **Immunoblotting**
992 Whole cell fractions were isolated using RIPA buffer and protein concentrations were
993 determined using a Bradford Assay (Bio-Rad). 20µg whole cell lysate were separated on 8-
994 12% SDS-PAGE gels (depending on the molecular weight of the target proteins), and
995 subsequently blotted on 0.2µm nitrocellulose membranes (Amersham). Membranes were
996 blocked in TBS or PBS containing 5% milk and 0.1% Tween-20 or 3% BSA and 0.1% Tween 20
997 for 1h. Primary antibodies were incubated overnight at 4°C, at a dilution of 1/1000 with the
998 following exceptions: Anti-Csnk1a1 1/200, Anti-Flag M2 1/1000 Anti-Gapdh 1/25000, Anti-
999 Tubulin 1/5000, Anti-Smg5 1/200, Smg7 1/2000. Secondary goat anti-mouse IgG HRP
1000 (1:10000) or goat anti-rabbit IgG HRP (1:15000) were incubated at RT for 1h in blocking
1001 solution. Antibody binding was detected using the ECL select detection kit (Amersham). A list
1002 of antibodies can be found in Supplementary Table 6.
1003
1004 **Parallel Differentiation**
1005 ES Cells were differentiated in seven batches, always including WT controls, control KO ESCs
1006 (no phenotype expected) and a range of ESCs with expected weak to strong phenotypes.
1007 Where applicable two independent KO clones were used, other KO ESCs were cultured in

1008 parallel replicates before intiation of differentiation. 24h before starting the differentiation
1009 assay, medium was changed from ESDMEM2i to N2B27 based 2i/LIF medium. Cells were then
1010 trypsinized, counted and plated in 2i (without LIF) at a density of 10^4 cells/cm^2 in two replicate
1011 wells of a 6-well plate for differentiation, one 6-well as undifferentiated control and 2 wells
1012 of a 12-well plate (undifferentiated control and differentiated sample) to monitor
1013 differentiation by FACS. 12h after plating, cells were washed twice with PBS and medium was
1014 changed to unsupplemented N2B27. For the undifferentiated controls medium was replaced
1015 with fresh 2i. 24h later, at N24, cultures were harvested in RLT buffer and stored at -80°C
1016 before isolation of RNA using the RNeasy Mini Kit (Qiagen). mRNA was isolated from 1 ug total
1017 RNA by poly-dT enrichment using the NEBNext Poly**a,** mRNA Magnetic Isolation Module
1018 according to the manufacturer's instructions. Samples were eluted in 15ul 2x first strand
1019 cDNA synthesis buffer (NEBnext, NEB). After chemical fragmentation by incubating for 15 min
1020 at 94°C the sample was directly subjected to the workflow for strand specific RNA-Seq library
1021 preparation (Ultra Directional RNA Library Prep, NEB). For ligation custom adapters were
1022 (Supplementary Table 6). After ligation, adapters were depleted by an XP bead purification
1023 (Beckman Coulter) adding beads in a ratio of 1:1, followed by an index PCR (15 cycles) using
1024 Illumina compatible index primer. After double XP beads purifications (with beads added in a
1025 ratio 1:1) libraries were quantified on a Fragment Analyzer run with a NGS Assay Kit (Agilent)
1026 and loaded on a HiSeq 3000 flowcell with 50 cycles single end sequencing by pooling the
1027 samples based on molarity aiming for 30 mio. reads per sample.
1028
1029 **2h resolved differentiation timecourse**
1030 Medium was changed to N2B27-2iLIF 24h before the assay. Cells were trypsinized, counted
1031 and plated at a density of 10^4 cells/cm^2 in 6-well plates in 2i (without LIF). 12h later medium

1032 was changed to unsupplemented N2B27 (or 2i for the undifferentiated controls) and cells
1033 were harvested in buffer RLT every 2 hours for the next 32h (N32). RNA was isolated using
1034 the RNeasy Mini Kit (Qiagen). All libraries were generated from 500ng RNA using the TruSeq
1035 Stranded mRNA Library Prep Kit (Illumina) and analysed on a HiSeq4000, Paired End 75bp.
1036

1037 **LIF signature**

1038 $1x10^{4c}$ells/cm^2 were plated in N2B27-2i and N2B27-2iLIF and grown for 24h after which RNA
1039 was isolated using the RNeasy Mini Kit (Qiagen). Libraries were made from 1µg high quality
1040 RNA using the Quantseq 3' mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen) according
1041 to the manufacturer's instructions. Libraries were quantified using qPCR according to the
1042 Quantseq Library Prep Kit and with the included primers, pooled and quality checked using
1043 an Agilent Bioanalyzer DNA High Sensitivity Assay (Agilent). Multiplexes were sequenced on
1044 a HiSeq4000, Single End 50bp.
1045

1046 **RNAi experiments**

1047 For RNAi FlexiTube siRNAs against Klf2 were used with AllStars Negative Control siRNAs
1048 (Qiagen). 20ng siRNAs/$4x10^4$ cells were transfected in 2i using DharmaFect (Dharmacon). 12h
1049 later medium was changed to N2B27 after two PBS washes. At N24 and N32 Rex1GFPd2
1050 expression was determined by flowcytometry (see Flowcytometry Analysis) and cells were
1051 harvested in RLT buffer and isolated with the RNeasy mini kit (Qiagen) according to the
1052 manufacturer's recommendations. cDNA was transcribed using the SensiFAST cDNA Synthesis
1053 Kit (Bioline). Naïve and primed marker expression was determined by qPCR using the Sensifast
1054 SYBR No Rox-Kit (Bioline). Primers are listed in Supplementary Table 6.
1055

1056 **Flow cytometry**

1057 Cells were dissociated in 0.25% trypsin/EDTA and trypsin was neutralized with DMEM
1058 supplemented with 15% FCS. After passing through a 40µm mesh, Rex1-reporter activity was
1059 measured using a BD Fortessa machine. High-throughput-measurements were performed in
1060 96-well plates using the HTS unit of the BD Fortessa. Data were analyzed using FlowJo
1061 software.

Data Analysis

Transposon Mutagenesis Screen in Haploid Murine Embryonic Stem Cells

Integration mapping was performed as previously described [12]. Most gene-trap insertions were found in intronic sequences, consistent with the relative length distributions of genomic features. However, after normalising to available TTAA sites per region, TTAA sites in 5'UTRs and promoters (<-500bp) were overrepresented, whereas TTAA sites in introns were relatively depleted. 7760 independent integrations passing cut-off criteria were mapped to 3469 genes. Of those, 232 Genes (2474 independent integration sites) were hit in five or more independent screens, indicating that those are integrations causative for the detected differentiation defect. To complement the candidate list and correct for gene specific biases we further developed a strategy to assign statistical strength to candidate genes by assessing the number of integrations per gene in relation to available TTAA transposon integration sites We utilized the number of different TTAA-sites within each gene to calculate the probability of having integrations in *k* or more different locations by chance. All TTAA sites in mm10 were assessed using bowtie and the analysis was restricted to genes classified as protein coding, long non-coding RNAs and micro RNAs. We computed the probability that a gene is hit *k* times given that it contains *n* TTAA transposon integration sites assuming a binomial distribution (while *k* is the observed number of integrations).

$$P[X \geq i] = \sum_{i=k}^{n} P[X = i] = \sum_{i=k}^{n} B(i|p_0, n) = \sum_{i=k}^{n} \binom{n}{i} p_0^i (1 - p_0)^{n-i}$$

p_0 was then calculated by $\frac{All\ hit\ TTAA-sites\ in\ genes}{All\ TTAA-sites\ in\ genes}$

1082 We applied this test to genes with at least 26 TTAA sites, as genes with fewer TTAA sites would
1083 be significant (p-value < 0.05) with a single integration. Candidate gene lists between both
1084 analyses showed overlapping results and top hits were largely identical. This resulted in a list
1085 of 421 significant genes (adjusted p-value < 0.01 and hit in at least two independent screens).

1086 **Guide RNA design criteria**

1087 To automate the design of sgRNAs as far as possible we used the R package CRISPRseek (Zhu
1088 et al., 2014) to score the efficiency and likely off-target binding of candidate sgRNA sequences.
1089 For a selected gene the algorithm proceeds as follows:
1090

1091 1. Retrieve the sequence of the first protein coding exon
1092 2. Sequence has to be at least 50 bases long
1093 3. Find sgRNAs within the sequence and score them.
1094 4. If no appropriate sgRNA could be found, move to next exon
1095 5. Still in the first third of the coding sequence?
1096 6. Find a second sgRNA in an intron at least 5kb away, but no more than 30kb
1097 7. Score the sgRNAs

1098 A minimal efficiency of 0.2 and an off-target score < 100 was required.

1099 **PCR Primer Design Strategy for Simplified Knock-out Validation**

1100 The surrounding regions of the sgRNAs (min distance 20bp) were used as input for PRIMER3
1101 primer predictions. Primers were designed to have a length of 18-27 bases with an optimum
1102 of 20 and to have melting temperatures between 57°C and 63°C with an optimum of 60°C. All

1103 primers were checked for off-target binding sites by using Blastn. Blastn was called with the
1104 parameters:
1105 -task megablast -use_index false -word_size 7 -perc_identity 65 –evalue 30000 -
1106 max_target_seqs 50 -num_threads 4 –outfmt 7.

1107 **RNA quantification, RNAseq**

1108 Quality control was performed using fastQC (version 0.11.5). Transcripts were mapped to the
1109 mm10 mouse reference genome and counts determined using STAR (version 2.5.3). RPKM
1110 values were calculated using DeSEQ2 (version 1.20.0).

1111 **RNA quantification, QuantSeq**

1112 Transcripts from RC9 cells in 2i and 2i/LIF medium were measured using QuantSeq (Lexogen)
1113 and quantified as described. Reads were trimmed with bbduk (version 35.92). Quality control
1114 was performed using fastQC (version 0.11.5). Transcripts were mapped to the mm10 mouse
1115 reference genome using STAR (version 2.5.3). After indexing with samtools (version 1.3) reads
1116 in genes were counted using HTSeq-count (version 0.6.0). Counts were subjected to
1117 differential expression analysis.

1118 **Differential expression analysis**

1119 Differential expression analysis was carried out using limma (version 3.30.13), after
1120 transforming transcript counts using the included voom function. We then fitted a linear
1121 model to sample measurements from each combination of knockouts (including RC9 control)
1122 and conditions (2i and N24). RNAseq batches were included in the model as a confounding
1123 factor. Contrasts were fitted to determine the expression differences between a) each

knockout sample in 2i and the RC9 control, b) each knockout sample in N24 and the RC9 control, and c) the difference between the log-fold-changes calculated in a) and b) (interaction effect). We also fitted a contrast to determine changes during normal differentiation, i.e. the differential expression between RC9N24 *versus* RC92i. All fitted contrasts were tested for differential expression using moderated t-statistics through the limma functions eBayes (H0 of no change) or treat (H0 interval [-log2(1.5), log2(1.5)]). p-Values were corrected for multiple testing using the Benjamini Hochberg method (FDR). For subsequent analyses the more stringent p-values and FDR values as calculated by treat were used (unless noted otherwise). Differential expression analysis of 2i/LIF vs. 2i RC9 measurements was carried out separately, but in an identical fashion.

GO enrichment analysis

Mouse gene GO annotations were extracted from the R package org.Mm.eg.db (version 3.4.0). GO terms that included between 5 and 500 genes were selected for further analysis. For each gene list of interest, significance of functional enrichment of GO terms compared to the background list was determined using Fisher's exact tests. The background was restricted to genes whose transcripts were detected at a median count above 5. To reduce redundancy of functional categories, we clustered GO terms that differed in 5 or fewer genes of interest (using the R base function hclust on the L1-distance of the binary membership matrix). The smallest GO term by total annotations was selected independently from Fisher test results as the primary, i.e. most specific term to represent each cluster. We then carried out correction for multiple hypothesis testing using the Benjamini-Hochberg method on all primary GO terms to obtain adjusted p-values.

1146 **Naive marker-dependency of N24 KO expression patterns (NAGs)**

1147 To quantify the similarity of each gene's differential expression (DE) pattern across KOs at N24
1148 with the DE pattern of seven core pluripotency markers (Nanog, Esrrb, Tbx3, Tfcp2l1, Klf4,
1149 Prdm14 and Zfp42), multiple regression analysis was performed. I.e., each gene's DE pattern
1150 was modelled as a linear combination of marker DE patterns, quantifying the strength of
1151 association as their shared variance (multiple-R^2). NAGs were defined as having a naïve
1152 marker R^2 of ≥ 0.7.

1153 **Cluster analysis (constitutive & N24 induced)**

1154 Based on the responses to the knockout conditions (excluding *Suz12*, *Eed* & *Csnk1a1* KOs) in
1155 both N24 and 2i and their interactions with the RC9 control, we defined two main categories
1156 of responsive genes: (1) the constitutive knockout response: genes that were significantly
1157 changed (adj. p-≤ 0.05) in the same direction in both N24 and 2i, in at least one knockout
1158 condition. Genes from the N24 induced KO response category were excluded. (2) N24 induced
1159 knockout response: genes that were significantly changed in both N24 and the knockout:RC9
1160 interaction term, in at least one knockout condition. To determine whether genes in the N24
1161 induced knockout response subset were overall more correlated with naive pluripotency
1162 markers, we checked the distribution of naive marker multiple-R2 values in the 'N24 induced
1163 knockout response' set compared to both the background distribution and genes classified as
1164 part of the constitutive KO response.

1165

1166 To determine if distinct functional clusters of genes that may be activated in all, single, or sub-
1167 groups of knockouts, we clustered genes in either of the categories (constitutive and N24
1168 induced response) based on their KO^{N2} vs. $RC9^{N24}$ log-fold-changes, using the R base package

1169 hclust with the “Ward.D2” method. We determined the total within-cluster variance at
1170 different numbers of clusters. The number of clusters was chosen such that the model
1171 improved strongly by increasing the number of clusters up to that value, but only marginal
1172 gains result from increasing it further. Thereby 12 gene clusters in the constitutive response
1173 and 12 clusters in the induced response were identified. GO enrichment analysis was carried
1174 out as described above for member genes of each cluster

1175 **Differentiation pathway correlation analysis**

1176 We defined downstream responses of key differentiation pathways using the most specific
1177 responses to the knockouts of their upstream regulators: *Tsc2* (mTOR), *Ptpn11* & *Fgfr1*
1178 (Fgf/Erk signalling), *Tcf7l1* (Wnt signalling), and *Rbpj* (Notch signalling). In addition, LIF-specific
1179 changes were defined as the response of genes in RC9 cells in the 2i/LIF compared to the 2i
1180 condition. To ensure specificity, we required that the log-fold-change of each response gene
1181 for a given regulator (in N2B27 medium) exceeded that observed in the remaining conditions
1182 by a factor of two. In the case of Ptpn11 & Fgfr1, this criterion had to be met by the lower of
1183 the two log-fold-changes for each gene. Additionally, the genes were required to satisfy a
1184 FPKM (CPM in case of LIF) cut-off of at least 1 in either the KO or the WT. From all remaining,
1185 significantly expressed genes (FDR ≤ 0.05) we then selected, by absolute log-fold-change in
1186 descending order, the top 50 most specific genes for each pathway (Supplementary Figure
1187 6a).

1188

1189 Pathway activity was determined across all other knockouts at N24. For each pathway, the
1190 log-fold-changes of the 50 pathway-specific response genes from the associated regulator
1191 knockout (KO^{N24} vs. $RC9^{N24}$) were selected. We then correlated these log-fold-changes to the

1192 log-fold-changes of the same genes in all other knockouts. A positive correlation indicates
1193 that the pathway is disrupted in a manner similar to knocking out its upstream regulator,
1194 whereas a negative correlation indicates the opposite. In the case of LIF-specific genes, a
1195 positive correlation indicates an effect similar to that of adding LIF. The analysis was repeated
1196 for all 2i experiments (using the same lists of specific genes as in N24). P-values of correlation
1197 values were calculated based on Fisher's Z transform. P-values were then corrected for
1198 multiplicity using the Bonferroni method.
1199
1200 To determine how predictive the calculated correlations were of the differentiation
1201 phenotype (mean naive marker logFC vs. RC9 in N2B27), we carried out multiple linear
1202 regression and extracted the coefficients of the model to establish the relative contribution
1203 of each pathway signature to naïve marker deregulation.

1204 **Quantification of differentiation delay in knockouts**

1205 Raw RNA sequencing data was aligned to the mm10 genome and read counts were obtained
1206 using the STAR aligner (version2.5.2b). Gene expression for the following biotypes was
1207 quantified: protein_coding, misc_RNA, miRNA, scaRNA, scRNA, snoRNA, snRNA, sRNA. Read
1208 counts were normalized using DESeq2 (version 1.22.2) and subsequently the log2fc between
1209 all samples and the mean expression of the 2i samples was calculated. This resulted in
1210 expression profiles for each gene over 32 hours of differentiation.
1211
1212 Gaussian process regression was used to smoothen the profiles of each gene (R package tgp
1213 version 2.4-14). The expression profiles of 73 knockouts at N24 were mapped on the
1214 differentiation axis from the time course to quantify differentiation delays per KO, i.e. we

aimed to quantify how many hours the KO expression pattern is 'behind' the expected differentiation in WT cells. This 'mapping' to the time axis was done first computing the logFCs between N24 versus 2i for each KO. The resulting logFC profiles were compared to the logFCs at each time point during the WT differentiation. We assumed that the time point at which the difference between the WT profile and the KO profile is minimized best reflects the molecular differentiation state of the respective KO.

We used the Euclidean distance for this purpose. To make distances better interpretable between the knockouts the distances were set to:

$$\frac{Euclidean\ distance - min\ Euclidean\ distance}{max\ Euclidean\ distance - min\ Euclidean\ distance}$$

where the maximum and minimum distances refer to the respective maximum and minimum Euclidean distances of the respective KO across all time points (i.e. the worst and the best matching time point). Thus, the best matching time point will get a distance score of 0.

In order to further increase the time resolution, additional time points were imputed via linear interpolation. Here we interpolated expression profiles every 15 minutes between the two neighbouring time points of the time point with the smallest Euclidean distance. Timing of knockouts was repeated on the interpolated time points and the new minimal distance was used to quantify the differentiation delay.

Prediction of cluster expression by pathway activity

Pathway activities for each pathway in all KOs were calculated using following formulas:

1237 $$directed\ FC(KO) = FC\left(\frac{KO}{WT}\right) * \begin{cases} LIF \begin{cases} +1\ same\ direction\ as\ in\ \frac{LIF}{2i}\ comparison \\ -1\ opposite\ direction\ as\ in\ \frac{LIF}{2i}\ comparison \end{cases} \\ Tsc2, P11F1, Tcf7l1, Rbpj \begin{cases} +1\ same\ direction\ as\ in\ \frac{rep.KO}{WT}\ comparison \\ -1\ opposite\ direction\ as\ in\ \frac{rep.KO}{WT}\ comparison \end{cases} \end{cases}$$

1238

1239 $$directed\ avg.logFC_{KO} = \frac{\sum_n^{PW\ specific\ marker\ genes} directed\ FC(KO)_n}{\sum_n^{PW\ specific\ marker\ genes} directed\ FC(rep.KO)_n}$$

1240

1241 $$pathway\ activity_{KO} = directed\ avg.logFC_{KO} * \begin{cases} 1\ for\ LIF, Tsc2 \\ -1\ for\ P11F1, Tcf7l1, Rbpj \end{cases}$$

1242

1243 The combinations of pathway activities and average cluster expression changes over all KOs
1244 were used to build a multiple regression model to predict the expression changes in each
1245 cluster by the five pathway activities. Connections between pathway activities and cluster
1246 expression changes not making a significance cut-off (padj <= 0.01) of the multiple regression
1247 models were excluded from each model. The remaining model was tested in the time course
1248 data. Connections from clusters where prediction of expression changes based on pathway
1249 activities did follow or contradict the observed expression changes in the time course were
1250 excluded.

1251

1252 **Embryo comparison**

1253 Sequencing data were obtained from the European Nucleotide Archive [69] from single-cell
1254 mouse embryo profiling studies [10, 70]. Human capacitation dataset was downloaded from
1255 Gene Expression Omnibus website (GSE123055). The macaque FKPM expression dataset was
1256 provided by [33]. Orthologues genes mapping in 1-to-1 fashion were used during comparative
1257 analysis. Alignments to gene loci were quantified with htseq-count [71] based on annotation

1258 from Ensembl 87 [69]. Principal component and cluster analysis were based on log2 FPKM values
1259 computed with the Bioconductor package DESeq [72], custom scripts and FactoRmineR
1260 packages [73]. High variable genes were calculated by fitting a non-linear regression curve
1261 between average log2 FPKM and the square of coefficient of variation. Specific thresholds
1262 were applied along the x-axis (log2FPKM) and y-axis (coefficient of variation) to identify the
1263 most variable genes. Fractional identity between the bulk RNAseq (2i, N24) and the mouse,
1264 macaque embryo or human capacitation dataset were computed using R package
1265 DeconRNASeq [74].

1266 **GENEES (Genetic Network Explorer for Embryonic Stem Cell Differentiation)**

1267 To both visualize and interrogate gene expression analysis of all KO ES cells we have set up
1268 the online tool GENEES (Genetic Network Exploration of Embryonic Stem Cell Differentiation).
1269 It is accessible at [http://shiny.cecad.uni-koeln.de:3838/stemcell_network/; User: stemcells;
1270 Pw: S3m7KmIfHr9]. The app integrates multiple visualisations and summary statistics to make
1271 all information, from the gene level to network modules / pathways / GO terms, directly
1272 available for exploratory analysis. The app has 4 different tabs for inspection of single genes,
1273 predefined genesets, pre-computed clusters and custom genesets. Single genes can be
1274 inspected in the “Genes tab”, while predefined genesets are available for selection under the
1275 “Genesets” tab. The previously described constitutive clusters and InteractionN2 clusters can
1276 be further inspected in the “Pre-computed clusters” tab. In the “Select custom geneset” tab,
1277 a set of genes of interest subjected to analysis.

References

1278 1. Smith, A. Formative pluripotency: the executive phase in a developmental continuum.
1279 *Development* **144**, 365-373 (2017).

1280 2. Kalkan, T. *et al.* Tracking the embryonic stem cell transition from ground state pluripotency.
1281 *Development* **144**, 1221-1234 (2017).

1282 3. Buecker, C. *et al.* Reorganization of Enhancer Patterns in Transition from Naive to Primed
1283 Pluripotency. *Cell Stem Cell* **14**, 838-853 (2014).

1284 4. Dunn, S.J., Martello, G., Yordanov, B., Emmott, S. & Smith, A.G. Defining an essential
1285 transcription factor program for naive pluripotency. *Science* **344**, 1156-1160 (2014).

1286 5. Niwa, H. The principles that govern transcription factor network functions in stem cells.
1287 *Development* **145**, dev157420 (2018).

1288 6. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional
1289 network in embryonic stem cells. *Cell* **133**, 1106 - 1117 (2008).

1290 7. Ying, Q.L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-523
1291 (2008).

1292 8. Boroviak, T., Loos, R., Bertone, P., Smith, A. & Nichols, J. The ability of inner-cell-mass cells to
1293 self-renew as embryonic stem cells is acquired following epiblast specification. *Nature cell*
1294 *biology* **16**, 516-528 (2014).

1295 9. Acampora, D. *et al.* Loss of the Otx2-Binding Site in the Nanog Promoter Affects the Integrity
1296 of Embryonic Stem Cell Subtypes and Specification of Inner Cell Mass-Derived Epiblast. *Cell*
1297 *reports* **15**, 2651-2664 (2016).

1298 10. Mohammed, H. *et al.* Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate
1299 Decisions during Mouse Early Gastrulation. *Cell reports* **20**, 1215-1228 (2017).

1300 11. Leeb, M. & Wutz, A. Derivation of haploid embryonic stem cells from mouse embryos.
1301 *Nature* **479**, 131-134 (2011).

1302 12. Leeb, M., Dietmann, S., Paramor, M., Niwa, H. & Smith, A. Genetic exploration of the exit
1303 from self-renewal using haploid embryonic stem cells. *Cell Stem Cell* **14**, 385-393 (2014).

1304 13. Li, M. *et al.* Genome-wide CRISPR-KO Screen Uncovers mTORC1-Mediated Gsk3 Regulation
1305 in Naive Pluripotency Maintenance and Dissolution. *Cell reports* **24**, 489-502 (2018).

1306 14. Betschinger, J. *et al.* Exit from Pluripotency Is Gated by Intracellular Redistribution of the
1307 bHLH Transcription Factor Tfe3. *Cell* **153**, 335-347 (2013).

1308 15. Guo, G., Huang, Y., Humphreys, P., Wang, X. & Smith, A. A PiggyBac-based recessive
1309 screening method to identify pluripotency regulators. *PloS one* **6**, e18189 (2011).

1310 16. Mulas, C., Kalkan, T. & Smith, A. NODAL Secures Pluripotency upon Embryonic Stem Cell
1311 Progression from the Ground State. *Stem cell reports* **9**, 77-91 (2017).

1312 17. Kokubu, C. & Takeda, J. When half is better than the whole: advances in haploid embryonic
1313 stem cell technology. *Cell Stem Cell* **14** (2014).

1314 18. Elling, U. *et al.* Forward and reverse genetics through derivation of haploid mouse embryonic
1315 stem cells. *Cell Stem Cell* **9**, 563-574 (2011).

1316 19. Kalkan, T. & Smith, A. Mapping the route from naive pluripotency to lineage specification.
1317 *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **369** (2014).

1318 20. Ye, J. & Blelloch, R. Regulation of pluripotency by RNA binding proteins. *Cell Stem Cell* **15**,
1319 271-280 (2014).

1320 21. Kalkan, T. *et al.* Complementary Activity of ETV5, RBPJ, and TCF3 Drives Formative Transition
1321 from Naive Pluripotency. *Cell Stem Cell* **24**, 785-801.e787 (2019).

1322 22. Ursu, A. *et al.* Epiblastin A Induces Reprogramming of Epiblast Stem Cells Into Embryonic
1323 Stem Cells by Inhibition of Casein Kinase 1. *Cell Chem Biol* **23**, 494-507 (2016).

1324 23. Martello, G. & Smith, A. The nature of embryonic stem cells. *Annual review of cell and*
1325 *developmental biology* **30**, 647-675 (2014).

1326 24. Qiu, D. *et al.* Klf2 and Tfcp2l1, Two Wnt/β-Catenin Targets, Act Synergistically to Induce and
1327 Maintain Naive Pluripotency. *Stem cell reports* **5**, 314-322 (2015).

1328 25. Hall, J. *et al.* Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem
1329 cell self-renewal. *Cell Stem Cell* **5**, 597-609 (2009).

1330 26. Yeo, J.C. *et al.* Klf2 is an essential factor that sustains ground state pluripotency. *Cell Stem*
1331 *Cell* **14**, 864-872 (2014).

1332 27. Martello, G. *et al.* Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem
1333 cell self-renewal. *Cell Stem Cell* **11**, 491-504 (2012).

1334 28. Moreira, S. *et al.* Endogenous BioID elucidates TCF7L1 interactome modulation upon GSK-3
1335 inhibition in mouse ESCs. *bioRxiv*, 431023 (2018).

1336 29. Wray, J. *et al.* Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the
1337 pluripotency network and increases embryonic stem cell resistance to differentiation.
1338 *Nature cell biology* **13**, 838-845 (2011).

1339 30. Shy, B.R. *et al.* Regulation of Tcf7l1 DNA binding and protein stability as principal
1340 mechanisms of Wnt/β-catenin signaling. *Cell reports* **4**, 1-9 (2013).

1341 31. Mayer, D. *et al.* Zfp281 orchestrates interconversion of pluripotent states by engaging Ehmt1
1342 and Zic2. *The EMBO journal* **39**, e102591 (2020).

1343 32. Rossant, J. & Tam, P.P.L. Exploring early human embryo development. *Science* **360**, 1075-
1344 1076 (2018).

1345 33. Nakamura, T. *et al.* A developmental coordinate of pluripotency among mice, monkeys and
1346 humans. *Nature* **537**, 57-62 (2016).

1347 34. Rostovskaya, M., Stirparo, G.G. & Smith, A. Capacitation of human naïve pluripotent stem
1348 cells for multi-lineage differentiation. *Development* **146**, dev172916 (2019).

1349 35. Artus, J., Kang, M., Cohen-Tannoudji, M. & Hadjantonakis, A.-K. PDGF signaling is required
1350 for primitive endoderm cell survival in the inner cell mass of the mouse blastocyst. *Stem cells*
1351 *(Dayton, Ohio)* **31**, 1932-1941 (2013).

1352 36. Boroviak, T. *et al.* Single cell transcriptome analysis of human, marmoset and mouse
1353 embryos reveals common and divergent features of preimplantation development.
1354 *Development* **145** (2018).

1355 37. Plusa, B., Piliszek, A., Frankenberg, S., Artus, J. & Hadjantonakis, A.-K. Distinct sequential cell
1356 behaviours direct primitive endoderm formation in the mouse blastocyst. *Development* **135**,
1357 3081-3091 (2008).

1358 38. Smith, A.G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified
1359 polypeptides. *Nature* **336**, 688-690 (1988).

1360 39. Watanabe, S. *et al.* Activation of Akt signaling is sufficient to maintain pluripotency in mouse
1361 and primate embryonic stem cells. *Oncogene* **25**, 2697-2707 (2006).

1362 40. Kunath, T. *et al.* FGF stimulation of the Erk1/2 signalling cascade triggers transition of
1363 pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development*
1364 **134**, 2895-2902 (2007).

1365 41. Lowell, S., Benchoua, A., Heavey, B. & Smith, A.G. Notch promotes neural lineage entry by
1366 pluripotent embryonic stem cells. *PLoS biology* **4**, e121 (2006).

1367 42. Molotkov, A., Mazot, P., Brewer, J.R., Cinalli, R.M. & Soriano, P. Distinct Requirements for
1368 FGFR1 and FGFR2 in Primitive Endoderm Development and Exit from Pluripotency.
1369 *Developmental cell* **41**, 511-526 e514 (2017).

1370 43. Lake, R.J., Tsai, P.-F., Choi, I., Won, K.-J. & Fan, H.-Y. RBPJ, the Major Transcriptional Effector
1371 of Notch Signaling, Remains Associated with Chromatin throughout Mitosis, Suggesting a
1372 Role in Mitotic Bookmarking. *PLoS genetics* **10**, e1004204 (2014).

1373 44. Ohtsuka, S., Nakai-Futatsugi, Y. & Niwa, H. LIF signal in mouse embryonic stem cells. *JAK-*
1374 *STAT* **4**, e1086520-e1086520 (2015).

1375 45. Villegas, F. *et al.* Lysosomal Signaling Licenses Embryonic Stem Cell Differentiation via
1376 Inactivation of Tfe3. *Cell Stem Cell* **24**, 257-270 e258 (2019).

1377 46. Niwa, H., Ogawa, K., Shimosato, D. & Adachi, K. A parallel circuit of LIF signalling pathways
1378 maintains pluripotency of mouse ES cells. *Nature* **460**, 118-122 (2009).

1379 47. Hoffman, J.A., Wu, C.I. & Merrill, B.J. Tcf7l1 prepares epiblast cells in the gastrulating mouse
1380 embryo for lineage specification. *Development* **140**, 1665-1675 (2013).

1381 48. Xu, T. *et al.* RBPJ/CBF1 interacts with L3MBTL3/MBT1 to promote repression of Notch
1382 signaling via histone demethylase KDM1A/LSD1. *The EMBO journal* **36**, 3232-3249 (2017).

1383 49. Amano, T. *et al.* Zscan4 restores the developmental potency of embryonic stem cells. *Nature*
1384 *Communications* **4**, 1966 (2013).

1385 50. Macfarlan, T.S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus
1386 activity. *Nature* **487**, 57-63 (2012).

1387 51. Chen, E.Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis
1388 tool. *BMC Bioinformatics* **14**, 128 (2013).

1389 52. Quenneville, S. *et al.* In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated
1390 Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions.
1391 *Molecular Cell* **44**, 361-372 (2011).

1392 53. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and
1393 adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).

1394 54. Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise Reprogramming of B Cells into Macrophages. *Cell*
1395 **117**, 663-676 (2004).

1396 55. Lassar, A.B., Paterson, B.M. & Weintraub, H. Transfection of a DNA locus that mediates the
1397 conversion of 10T1 2 fibroblasts to myoblasts. *Cell* **47**, 649-656 (1986).

1398 56. Ramos-Ibeas, P. *et al.* Pluripotency and X chromosome dynamics revealed in pig pre-
1399 gastrulating embryos by single cell analysis. *Nature Communications* **10**, 500 (2019).

1400 57. Bredenkamp, N. *et al.* Efficient RNA-mediated reprogramming of human somatic cells to
1401 naïve pluripotency facilitated by tankyrase inhibition. *bioRxiv*, 636670 (2019).

1402 58. Chambers, I. & Tomlinson, S.R. The transcriptional foundation of pluripotency. *Development*
1403 **136**, 2311-2322 (2009).

1404 59. Nichols, J. & Smith, A. Pluripotency in the embryo and in culture. *Cold Spring Harb Perspect*
1405 *Biol* **4**, a008128 (2012).

1406 60. Boroviak, T. & Nichols, J. The birth of embryonic pluripotency. *Philosophical Transactions of*
1407 *the Royal Society B: Biological Sciences* **369** (2014).

1408 61. Boroviak, T. & Nichols, J. Primate embryogenesis predicts the hallmarks of human naïve
1409 pluripotency. *Development (Cambridge, England)* **144**, 175-186 (2017).

1410 62. Rossant, J. & Tam, P.P.L. New Insights into Early Human Development: Lessons for Stem Cell
1411 Derivation and Differentiation. *Cell Stem Cell* **20**, 18-28 (2017).

1412 63. Baker, O. *et al.* RAC-tagging: Recombineering And Cas9-assisted targeting for protein tagging
1413 and conditional analyses. *Sci Rep* **6**, 25529 (2016).

1414 64. Sarov, M. *et al.* A genome-scale resource for in vivo tag-based protein function exploration
1415 in C. elegans. *Cell* **150**, 855-866 (2012).

1416 65. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823-826
1417 (2013).

1418 66. Leeb, M., Perry, A.C.F. & Wutz, A. Establishment and Use of Mouse Haploid ES Cells, in
1419 *Current Protocols in Mouse Biology* (John Wiley & Sons, Inc., 2015).

1420 67. Leeb, M. *et al.* Germline potential of parthenogenetic haploid mouse embryonic stem cells.
1421 *Development* **139**, 3301-3305 (2012).

1422 68. Yusa, K., Zhou, L., Li, M.A., Bradley, A. & Craig, N.L. A hyperactive piggyBac transposase for
1423 mammalian applications. *Proceedings of the National Academy of Sciences of the United*
1424 *States of America* **108**, 1531-1536 (2011).

1425 69. Toribio, A.L. *et al.* European Nucleotide Archive in 2016. *Nucleic Acids Res* **45**, D32-D36
1426 (2017).

1427 70. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-Cell RNA-Seq Reveals Dynamic,
1428 Random Monoallelic Gene Expression in Mammalian Cells. *Science* **343**, 193-196 (2014).

1429 71. Anders, S., Pyl, P.T. & Huber, W. HTSeq—a Python framework to work with high-throughput
1430 sequencing data. *Bioinformatics* **31**, 166-169 (2014).

1431 72. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome*
1432 *biology* **11**, R106 (2010).

1433 73. Lê, S., Josse, J. & Husson, F. FactoMineR: An R Package for Multivariate Analysis. *2008* **25**, 18
1434 (2008).

1435 74. Gong, T. & Szustakowski, J.D. DeconRNASeq: a statistical framework for deconvolution of
1436 heterogeneous tissue samples based on mRNA-Seq data. *Bioinformatics* **29**, 1083-1085
1437 (2013).