

1 **Fibroblasts-derived from Pluripotent Cells Harboring a Single Allele**
2 **Knockout in Two Pluripotency Genes Exhibit DNA Methylation**
3 **Abnormalities and pluripotency induction Defects**
4

5 Rachel Lasry^{1,3}, Noam Maoz^{1,3}, Albert W. Cheng², Nataly Yom Tov¹, Elisabeth Kulenkampff², Meir
6 Azagury¹, Hui Yang², Cora Ople², Styliani Markoulaki², Dina A. Faddah², Kirill Makedonski¹, Ofra Sabbag¹,
7 Rudolf Jaenisch² and Yosef Buganim^{1,*}

- 8
9
- 10 1. Department of Developmental Biology and Cancer Research, The Institute for Medical Research
 - 11 Israel-Canada, The Hebrew University-Hadassah Medical School, Jerusalem 91120, Israel
 - 12 2. Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA; Department of
 - 13 Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
 - 14 3. This authors contributed equally to this work

15
16 *Correspondence should be addressed to Y.B. (yossib@ekmd.huji.ac.il)

17 **ABSTRACT**

18 A complete knockout (KO) of a single key pluripotency gene has been shown to drastically affect
19 embryonic stem cell (ESC) function and epigenetic reprogramming. However, knockin (KI)/KO of a
20 reporter gene only in one of two alleles in a single pluripotency gene is considered harmless and is
21 largely used in the stem cell field. Here, we sought to understand the impact of simultaneous
22 elimination of a single allele in two ESC key genes on pluripotency potential and acquisition. We
23 established multiple pluripotency systems harboring KI/KO in a single allele of two different pluripotency
24 genes (i.e. *Nanog*^{+/-}; *Sall4*^{+/-}, *Nanog*^{+/-}; *Utf1*^{+/-}, *Nanog*^{+/-}; *Esrrb*^{+/-} and *Sox2*^{+/-}; *Sall4*^{+/-}). Interestingly,
25 although these double heterozygous mutant lines maintain their stemness and contribute to chimeras
26 equally to their parental control cells, fibroblasts derived from these systems show a significant
27 reduction in their capability to induce pluripotency either by Oct4, Sox2, Klf4 and Myc (OSKM) or by
28 nuclear transfer (NT). Tracing the expression of *Sall4* and *Nanog*, as representative key pluripotency
29 targeted genes, at early phases of reprogramming could not explain the seen delay/blockage. Further
30 exploration identifies abnormal methylation landscape around pluripotent and developmental genes in
31 the double heterozygous mutant fibroblasts. Accordingly, treatment with 5-azacytidine two days prior to
32 transgene induction rescues the reprogramming defects. This study emphasizes the importance of
33 maintaining two intact alleles for pluripotency induction and suggests that insufficient levels of key
34 pluripotency genes leads to DNA methylation abnormalities in the derived-somatic cells later on in
35 development.

36

37 INTRODUCTION

38 Fluorescent reporter genes are widely used to monitor cell states such as stemness, differentiation, cell
39 cycle and migration (Benchetrit et al., 2019; Eastman et al., 2020). One common strategy to introduce a
40 reporter gene is by a knockin/knockout (KI/KO) approach. In this strategy, a fluorescent gene is
41 introduced into a *locus* of interest by replacing the endogenous gene with a fluorescent reporter, leaving
42 the targeted gene with only one functional allele. Many fluorescent reporter cell lines have been
43 generated over the years using this approach, targeting either pluripotency genes such as *Sox2* (Arnold
44 et al., 2011; Avilion et al., 2003), *Nanog* (Meissner et al., 2007; Wernig et al., 2008) and *Utf1* (Morshedi
45 et al., 2013) or early differentiation genes such as *Gata6* (Heslop et al., 2021). Such reporter lines are
46 useful among others in studying the mechanisms underlying exit from pluripotency and somatic nuclear
47 reprogramming. By producing engineered fibroblasts from these embryonic stem cell (ESC) reporter
48 lines one can easily monitor pluripotency acquisition following the transduction of a set of transcription
49 factors such as OCT4, SOX2, KLF4 and MYC (OSKM) (Buganim et al., 2012; Buganim et al., 2014) or
50 following nuclear transfer (Boiani et al., 2002).

51 While elimination of one allele of one gene is considered harmless to the cell, a complete KO may be
52 detrimental to the cell as seen in the case of *Oct4* and *Sox2* KO for pluripotent cells (Masui et al., 2007;
53 Nichols et al., 1998). In contrast, a complete elimination of other important pluripotent genes such as
54 *Nanog*, partially maintains the pluripotent state and contributes to chimeras, but shows a dramatic
55 reduced reprogramming efficiency by their fibroblast derivatives that can only be partially overcome by
56 high levels of exogenous OSKM factors (Carter et al., 2014; Schwarz et al., 2014). Although KI/KO of one
57 allele is a widely used approach to introduce a reporter gene, our previous study suggests that the
58 quality of the reprogramming process to pluripotency might be affected by the loss of even one allele as
59 fibroblasts with only one intact allele of *Nanog* generated lower quality iPSCs compared to controls as
60 assessed by the stringent pluripotency test, the tetraploid complementation (4n) assay (Buganim et al.,
61 2014).

62 During the maturation phase of the reprogramming process, which is thought to be the bottleneck of
63 the process, epigenetic changes happen stochastically to eventually allow expression of the first
64 pluripotent-related genes (Buganim et al., 2013). Using single-cell analyses, it has been shown that
65 stochastic low expression of pluripotent genes such as *Utf1*, *Esrrb*, *Sall4* (Buganim et al., 2012) and
66 *Nanog* (Polo et al., 2012) can be observed early on in the process in a small fraction of induced cells
67 which is correlated to the low efficiency of reprogramming. The stochastic behavior of the maturation
68 phase ends with the activation of late pluripotent genes such as *Sox2*, *Dppa4*, *Prdm14* and *Gdf3*

69 (Buganim et al., 2012; Soufi et al., 2012) which unleashes the final deterministic phase of
70 reprogramming which leads to stabilization by activation of the core pluripotency network, transgene
71 silencing and complete epigenetic resetting (Buganim et al., 2013).
72 While major efforts have been put to decipher how the identity and levels of the exogenous pluripotent
73 reprogramming factors are linked to efficiency of reprogramming and quality of the resulting iPSCs
74 (Benchetrit et al., 2015; Buganim et al., 2013; Theunissen and Jaenisch, 2014), studies dealing with the
75 effect of reduced levels of endogenous pluripotency genes during development or during pluripotency
76 induction are mostly based on one gene KO approach, that eliminates completely the expression of the
77 targeted gene, or on haploid ESC systems that become diploid very soon during development (Elling et
78 al., 2019; Leeb and Wutz, 2011). Thus, there is very little knowledge of how reduced levels of multiple
79 endogenous pluripotency genes in pluripotent cells affects their developmental potential and their
80 somatic cell derivatives.
81 Here, we sought to examine how elimination of one allele of two pluripotency genes in different
82 pluripotent systems affects their developmental potential and the efficiency of the reprogramming
83 process of their fibroblast derivatives. We produced three secondary systems where two pluripotency
84 genes were KO only in one of the two alleles. These double heterozygous mutant lines include NGFP2
85 ($Nanog^{+/-};Sall4^{+/-}$, $Nanog^{+/-};Esrrb^{+/-}$ and $Nanog^{+/-};Utf1^{+/-}$), NGFP1 ($Nanog^{+/-};Sall4^{+/-}$) and SGFP1 ($Sox2^{+/-}$
86 $;Sall4^{+/-}$). Interestingly, while all double heterozygous mutant lines were capable of contributing to
87 chimeras in a comparable manner to their parental secondary iPSC systems (i.e. NGFP2 ($Nanog^{+/-}$),
88 NGFP1 ($Nanog^{+/-}$) and SGFP1 ($Sox2^{+/-}$)), multiple derivations of fibroblasts from these lines resulted in
89 poor reprogramming efficiency ranging from a complete blockage at the mesenchymal to epithelial
90 (MET) transition (NGFP2 line) to a late blockage at the stabilization step just before the acquisition of
91 pluripotency (NGFP1 and SGFP1 lines). This reduced efficiency was not limited to reprogramming by
92 defined factors but also was evident in nuclear transfer (NT). To understand whether reduced early
93 stochastic expression of these key pluripotency genes can explain the low efficiency of the
94 reprogramming process we generated tracing systems for *Sall4* and *Nanog* as major determinants for
95 the reprogramming process. Tracing *Sall4* or *Nanog* locus activation along the reprogramming process
96 revealed that only a very small fraction of cells activated these *loci*, at one point during the stochastic
97 phase, a result that cannot explain the global blockage seen during the reprogramming process with
98 these double heterozygous mutant lines. To further understand this phenomenon, we profiled the CpG-
99 riched methylation landscape of fibroblasts derived from SGFP1 double heterozygous mutant line and
100 their parental control. Interestingly, a clear difference in the methylation levels of multiple

101 developmental and pluripotent *loci* was observed between the double heterozygous mutant fibroblasts
102 and their parental control cells. In agreement with that, treating double heterozygous mutant
103 fibroblasts for two days prior to factor induction with 5-azacytidine rescued the reprogramming
104 blockage and allowed the induction of pluripotency. This study emphasizes the importance of having
105 two intact alleles for proper pluripotency induction and for normal embryonic development and raises a
106 concern regarding the often used approach of reporter introduction using a KI/KO targeting technique.

107

108 **RESULTS**

109 ***Double heterozygous mutant pluripotent cells contribute to chimeras and exhibit modest*** 110 ***transcriptional changes***

111 Given the importance of properly functioning core ESC circuitry for the establishment and maintenance
112 of pluripotency, we hypothesized that even a small reduction in gene expression of few key pluripotency
113 genes might hold a dramatic effect on the developmental potential of the cells or on their derivatives to
114 undergo nuclear reprogramming.

115 We decided to focus our research on secondary iPSC systems as these systems on the one hand
116 contribute to chimeras and on the other hand exhibit stable and reproducible reprogramming efficiency
117 by minimizing cell heterogeneity (Wernig et al., 2008). Moreover, it allows us to compare a single allele
118 KO of one gene to a single allele KO of two genes with minimal background heterogeneity (Haenebalcke
119 et al., 2013).

120 We started by targeting the NGFP2 secondary system as it already contains a single KI/KO allele of
121 *Nanog* (Wernig et al., 2008). We chose to eliminate a single allele of *Esrrb*, *Utf1* or *Sall4* as they have all
122 been shown to be important for pluripotency and to play a role in reprogramming during the stochastic
123 phase (Buganim et al., 2012; Feng et al., 2009; Tsubooka et al., 2009). To produce a single allele KO and
124 to be able to monitor the activity of the targeted allele, we designed donor vectors that fused, in frame,
125 to the first or second exon a tdTomato reporter (**Figure 1A-B**). Moreover, although a stop codon at the
126 end of the tdTomato was introduced to the targeted allele, to avoid exon skipping and to completely
127 destabilizing the mRNA of the targeted allele we did not add polyA to the targeting vectors. NGFP2 iPSCs
128 were electroporated with either of the three targeting vectors and treated with neomycin for a week.
129 Stable colonies were isolated, expanded and examined for correct targeting by southern blots using
130 external or internal probes (**Figure 1C**, correctly targeted clones are marked by red asterisks). Overall,
131 we isolated two correctly targeted colonies for each combination of manipulated genes: $Nanog^{+/-}$;
132 $Esrrb^{+/-}$ (NGFP2^{N+/-;E+/-}), $Nanog^{+/-}$; $Utf1^{+/-}$ (NGFP2^{N+/-;U+/-}) and $Nanog^{+/-}$; $Sall4^{+/-}$ (NGFP2^{N+/-;S+/-}). To validate

133 the reduced levels of the targeting genes, we cultured the cells in 2i/L medium that recapitulates the
134 ground pluripotent state and facilitates gene expression from both alleles (Miyazari and Torres-Padilla,
135 2012). qPCR and western blot analyses clearly demonstrated a reduction in about 50% of the total
136 mRNA or protein levels of all targeted alleles (**Figures 1D and S1A**), but not in other key pluripotency
137 genes such as *Oct4*, *Sox2*, *Lin28*, *Fbxo15* and *Fgf4* as assessed by qPCR (**Figure S1A**). It is important
138 however to note that out of the examined genes some further reduction in the protein level of NANOG
139 and ESRRB was seen in NGFP2^{N+/-;U+/-} and NGFP2^{N+/-;S+/-} iPSC lines (**Figure 1D**) and in the mRNA of the
140 *Dppa3* gene in NGFP2^{N+/-;S+/-} line (**Figure S1A**). These results suggest that NANOG and ESRRB are either
141 direct or indirect targets of *Sall4* and *Utf1* and that *Dppa3* is regulated by SALL4. To test the stability of
142 the mRNA of the targeted allele we grew the various double heterozygous mutant lines either in S/Lif or
143 2i/Lif conditions and subjected them to flow cytometry analysis for GFP and tdTomato activity. As
144 expected, and in agreement with the western blot analysis, cells grown under S/Lif conditions (i.e.
145 conditions that mostly facilitate mono-allelic expression of Nanog) exhibited 68% GFP reporter activity
146 (reporter that was introduced in frame and contains polyA) in NGFP2^{N+/-} control and NGFP2^{N+/-;E+/-} iPSC
147 lines, and 55% and 58% in NGFP2^{N+/-;S+/-} and NGFP2^{N+/-;U+/-} iPSC lines, respectively (**Figure 1E**). In contrast
148 to the Nanog-GFP reporter and in accordance with our strategy, tdTomato activity for all targeted genes
149 was minor due to the absence of a polyA which resulted in the destabilization of the targeted mRNA
150 (**Figure 1E**). A better activation of the Nanog-GFP reporter was noted under 2i/Lif conditions in all clones
151 but a reduced percentage was still evident in all heterozygous mutant iPSC lines (**Figure S1C**). As in the
152 S/Lif conditions, the activation of the tdTomato reporter in 2i/Lif conditions was minor but still showed a
153 stronger activation than S/Lif conditions. These results validate our strategy of eliminating a single allele
154 in several combinations of two pluripotency genes.

155 We first wished to investigate the impact of eliminating a single allele in two different pluripotent genes
156 on the developmental potential of the cells. To that end, we injected the three double heterozygous
157 mutant lines as well as their parental control cells into blastocysts and measured their potential to form
158 chimeric mice. As can be seen in representative images in Figure S2A, a comparable grade of chimerism
159 was noted between all double heterozygous mutant lines and control iPSC line, suggesting that
160 elimination of a single allele in these combinations of two pluripotent genes does not exert a significant
161 developmental barrier (**Figure S2A**).

162 In our previous study we showed that a gene list of 1716 genes can distinguish between iPSCs with poor,
163 low and high quality as assessed by grade of chimerism and 4n complementation assay (Buganim et al.,
164 2014). Thus, we next profiled the transcriptome of the three heterozygous mutant lines, the parental

165 NGFP2^{N+/-} cells and wild type (WT) ESC (V6.5) control line under S/Lif or 2i/Lif conditions using RNA-
166 sequencing (RNA-seq). Correlation heatmap clustered the cells into two main groups based on the
167 culture conditions used. Nevertheless, within the S/Lif group some changes in gene expression were
168 noted in NGFP2^{N+/-;S+/-} and NGFP2^{N+/-;U+/-} compared to NGFP2^{N+/-;E+/-}, parental NGFP2^{N+/-} and control ESC
169 line (**Figure S2B**). As *Esrrb* was shown to be a downstream target gene of NANOG and to exert a positive
170 feedback loop with it (Festuccia et al., 2012; Sevilla et al., 2021), it is not surprising that the parental
171 NGFP2^{N+/-} and NGFP2^{N+/-;E+/-} exhibited minimal transcriptional changes between them and clustered very
172 close to each other. Principal component analysis (PCA) validated the results seen in the correlation
173 heatmap, separating S/Lif conditions from 2i/Lif conditions by PC1 and NGFP2^{N+/-;S+/-} and NGFP2^{N+/-;U+/-}
174 that grown under S/Lif condition from the rest of the samples by PC2 (**Figure S2C**). Interestingly,
175 NGFP2^{N+/-;U+/-} grown under S/Lif conditions, clustered closer to samples that grew under 2i conditions as
176 indicated by PC1 (**Figure S2C**). These results are in accordance with the notion that UTF1 is mostly
177 implicated in a more primed pluripotent state and less in the ground state (Martinez-Val et al., 2021).
178 As expected, cells grown under 2i/Lif conditions clustered together with minimal transcriptional changes
179 between them (**Figure S2C**). These results suggest that elimination of one allele of two different
180 pluripotent genes from the tested combinations, although show some small transcriptional change
181 under S/Lif conditions, can still maintain a functional pluripotency state with minimal variation in gene
182 expression under ground pluripotency state.

183 184 ***Fibroblasts derived from NGFP2 double heterozygous mutant iPSC lines fail to induce pluripotency***

185 Given that the reprogramming process involves a stochastic phase of activation of pluripotency genes
186 (Buganim et al., 2012), we hypothesized that MEFs harboring double heterozygous mutant alleles might
187 exhibit reprogramming delay due to difficulties in the activation of the core pluripotency circuitry.

188 To that end, secondary MEF systems were established from all the three NGFP2 double heterozygous
189 mutant lines, as well as from the parental NGFP2 control. These secondary MEF systems contain a
190 unique integration pattern of OSKM transgenes under Tet-on promotor and a M2rtTA transactivator in
191 the *Rosa26 locus* (Wernig et al., 2008). To initiate reprogramming, MEFs were exposed to dox for 13
192 days followed by dox withdrawal for 3 more days to stabilize any iPSC colony, and the percentage of
193 Nanog-GFP-positive cells was scored by flow cytometry.

194 In accordance with our hypothesis, while NGFP2^{N+/-} control cells exhibited the expected ~2% of Nanog-
195 GFP-positive cells (Buganim et al., 2012; Wernig et al., 2008), all the double heterozygous mutant lines
196 showed a blockage in reprogramming (**Figure 2A**). This blockage was not due to cell death or

197 proliferation arrest as all double heterozygous mutant and control plates stained equally to crystal
198 violet, indicating a comparable number of cells between all mutant lines and control (**Figure 2B**).
199 However, in agreement with the flow cytometry results, although all the double heterozygous mutant
200 lines stained positive to the early reprogramming marker alkaline phosphatase (AP), implying that the
201 cells initiated reprogramming, their AP staining was significantly lower compared to the control plates
202 (**Figure 2C**). By extending the dox exposure time to 20 days, a small percentage of Nanog-GFP-positive
203 cells (i.e. ~0.3-0.4% Nanog-GFP-positive cells in the mutant lines compared to ~10% in the control line)
204 did emerge in all double heterozygous mutant lines, suggesting that some cells can overcome this
205 blockage when prolonged exposure of the 4 factors is triggered (**Figure 2D**).

206 We then asked whether the observed reprogramming blockage within these double heterozygous
207 mutant MEF lines is specific to reprogramming by defined factors or whether the loss of the two alleles
208 would show reprogramming defects in other reprogramming techniques as well. We chose to use the
209 nuclear transfer (NT) technique as it utilizes the entire array of reprogramming proteins within the egg
210 as opposed to the Yamanaka's approach that uses very few selected reprogramming factors. Enucleated
211 eggs were injected with MEF nuclei from each of the three double heterozygous mutant MEF lines and
212 control, and blastocyst formation and the establishment of ESC lines from these blastocysts were
213 scored. Notably, while all lines exhibited a comparable and expected efficiency in producing blastocysts,
214 the efficiency of ESC line derivation was significantly lower in the double heterozygous mutant lines
215 compared to control (i.e. 0-4% in the double heterozygous mutant lines vs 11% in control line, **Figure**
216 **2F**). Taken together, these results suggest that the elimination of two alleles from two different key
217 pluripotency genes affects the somatic nucleus in a way that interferes with its capability to undergo
218 reprogramming to pluripotency by various techniques.

219 We then asked whether the loss of the two alleles causes a permanent reprogramming defect or
220 whether it can be rescued by exogenously expressing one of the targeted genes. To that end, each
221 double heterozygous mutant MEF line was transduced with either *Nanog* or with its corresponding
222 targeted gene (i.e. *Sall4*, *Utf1* or *Esrrb*) or with additional viruses encoding for OSK. Overall, both *Nanog*
223 or each of the corresponding factor showed either partial (i.e. *Esrrb* in NGFP2^{N+/-;E+/-} cells and *Nanog* in
224 NGFP2^{N+/-;S+/-} cells) or complete rescue of the reprogramming blockage, while additional OSK further
225 boosted the reprogramming process (**Figures 2E and S2D-E**). The fact all the double heterozygous
226 mutant lines could be rescued by the addition of different pluripotent factors, suggests that the seen
227 blockage is not specific to the unique function of the eliminated alleles' but rather is associated with a
228 more general effect.

229

230

231 ***NGFP2^{N+/-} double heterozygous mutant lines show an early defect in the activation of epithelial***
232 ***markers***

233 Given that double heterozygous mutant MEF lines are capable, to some extent, to activate the AP
234 enzyme (**Figure 2C**), we next sought to understand at which time point during the reprogramming
235 process the double heterozygous mutant cells lose their capacity to undergo reprogramming. To that
236 end, we examined the expression level of early reprogramming markers (*Fgf4* and *Fbxo15*), intermediate
237 markers (*endogenous Oct4* and *Sall4*) and late and predictive markers (*Sox2*, *Utf1*, *Esrrb* and *Lin28*)
238 following 13 days of dox addition (Buganim et al., 2012; Buganim et al., 2013)). While the double
239 heterozygous mutant induced cells showed some activation of the early markers and very low
240 expression of intermediate markers, no activation at all was seen in the late and predictive markers
241 compared to control cells (**Figure S2F**). These results suggest that the blockage seen in these double
242 heterozygous mutant cells during reprogramming occurs relatively early in the reprogramming process.
243 It is interesting to note that out of the three double heterozygous mutant lines, NGFP2^{N+/-;S+/-} showed
244 the strongest inhibitory effect as assessed by marker expression (**Figure S2F**), Nanog-GFP activation in
245 the *Nanog* rescue experiment (**Figure S2D**), and AP staining (**Figure 2C**).

246 We next profiled the transcriptome of the three double heterozygous mutant lines and control lines (i.e.
247 NGFP2^{N+/-} cells, and NGFP2^{N+/-} cells that were infected with empty vector (EV)) following six days of
248 reprogramming. We chose this time point as it showed a clear reprogramming delay in the double
249 heterozygous mutant plates compared to control plates. NGFP2^{N+/-} MEFs and the parental NGFP2^{N+/-}
250 iPSCs were profiled as well. Hierarchical clustering analysis showed that all the double heterozygous
251 mutant lines clustered together and different than the control lines (**Figure 3A**). PCA and scatter plots
252 emphasize even more the transcriptomic differences between the double heterozygous mutant lines
253 and NGFP2^{N+/-} control lines after 6 days of dox. While the control lines demonstrated significant
254 transcriptional changes by day 6 of reprogramming (represented by PC2) compared to parental
255 NGFP2^{N+/-} MEFs, all the double heterozygous mutant lines showed minimal transcriptional changes
256 between themselves or when compared to NGFP2^{N+/-} MEFs (**Figure 3B-D**). These results suggest that an
257 early reprogramming defect is responsible for the delay seen in the NGFP2^{N+/-}-double heterozygous
258 mutant lines.

259 An important and essential early step in reprogramming is mesenchymal to epithelial transition (MET).
260 In this step the induced cells lose their fibroblastic characteristics and start acquiring an epithelial
261 identity. As the transcriptome of the double heterozygous mutant lines after 6 days of reprogramming
262 (**Figure 3B-D**) was still very close to MEFs, we next asked whether the MET process is impaired in the

263 double heterozygous mutant lines. Initially we examined the expression levels of four well-known
264 fibroblastic markers; *Thy1*, *Col5a1*, *Postn*, and *Des* and noticed that in all mutant lines these markers are
265 significantly downregulated in a comparable manner to that of the control lines (**Figure 3E**). Similarly, a
266 comparable downregulation in the expression of the EMT master regulators; *Twist1*, *Zeb1*, *Snai2*, and
267 *Foxc2* was noted as well (**Figure S3A**), implying that the loss of the fibroblastic identity is not impaired in
268 the double heterozygous mutant lines. However, when we tested whether the cells are capable of
269 activating the epithelial program, we noticed that while the control lines are able to activate epithelial
270 markers such as *Cdh1*, *Dsp*, *Epcam*, *Cldn4*, and *Cldn7*, the double heterozygous mutant lines fail in doing
271 so (**Figure 3F and S3B**). These results suggest that there is an inherent blockage acquired specifically in
272 the double heterozygous mutant MEF lines that prevents/delays them from activating a robust epithelial
273 program as occurs during intact cellular reprogramming.

274
275 ***Reprogramming impairment caused by double heterozygous allele elimination is not restricted to a***
276 ***system nor to the identity of the modified alleles***

277 To exclude the possibility that the observed effect is system-specific, we produced additional secondary
278 double heterozygous mutant iPSC systems that differ in their reprogramming efficiency and dynamics,
279 developmental potential, allele-specific elimination and reprogramming factor stoichiometry.

280 We decided to produce a double heterozygous mutant line from NGFP1^{N⁺/-} system as it was generated in
281 parallel to the NGFP2^{N⁺/-} system but demonstrated different reprogramming efficiency and dynamics
282 and reprogramming factor induction levels (Wernig et al., 2008).

283 As NGFP2^{N⁺/-;S⁺/-} double heterozygous mutant line demonstrated the strongest delay in pluripotency
284 induction, we decided to eliminate one allele of *Sall4* in NGFP1^{N⁺/-} as well. Initially, we confirmed by
285 single molecule mRNA-FISH that the strong effect seen in NGFP2^{N⁺/-;S⁺/-} is not a result of *Sall4* reduction
286 that is greater than 50%. In agreement with the protein level (**Figure 1D**), Figure 4A shows the
287 distribution of *Sall4* transcript level in NGFP2^{N⁺/-} cells (n=57) compared to NGFP2^{N⁺/-;S⁺/-} cells (n=49),
288 validating transcript reduction of *Sall4* in about 50% within NGFP2^{N⁺/-;S⁺/-} cells (**Figure 4A**).

289 Then, we targeted a tdTomato reporter gene into the *Sall4* locus of NGFP1^{N⁺/-} as described above to
290 produce NGFP2^{N⁺/-;S⁺/-} (**Figure 4B**). Correctly targeted NGFP1^{N⁺/-;S⁺/-} iPSC double heterozygous colonies
291 were validated by PCR and Western blot (**Figure 4C-D**). We also produced a *Nanog* KO NGFP1^{N⁻/-} line as a
292 single KO gene control (**Figures 4E-F and S4A**). Secondary MEFs were produced from NGFP1^{N⁺/-},
293 NGFP1^{N⁺/-;S⁺/-}, and NGFP1^{N⁻/-} which were then exposed to dox for 13 days following by dox withdrawal for
294 3 more days. Flow cytometry analysis of the various reprogramming plates showed a clear and

295 comparable reduction in the percentage of Nanog-GFP-positive cells in the double heterozygous mutant
296 cells and in the *Nanog* KO fibroblasts compared to control NGFP1^{N+/-} cells (**Figure 4G**). Exogenous
297 expression of *Nanog*, from the onset of the reprogramming process, rescued both *Nanog* KO cells and
298 the NGFP1^{N+/-;S+/-} double heterozygous mutant cells (**Figure 4G-H**). These intriguing results suggest that
299 even a reduction of 50% in the levels of two important pluripotency genes (i.e. Nanog and Sall4) is
300 crucial for the establishment of the core pluripotency circuitry in a way that is comparable to a complete
301 KO of key pluripotency genes such as *Nanog*.

302 We were interested to examine whether the pluripotency induction impairment seen in the double
303 heterozygous mutant lines is restricted to combinations that harbor allele elimination of *Nanog*.

304 To that end, we eliminated one allele of *Sall4* in SGFP1^{S2+/-} line, a secondary iPSC system that was
305 generated in our lab and contains GFP reporter instead of one allele of *Sox2*. Correctly targeted
306 SGFP1^{S2+/-;S4+/-} iPSC colonies were validated by PCR, Western blot and immunostaining (**Figures 4I-K and**
307 **S4B**). In agreement with the other systems, a significant reduction in reprogramming efficiency was
308 noted in SGFP1^{S2+/-;S4+/-} cells compared to SGFP1^{S2+/-} control as assessed by flow cytometry and number of
309 Sox2-GFP-positive colonies (**Figure 4K-M**). It is interesting to note that while all the double heterozygous
310 NGFP^{N+/-} lines produced a neglectable number of iPSCs following 13 days of reprogramming (i.e. 0.0%-
311 0.2%), the SGFP1^{S2+/-;S4+/-} double heterozygous mutant cells produced about 2%-2.5% of iPSCs. This
312 difference can be explained by the levels of the OSKM transgenes that is much higher in SGFP1^{S2+/-} cells
313 than in NGFP^{N+/-} cells (**Figure S4C**) as additional levels of OSK can rescue the phenotype of the double
314 heterozygous mutant lines (**Figure S2E**). Taken together, these results suggest that the double
315 heterozygous phenotype is not system nor gene- specific. This observation raises a real concern as to
316 the KI/KO targeting approach when cellular state induction is studied.

317
318 ***Reduced early stochastic expression of the targeted genes cannot explain the reprogramming***
319 ***blockage seen in the double heterozygous mutant lines***

320 Stochastic expression of pluripotency genes during early stages of reprogramming was evident by
321 multiple single-cell studies (Buganim et al., 2012; Buganim et al., 2013; Guo et al., 2019). Thus, we
322 hypothesized that the lack of two key pluripotency alleles in the double heterozygous mutant cells might
323 impair their ability to successfully pass the early stochastic phase, resulting in a blockage/delay in
324 reprogramming. To explore this possibility, we generated tracing system for *Nanog* and *Sall4* as two
325 representative genes out of the 5 targeted ones (i.e. *Nanog*, *Sall4*, *Utf1*, *Esrrb* and *Sox2*). We chose
326 *Nanog* because it appears in most of our double heterozygous mutant lines and *Sall4* because it exhibits

327 the highest levels of stochastic expression at early phases of reprogramming compared to the other
328 targeted genes (Buganim et al., 2012).

329 To that end, we targeted a *2A-EGFP-ERT-CRE-ERT* cassette into *Sall4* or *Nanog* locus using gRNA that is
330 located at the 3' UTR of the gene. We targeted an ESC line (RL) that contains a lox-STOP-lox (L-S-L)
331 cassette upstream to a tdTomato reporter gene and M2rtTA transactivator at the *Rosa26* locus (each
332 cassette at different allele, **Figure 5A, 5B**). Upon *Sall4* or *Nanog* expression and the addition of
333 tamoxifen, CRE recombinase is translocated to the nucleus and removes the L-S-L cassette, leading to
334 irreversible activation of the tdTomato reporter. Given that both *Nanog* and *Sall4* are expressed in ESCs,
335 transfected colonies were sorted based on EGFP expression and correct targeting was validated by PCR
336 (**Figure 5C-D**). To assess the efficiency of the tracing system, correctly targeted ESC clones (i.e. RL8 for
337 *Sall4* and RL9 for *Nanog*) were exposed to Tamoxifen (Tam) and the percentage of tdTomato-positive
338 cells were scored under the microscope and by flow cytometry, demonstrating very high L-S-L cassette
339 removal efficiency (**Figures 5E-F and S5A-D**).

340 The observation that NGFP2^{N+/-} double heterozygous mutant induced cells could not activate the
341 epithelial program (**Figure 3F and S3B**) suggests that the blockage in reprogramming in these cells is
342 general and not restricted to the small fraction of cells that are destined to become iPSCs. Thus, in order
343 to correlate the stochastic expression of the targeted alleles to the observed blockage/delay, most
344 induced cells should show some activation of the targeted alleles at early time point of reprogramming.
345 MEFs produced from *Sall4* and *Nanog* tracing ESC systems were transduced with OSKM cassette and
346 tdTomato activation was assessed in the induced cells after 6 days and following 14 days of
347 reprogramming followed by 3 days of dox removal. We chose day 6 since it is an early time point that
348 exhibits high stochastic expression of pluripotency genes (Buganim et al., 2012; Buganim et al., 2013;
349 Guo et al., 2019). However, only up to 0.24% of the *Sall4* tracing cells and up to 0.62% of *Nanog* tracing
350 cells were tdTomato-positive at day 6 of reprogramming, ruling out the hypothesis that *Sall4* or *Nanog*
351 stochastic expression early in the reprogramming process is responsible for the double heterozygous
352 mutant phenotype (**Figure 5G-I, 5J-L**). 7.42% of *Sall4*-EGFP in conjunction with 7.96% of tdTomato-
353 positive cells for the *Sall4* tracing system and 2.8% of *Nanog*-EGFP together with 6.7% of tdTomato-
354 positive cells for the *Nanog* tracing system at the end of the reprogramming process confirmed
355 successful reprogramming, refuting the possibility that the low percentage of tdTomato-positive cells
356 observed at day 6 of reprogramming in the *Sall4* and *Nanog* tracing systems is due to low
357 reprogramming efficiency (**Figure 5M-N**). In conclusion, this set of experiments, challenges the

358 hypothesis that reduced stochastic expression of the targeted pluripotent alleles is responsible for the
359 early blockage seen in the double heterozygous mutant lines.

360

361 ***Methylation abnormalities in the double heterozygous mutant fibroblasts is correlated with***
362 ***reprogramming impairment***

363 The fact that additional exogenous expression of OSK factors could rescue the delay seen in the double
364 heterozygous mutant cells (**Figure S2E**), suggests that epigenetic abnormalities, rather than genetic
365 modifications (i.e. the elimination of the targeted alleles themselves), are responsible for the observed
366 blockage. This notion is also supported by the transcriptomic changes seen in the double heterozygous
367 NGFP2^{N+/-} iPSC mutant lines that grew under S/Lif condition, but not in 2i/Lif conditions, that force the
368 naïve ground state (**Figure S2B-C**). As 2i/Lif conditions induce DNA hypomethylation (Sim et al., 2017),
369 and since DNA methylation reshaping has been shown to be crucial for reprogramming (Buganim et al.,
370 2013; De Carvalho et al., 2010) we hypothesized that the double heterozygous mutant fibroblasts might
371 harbor abnormal DNA methylation that hinders their ability to undergo reprogramming. To test this
372 hypothesis, MEFs derived from double heterozygous mutant SGFP1^{S2+/-;S4+/-} iPSCs and from their parental
373 SGFP1^{S2+/-} iPSCs were subjected to methylation analyses using reduced representation bisulfite
374 sequencing (RRBS) to capture the CpG-enriched methylation landscape of the cells as a representation
375 for the global methylation changes.

376 RRBS analysis revealed that the two fibroblast lines are very similar in regard to their CpG-enriched
377 methylation landscape, suggesting that overall the double heterozygous mutant line harbor a correct
378 fibroblastic methylation landscape. However, read counts did vary between samples and so did reads-
379 per-site, clustering them as two different groups (**Figure 6A**). We then searched for differentially
380 methylated regions (DMRs) between the two fibroblast lines. DMRs were defined as CpG sites of
381 consecutive tiles that are 100bp long in size, include at least 15 reads and show at least 20% methylation
382 differences between the two fibroblast lines. All DMRs were adjusted to p-value of 1e-3 or lower. This
383 analysis yielded two groups of DMRs: (i) 1263 tiles that are more methylated and (ii) 1384 tiles that are
384 less methylated in the double heterozygous mutant line compared to control (**Figure 6B-C**). We then
385 associated each DMR to its neighboring gene and ran GO term analysis using EnrichR (Xie et al., 2021).
386 Interestingly, many of the differentially methylated *loci* were found to be associated with pluripotency
387 and developmental pathways (**Figure 6D-E**). Specifically, dataset derived from loss of function
388 experiments suggested that these genes are being upregulated in ESCs upon loss of function of *Oct4*,
389 and are associated with the Hippo signaling (**Figure 6D-E**), suggesting that the loss of indicated two

390 pluripotency alleles hinders the capability of the core pluripotency circuitry to maintain normal DNA
391 methylation of these *loci* later on in development.

392 To confirm that DNA methylation abnormalities is responsible for the reprogramming delay seen in the
393 double heterozygous mutant lines we next employed the DNA hypomethylation agent, the 5-Aza-2'-
394 deoxycytidine (5'azaDC). Double heterozygous mutant fibroblasts from all lines were treated for two
395 days with 5'azaDC and reprogramming experiments were carried out by the addition of dox for 13 days
396 followed by 3 days of dox removal. In agreement with the RRBS results, treatment of 5'azaDC for two
397 days prior to dox addition rescued the reprogramming defect seen in the double heterozygous mutant
398 lines (**Figure 6F**). These results suggest that reduced levels of pluripotency genes at the pluripotent state
399 leads to methylation abnormalities later on in their somatic cell derivatives. Moreover, these data
400 suggest that the KI/KO approach to introduce a reporter gene should be avoided to maintain normal
401 epigenetic state in the cells.

402

403 DISCUSSION

404 Fluorescent reporter genes are a widely used tool in science to monitor the activity of a gene, regulatory
405 element, non-coding RNA or other elements in the genome. One of the most common approach to
406 introduce a reporter gene in a *locus-specific* manner is by the KI/KO approach. In this technique, the
407 genomic sequence of the element of interest is being replaced by the coding sequence of the reporter
408 gene, leaving only one intact allele of the targeted element. While this approach is considered harmless
409 to the cells, up till now no thorough study has been conducted to support this claim. Here, by using
410 pluripotent stem cells as a tested model we aimed to understand how allele elimination affects cell's
411 function and potential. To increase our ability to detect abnormalities caused by the elimination of the
412 targeted alleles we deleted a single allele from various combinations of two pluripotency genes (i.e.
413 *Nanog*^{+/-};*Sall4*^{+/-}, *Nanog*^{+/-};*Esrrb*^{+/-}, *Nanog*^{+/-};*Utf1*^{+/-}, *Sox2*^{+/-};*Sall4*^{+/-}) and used different pluripotent stem
414 cell systems to exclude any system-specific effect. Interestingly, while examination of the
415 developmental potential of the cells did not reveal a significant difference between the double
416 heterozygous mutant cells and their parental controls, fibroblasts derived from these double
417 heterozygous mutant pluripotent cells, demonstrated a strong blockage in their capability to induce
418 pluripotency either by transcription factors or by nuclear transfer. The poor reprogramming efficiency
419 observed between the various pluripotent stem cell systems ranged from a complete blockage at the
420 mesenchymal to epithelial (MET) transition (NGFP2 line) to a later blockage at the stabilization step just
421 before the acquisition of pluripotency (NGFP1 and SGFP1 lines, data not shown). Given that the affected
422 genes were shown to play a major role during the stochastic phase of the reprogramming process
423 (Buganim et al., 2012; Guo et al., 2019; Morshedi et al., 2013), we next examined the possibility that
424 reduced stochastic expression of the targeted genes hinders the capability of the cells to pass the
425 stochastic phase and to induce pluripotency. As the blockage seen in the double heterozygous mutant
426 cells happened in the vast majority of the induced cells, to support the claim that reduced stochastic
427 expression is responsible to the observed blockage, we aimed to show that the activation of the *Sall4* or
428 *Nanog* allele is a frequent event and occurred in most induced cells at early stages of reprogramming. To
429 test this hypothesis, we generated a tracing system for *Nanog* and *Sall4* that is based on the activity of
430 Cre recombinase that unleashes an irreversible activation of a tdTomato reporter gene once activated.
431 However, only a small number of induced cells turned on the tdTomato reporter following six days of
432 factor induction, suggesting that reduced stochastic expression of these genes is not responsible for the
433 global reprogramming delay seen in the double heterozygous mutant cells.

434 Additional expression of multiple pluripotent factors (e.g. *Sall4*, *Nanog*, *Utf1*, *Esrrb*, OSK) can either
435 partially or fully rescue the observed blockage, thus, we next hypothesized that epigenetic barrier in the
436 double heterozygous mutant fibroblasts, but not the allele elimination itself, may cause the observed
437 delay. To that end, we subjected the parental and its double heterozygous mutant fibroblast
438 counterparts to CpG-enriched DNA methylation analysis. A clear difference in the DNA methylation
439 levels in regions within pluripotent and developmental genes was noted between the two fibroblast
440 lines, suggesting that even a 50% reduction in the levels of two pluripotent genes is sufficient to induce
441 aberrant DNA methylation during development. In fact, although *Oct4* expression level was not affected
442 in the iPSCs, GEO enrichment of the derived MEFs showed loss of function of *Oct4* as a core pluripotency
443 player which can be explained by the fact that key pluripotent genes such as *Nanog*, *Sox2*, *Sall4* and
444 *Esrrb* who have been shown to regulate and function with the core DNA methylation machinery were
445 missing (Adachi et al., 2018; Shanak and Helms, 2020; Tan et al., 2013).

446 The KI/KO approach is a well-accepted technique to introduce a reporter gene, however, there are other
447 methodologies to introduce a reporter gene into a gene of interest without eliminating the gene's
448 coding sequences. This includes the use of self-cleavage peptides such as 2A and the internal ribosome
449 entry site (IRES). Nevertheless, even in these, relatively safe techniques, in many occasions, a robust
450 degradation of the targeted allele is still observed due to destabilization of the targeted mRNA following
451 the introduction of the reporter cassette itself (Benchetrit et al., 2019). Overall, our study suggest that
452 the KI/KO approach should be used carefully when cell state establishment is studied and emphasizes
453 the importance of having two intact alleles for proper cellular functioning.

454 **Material and Methods**

455 **Cell culture**

456 Mouse embryonic fibroblasts (MEFs) were isolated as previously described (Wernig et al., 2008). MEFs
457 were grown in DMEM supplemented with 10% fetal bovine serum, 1% non-essential amino acids, 2 mM
458 L-Glutamine and antibiotics. ESCs and iPSCs were grown in S/Lif medium or 2i/Lif: DMEM supplemented
459 with 10% fetal bovine serum, 1% non-essential amino acids, 2 mM L-Glutamine, 2X106 units mLif,
460 0.1 mM β -mercaptoethanol (Sigma) and antibiotics with or without 2i- PD0325901 (1 mM) and
461 CHIR99021 (3 mM) (PeproTech). All the cells were maintained in a humidified incubator at 37°C and 6%
462 CO₂. All infections were performed on MEFs (passage 0-2) that were seeded at 50-70% confluency two
463 days before the first infection. During the reprogramming to iPSC, the cells were grown in S/Lif medium
464 with the addition of 2 μ g/ml doxycycline.

465

466 **Secondary MEF production**

467 Briefly, iPSC lines (NGFP2, NGFP1 and SGFP1 lines) were injected into blastocysts and chimeric embryos
468 were isolated at E13.5. For MEF production, embryos were dissected under the binocular removing
469 internal organs and heads. The remaining body was chopped thoroughly by scalpels and exposed to 1ml
470 Trypsin-EDTA (0.25%, GIBCO) for 30 minutes at 37°C. Following that, 10 mL of DMEM medium containing
471 10%FBS was added to the plate and the chopped tissue was subjected to thorough and intensive
472 pipetting resulting in a relatively homogeneous mix of cells. Each chopped embryo was seeded in 15cm
473 plate and cells were cultured in DMEM supplemented with 10%FBS, 2mM L-glutamine, and antibiotics
474 until the plate was full. Puromycin (2 μ g/ml) was added to each 15cm plate for positive selection for
475 NGFP2, NGFP1 and SGFP1 MEFs, eliminating only the host cells.

476

477 **Immunostaining and Western blot**

478 Cells were fixed in 4% paraformaldehyde (in PBS) for 20 minutes. The cells were rinsed 3 times with PBS
479 and blocked for 1hr with PBS containing 0.1% Triton X-100 and 5% FBS. The cells were incubated
480 overnight with primary antibodies (1:200) in 4C. The antibodies are: anti-SALL4 (Abcam, ab29112) and
481 anti-NANOG (Bethyl, A300-379A) in PBS containing 0.1% Triton X-100 and 1%FBS (1:200 dilution). The
482 next day, the cells were washed 3 times and incubated for 1hr with relevant (Alexa) secondary antibody
483 in PBS containing 0.1% Triton X-100 and 1% FBS (1:500 dilution). DAPI (1:1000 dilution) was added 10
484 minutes before the end of incubation. For western blot, cell pellets were lysed on ice in lysis buffer (20
485 mM Tris-HCl, pH 8, 1mM EDTA pH 8, 0.5% Nonidet P-40, 150mM NaCl, 10% glycerol, 1mM, protease

486 inhibitors (Roche Diagnostics) for 10 min, supernatant were collected and 40µg protein were suspended
487 with sample buffer and boiled for or 5 min at 100C, and subjected to western blot analysis. Primary
488 antibodies: anti-SALL4 (Abcam, ab29112), anti-NANOG (Bethyl, A300-379A), anti-ESRRB (Perseus
489 proteomics, PP-H6705-00), anti-UTF1 (Abcam, ab24273), anti-Actin (Santa cruz, sc-1616), anti-β-Tubulin
490 (Abcam, ab179513), anti-Vinculin (Abcam, ab129002). Blots were probed with anti-mouse, anti-goat or
491 anti-rabbit IgG-HRP secondary antibody (1:10,000) and visualized using ECL detection kit.

492

493 **Southern Blot**

494 Southern blot was performed as previously described (Carey et al., 2011).

495

496 **FACS analysis**

497 Cells were washed twice with PBS and trypsinized (0.25%) and filtered through mesh paper. Flow
498 cytometry analysis was performed on a Beckman Coulter and analyzed using Kaluza Software. All FACS
499 experiments were repeated at least three times, and the bar graph results are presented as a mean ±
500 standard deviation of two biological duplicate from a typical experiment. Flow cytometry analysis was
501 performed on a Beckman Coulter and analyzed using Kaluza Software.

502

503 **Quantitative real-time PCR**

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

511

512 **RNA sequencing**

513 Total RNA was isolated using Rneasy Kit (QIAGEN) and sent to the “Technion Genome Center”, Israel, for
514 library preparation and sequencing.

515

516 **Cleaning and filtering of raw reads**

517 Raw reads (fastq files) were inspected for quality issues with FastQC (v0.11.2,
518 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). According to the FastQC report, reads
519 were then trimmed to a length of 50 bases with fastx_trimmer of the FASTX package (version 0.0.13,
520 http://hannonlab.cshl.edu/fastx_toolkit/), and quality-trimmed at both ends, using in-house perl scripts,
521 with a quality threshold of 32. In short, the scripts use a sliding window of 5 base pairs from the read's
522 end and trim one base at a time until the average quality of the window passes the given threshold.
523 Following quality-trimming, adapter sequences were removed by Trim Galore (version 0.3.7,
524 http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), using the command "trim_galore -a
525 \$adseq -length 15" where \$adseq is the appropriate adapter sequence. The remaining reads were
526 further filtered to remove very low-quality reads, using the fastq_quality_filter program of the FASTX
527 package, with a quality threshold of 20 at 90 percent or more of the read's positions.

528

529 **Expression analysis**

530 The cleaned fastq files were mapped to the mouse transcriptome and genome, Ensembl version
531 GRCm38 from Illumina's iGenomes
532 (http://support.illumina.com/sequencing/sequencing_software/igenome.html), using TopHat (v2.0.11),
533 allowing up to 3 mismatches and a total edit distance of 8 (full command: tophat -G
534 Mus_musculus/Ensembl/GRCm38/Annotation/Genes/genes.gtf -N 3 --read-gap-length 5 --read-edit-dist
535 8 --segment-length 18 --read-realign-edit-dist 5 --b2-i S,1,0.75 --b2-mp 3,1 --b2-score-min L,-0.5,-0.5
536 Mus_musculus/Ensembl/GRCm38/Sequence/Bowtie2Index/genome_clean.fastq). Quantification and
537 normalization were done with the Cufflinks package (v2.2.1). Quantification was done with cuffquant,
538 using the genome bias correction (-b parameter), multi-mapped reads assignment algorithm (-u
539 parameter) and masking for genes of type IG, TR, pseudo, rRNA, tRNA, miRNA, snRNA and snoRNA (-M
540 parameter). Normalization was done with cuffnorm (using output format of Cuffdiff).

541

542 **Visualization**

543 The R package cummeRbund (version 2.8.2) was used to calculate and draw the figures (such as scatter
544 plots, MA plots, etc.) from the normalized expression values.

545

546 **Chimera Formation**

547 Blastocyst injections were performed using (C57/Bl6xDBA) B6D2F2 or CB6F1 host embryos. All injected
548 iPSC lines were derived from crosses of 129Sv/Jae to C57/Bl6 mice and could be identified by agouti coat

549 color. Embryos were obtained 24 hr (1 cell stage) or 40 hr (2 cell stage) posthuman chorionic
550 gonadotropin (hCG) hormone priming. Diploid embryos were cultured in EmbryoMax KSOM (Millipore)
551 or Evolve KSOMaa (Zenith Biotech) until they formed blastocysts (94–98 hr after hCG injection) at which
552 point they were placed in a drop of Evolve w/HEPES KSOMaa (Zenith) medium under mineral oil. A flat
553 tip microinjection pipette with an internal diameter of 16 μ m (Origio) was used for iPSC injections. Each
554 blastocyst received 8–12 iPSCs. Shortly after injection, blastocysts were transferred to day 2.5 recipient
555 CD1 females (20 blastocysts per female). Pups, when not born naturally, were recovered at day 19.5 by
556 cesarean section and fostered to lactating Balb/c mothers.

557

558 **Nuclear transfer**

559 Nuclear transfer was performed as described (Wakayama et al., 1998) with modifications. Briefly,
560 metaphase II-arrested oocytes were collected from superovulated B6D2F1 females (8-10 wks) and
561 cumulus cells were removed using hyaluronidase. The oocytes were enucleated in a droplet of HEPES-
562 CZB medium containing 5 μ g/ml cytochalasin B (CB) using a blunt Piezo-driven pipette. After enucleation,
563 the spindle-free oocytes were washed extensively and maintained in CZB medium up to 2 h before
564 nucleus injection. The CCs from mice (B6D2F1) were aspirated in and out of the injection pipette to
565 remove the cytoplasmic material and then injected into enucleated oocytes. The reconstructed oocytes
566 were cultured
567 in CZB medium for 1 h and then activated for 5-6 h in activation medium containing 10mM Sr²⁺, 5ng/ml
568 trichostatin A (TSA) and 5 μ g/ml CB. Following activation, all of the re constructed embryos were
569 cultured in KSOM medium supplemented with 5ng/ml TSA for another 3-4 hours and maintained in
570 KSOM medium with amino acids at 37°C under 5% CO₂ in air.

571

572 **Reduced-representation bisulfite sequencing (RRBS)**

573 DNA was isolated from MEFs and incubated in lysis buffer (25mM Tris-HCl (pH8), 2mM
574 ethylenediaminetetraacetic acid, 0.2% sodium dodecyl sulfate, 200mM NaCl) supplemented with
575 300 μ g/mL proteinase K (Roche) followed by phenol:chloroform extraction and ethanol precipitation
576 and RRBS libraries were prepared (Boyle et al., 2012) and run on HiSeq 2500 (Illumina) using 100bp
577 paired-end sequencing.

578 DNA methylation was analyzed by using 100bp paired-end sequencing reads from RRBS that were
579 trimmed and quality filtered by trim galore software using default parameters for RRBS. Read alignment
580 (genome build mm10) and extraction of single-base resolution methylation levels were carried out by

581 BSMAP. Differentially methylated regions (DMR) were explored with R methylKit package version 1.18.0
582 (Akalin et al., 2012). CpG sites featuring less than 10 reads were considered unreliable and discarded
583 from further analysis. CpG sites were then aggregated into consecutive tiles of size 100 bp and a
584 threshold of at least 15 reads per tile was applied. Differential methylation between the two lines, each
585 consisting of three samples, was determined by logistic regression and adjusted p-values are calculated
586 with SLIM (sliding linear model). A threshold of 1E-3 was set for adjusted p-value and a threshold of 20
587 methylation points was set between the two lines and further explored. DMRs were annotated with
588 Homer (Hypergeometric Optimization of Motif Enrichment) version 4.11.1 (Heinz et al., 2010) and
589 specifically its function annotatePeaks.pl. This function outputs a set of genes affiliated with DMR based
590 on the nearest promoter distance. Heatmaps were created with R package heatmap.2 version 3.1.1 and
591 dendrogram with R package dendextend version 1.15.2 (Galili, 2015).

592

593 **ACKNOWLEDGEMENTS**

594 Y.B. is supported by research grants from EMBO Young Investigator Programme (YIP), DKFZ-MOST (CA
595 177), Howard Hughes Medical Institute International Research Scholar (HHMI, #55008727) and by a
596 generous gift from Ms. Nadia Guth Biasini. We thank Yuval Nevo and huji core bioinformatics unit for
597 analyzing part of the RNA-seq data.

598

599 **Author Information**

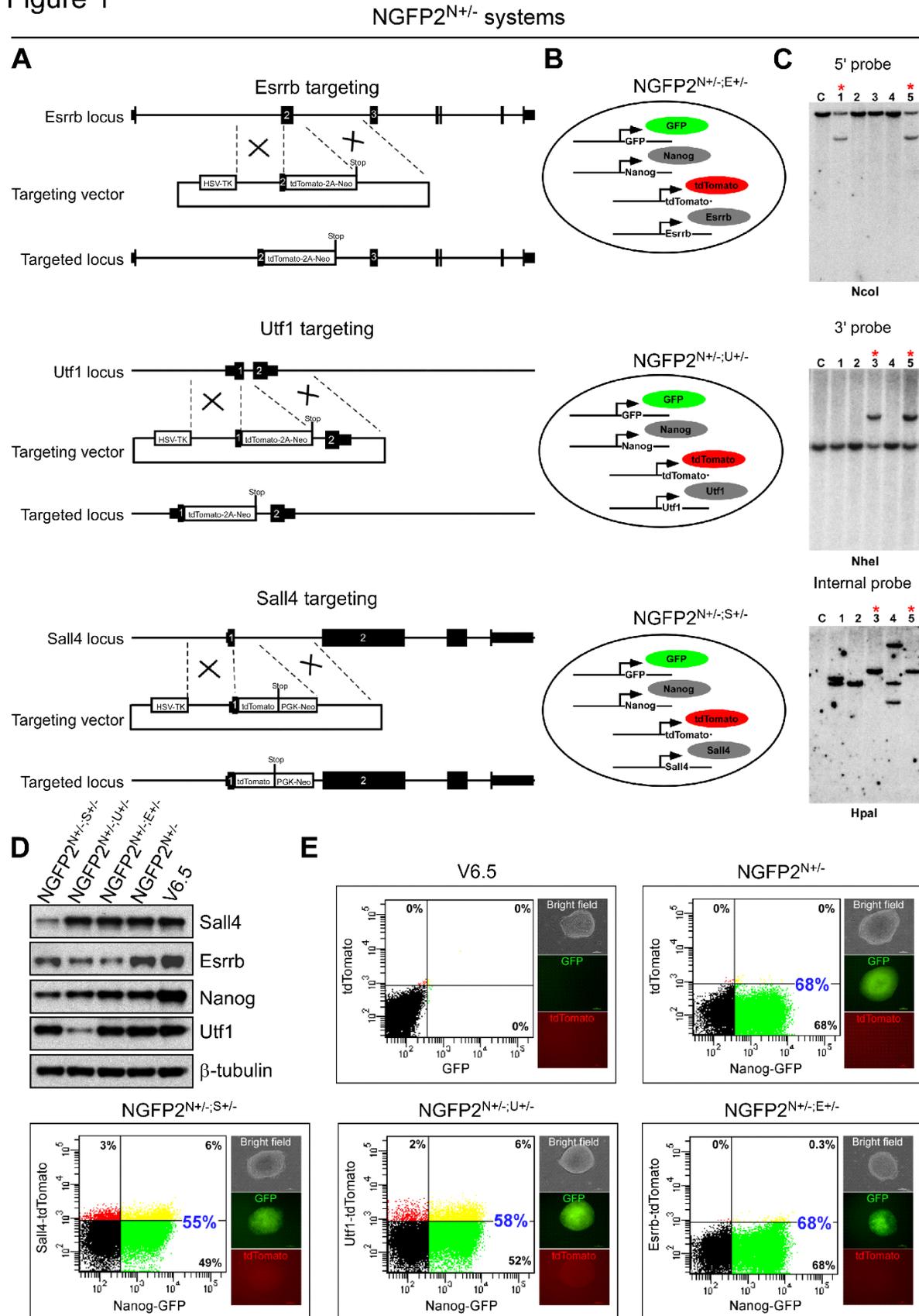
600 RNA-seq data for the various NGFP2^{N+/-} double heterozygous mutant and control iPSC lines (accession
601 number GSE182009) and RNA-seq for NGFP2^{N+/-} double heterozygous mutant and control MEF lines
602 after 6 days of reprogramming and RRBS for the SGFP1^{S2+/-} and SGFP1^{S2+/-;S4+/-} primary MEFs (accession
603 number GSE192655) has been deposited to the Gene Expression Omnibus database (GEO). The
604 reviewers can enter the deposited data using the following token: wxcbcokaplurhsn. Correspondence
605 and requests for materials should be addressed to Y.B. (yossib@ekmd.huji.ac.il).

606

607 **Author Contribution**

608 Y.B. and R.J. conceived the study. Y.B., R.L. designed the experiments, prepared the figures and wrote
609 the manuscript. Y.B. together with E.K., C.O., and D.F. generated the NGFP2^{N+/-} double heterozygous
610 mutant lines and ran the various reprogramming experiments on NGFP2^{N+/-} lines. R.L. generated the
611 tracing systems, with the help of N.M. for Nanog and Sall4 and the NGFP1^{N+/-;S+/-}, NGFP1^{N-/-} and SGFP1^{N+/-}
612 ^{S4+/-} lines, performed reprogramming experiments on these lines, immunostaining, flow cytometry and
613 5'azaDC experiments. N.M. prepared the samples for the RNA-seq at day 6 of reprogramming and
614 performed qPCR for the MET genes. N.M. together with N.YT. ran rescue reprogramming experiments
615 and performed sm-mRNA-FISH. A.W.C. analyzed the RNA-seq data from the various NGFP2^{N+/-} iPSC lines.
616 H.Y. performed SCNT experiments. S.M. and K.M. injected iPSC lines to produce secondary MEFs and
617 chimeric mice. M.A. helped R.L. to run reprogramming experiments and to analyze the flow cytometry
618 results.

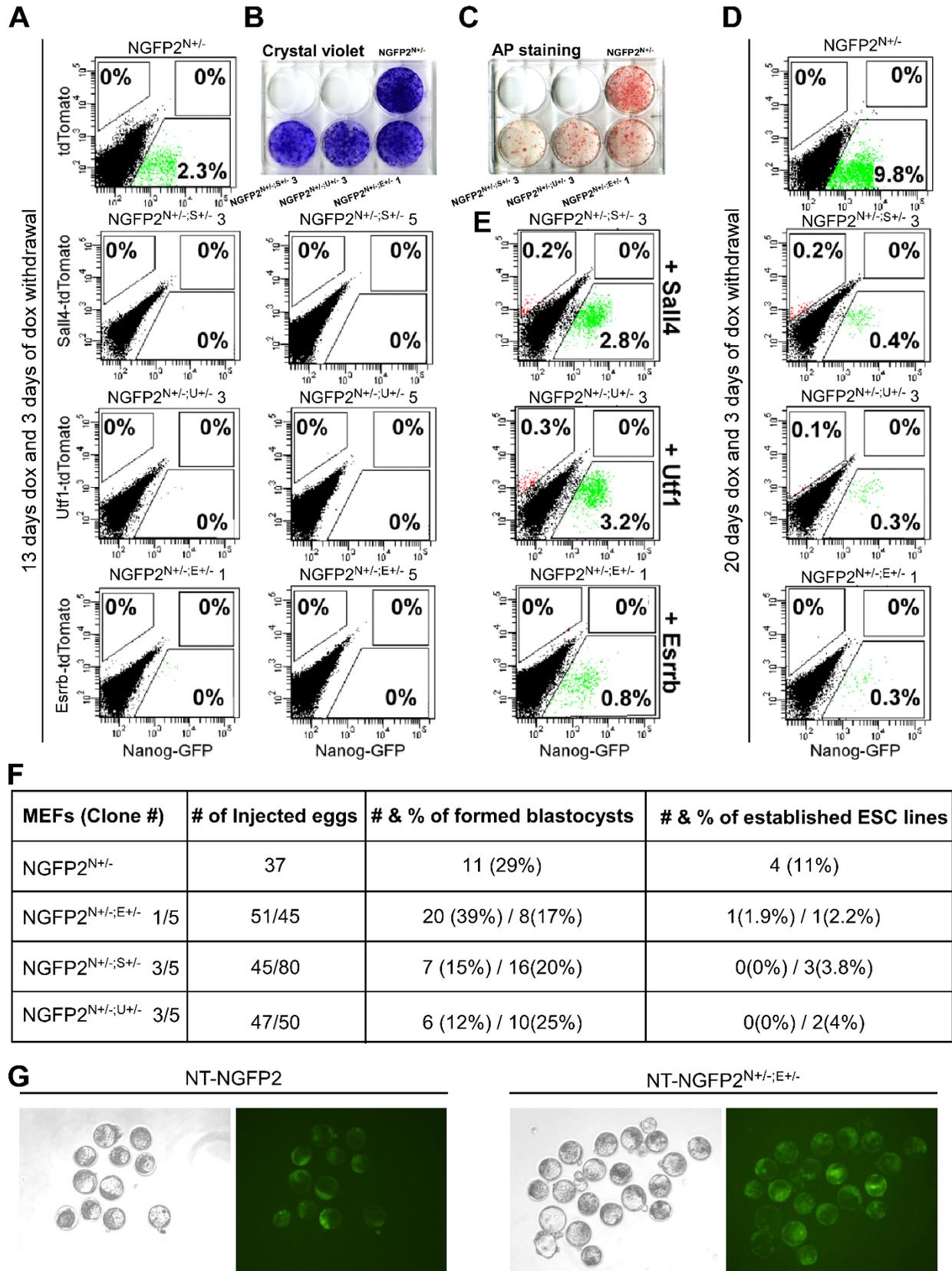
Figure 1



619

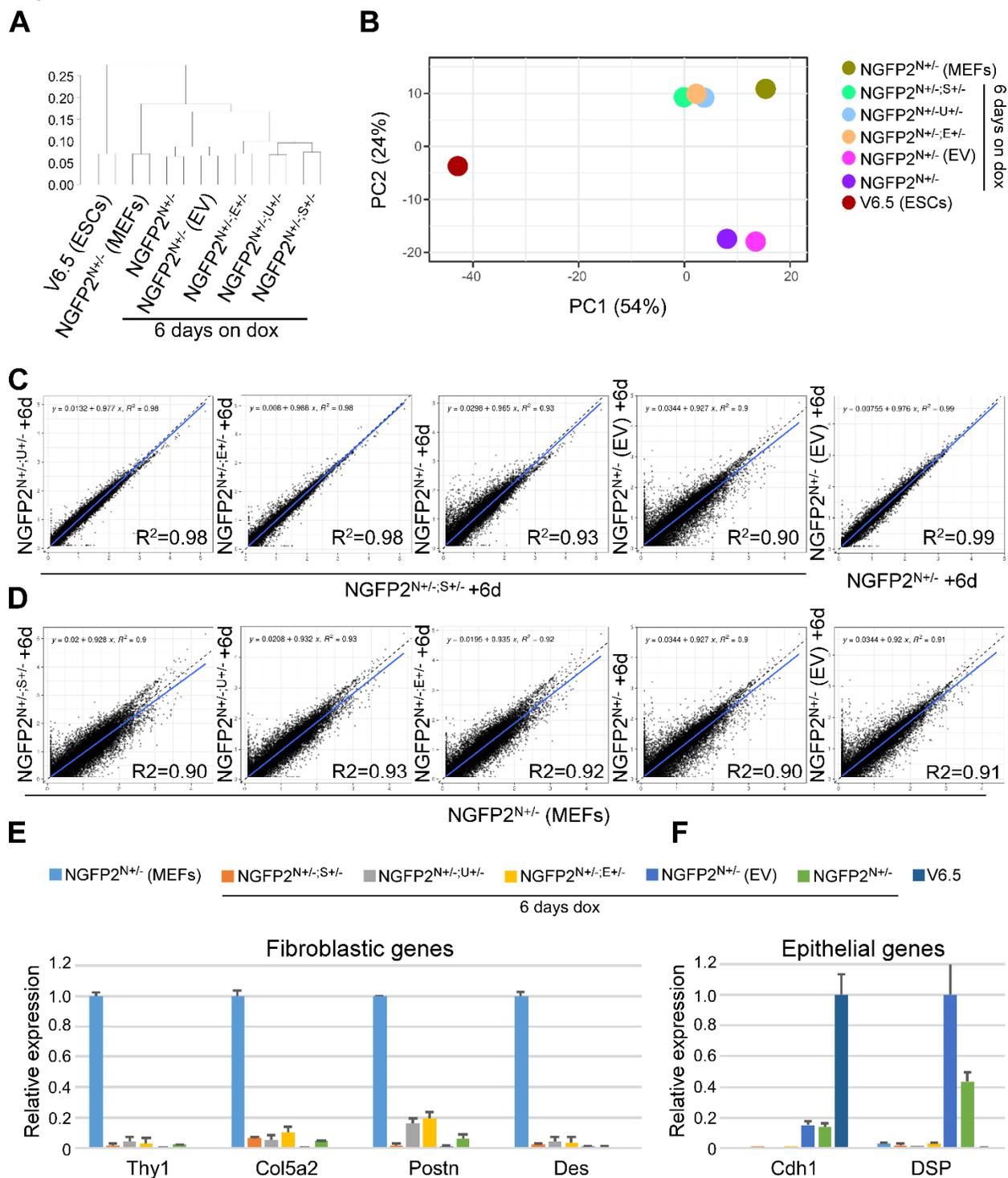
620 **Figure 1. Generation of double heterozygous mutant NGFP2^{N+/-} iPSC lines. (A, B)** Schematic
621 representation of the KI/KO targeting strategy for replacing one allele of *Esrrb*, *Utf1* or *Sall4* with a
622 tdTomato reporter in NGFP2^{N+/-} line. **(C)** Southern blot analysis for NGFP2^{N+/-} targeted iPSC clones
623 demonstrating heterozygous targeting for *Esrrb*, *Utf1* and *Sall4*. Correctly targeted clones are marked by
624 red asterisks. **(D)** Western blot analysis demonstrating a reduction of ~50% of the protein levels of the
625 targeted gene (*Esrrb*, *Utf1*, *Nanog* and *Sall4*) compared to ESC (V6.5) control. Cells were grown in 2i
626 condition to facilitate expression from both alleles. **(E)** Flow cytometry analysis for GFP (*Nanog*) and
627 tdTomato (*Utf1*, *Esrrb* or *Sall4*) in the various double heterozygous mutant lines that grew under S/Lif
628 conditions. Note that while cells express the GFP reporter due to functional polyA signal, tdTomato is
629 hardly detectable due to the absence of polyA. Similar to the western blot (D) GFP expression is reduced
630 even further in NGFP2^{N+/-;U+/-} and NGFP2^{N+/-;S+/-} compared to NGFP2^{N+/-;E+/-} and the parental cells
631 NGFP2^{N+/-}.

Figure 2



633 **Figure 2. NGFP2^{N+/-} double heterozygous mutant MEFs show strong reprogramming inhibition either**
634 **by OSKM or by somatic cell nuclear transfer (SCNT). (A)** Flow cytometry analysis of Nanog-GFP and
635 tdTomato-positive cells for two different clones from each of the NGFP2 double heterozygous mutant
636 MEFs and control following 13 days of dox exposure following by 3 days of dox withdrawal. **(B)** Crystal
637 violet staining of whole reprogramming plate for each of the double heterozygous mutant MEF line and
638 control at the end of the reprogramming process **(C)** Alkaline phosphatase staining of whole
639 reprogramming plate of each of the double heterozygous mutant MEF line and control at the end of the
640 reprogramming process. **(D)** Flow cytometry analysis of Nanog-GFP and tdTomato-positive cells for each
641 of the NGFP2^{N+/-} double heterozygous mutant MEFs and control following 20 days of dox exposure
642 following by 3 days of dox withdrawal. **(E)** Flow cytometry analysis of Nanog-GFP and tdTomato-positive
643 cells of each of the NGFP2^{N+/-} double heterozygous mutant MEFs and control following overexpression of
644 the targeted gene (*Sall4*, *Utf1* and *Esrrb*) and reprogramming process of 13 days and 3 days of dox
645 withdrawal. **(F)** Table summarizing the efficiency (i.e. blastocyst formation and ESC derivation) of the
646 somatic cell nuclear transfer (SCNT) process of MEF nuclei of the different double heterozygous mutant
647 NGFP2^{N+/-} lines. **(G)** Representative bright field and green channel images of NGFP2^{N+/-} and NGFP2^{N+/-;E+/-}
648 following SCNT. Note that both cell lines produced Nanog-GFP-positive blastocysts.

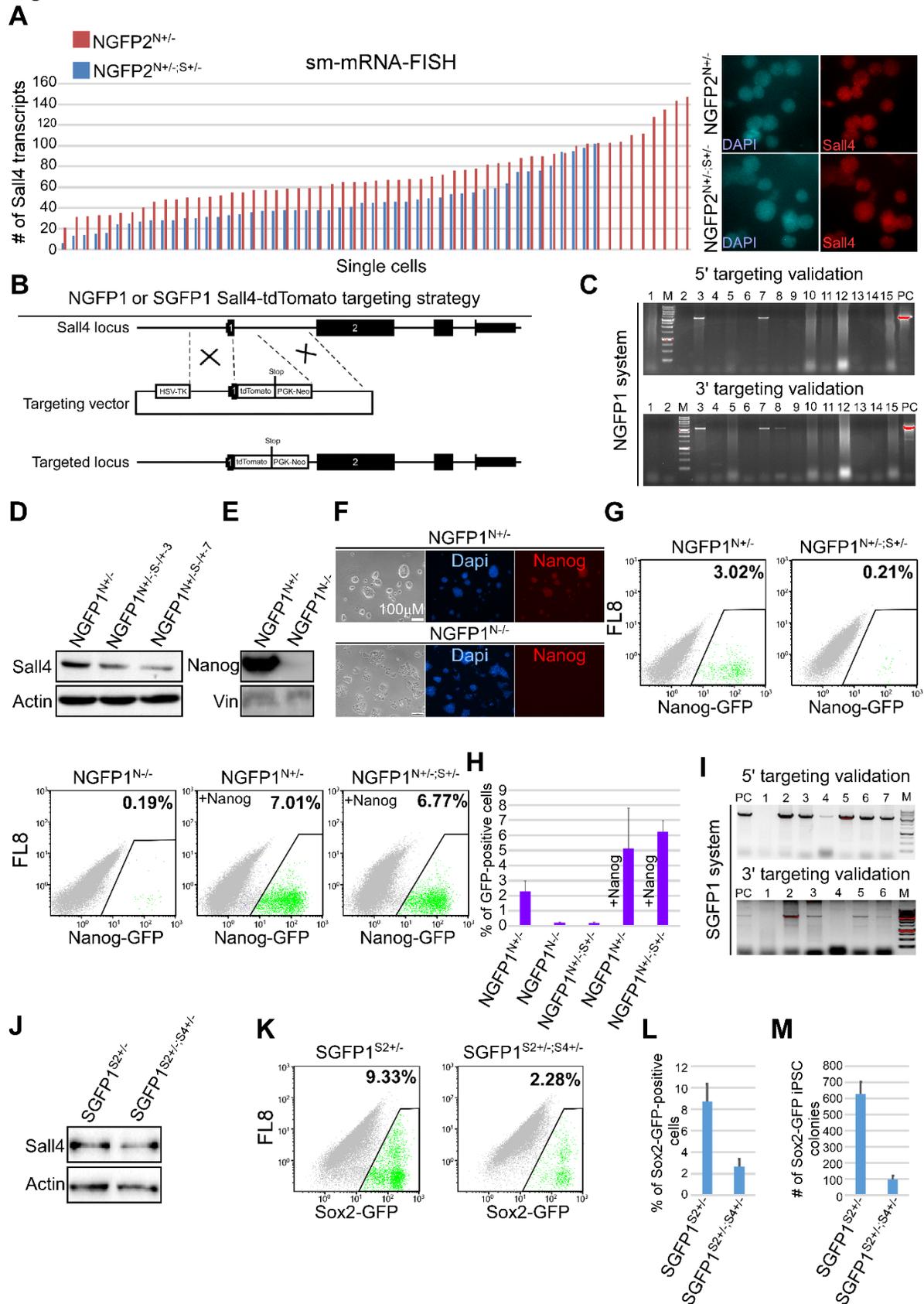
Figure 3



649
650

651 **Figure 3. Unbiased comparative transcriptome analyses after 6 Days of dox clusters NGFP2^{N+/-} double**
652 **heterozygote lines far from NGFP2^{N+/-} control. (A)** Hierarchical clustering of global gene expression
653 profiles for two RNA-seq replicates for NGFP2^{N+/-} iPSCs, NGFP2^{N+/-} MEFs and NGFP2^{N+/-}, NGFP2^{N+/-} (EV)
654 and the various NGFP2^{N+/-} double heterozygous mutant lines (NGFP2^{N+/-; E+/-}, NGFP2^{N+/-; U+/-} and NGFP2^{N+/-;}
655 ^{S4+/-}) after 6 days of reprogramming. Replicate pairs were assigned a shared numerical value. **(B)**
656 Principle component analysis for genes from (A). PC1, 54%; PC2, 24%. Each line is marked by a specific
657 color. The group names correspond to the names in (A). **(C, D)** Scatter plots comparing gene expression
658 between the indicated NGFP2^{N+/-} lines after 6 days of dox and controls. Blue line shows the linear
659 representation of the data, black line shows the $y = x$ line. **(E, F)** qPCR of the indicated fibroblastic genes
660 (E) and epithelial genes (F) in NGFP2^{N+/-} and the different NGFP2^{N+/-} double heterozygous mutant lines
661 after 6 days of dox, MEFs and V6.5 ESCs controls. mRNA levels were normalized to the housekeeping
662 control gene *Gapdh*. Error bars presented as a mean \pm SD of 2 duplicate runs from a typical experiment
663 out of 3 independent experiments.

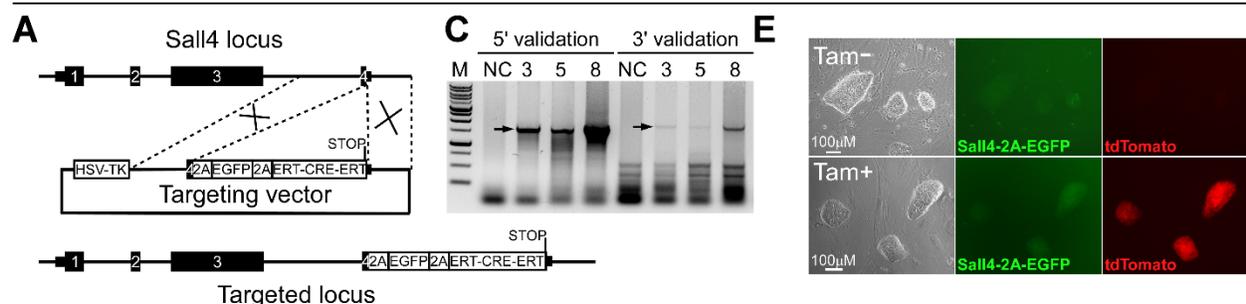
Figure 4



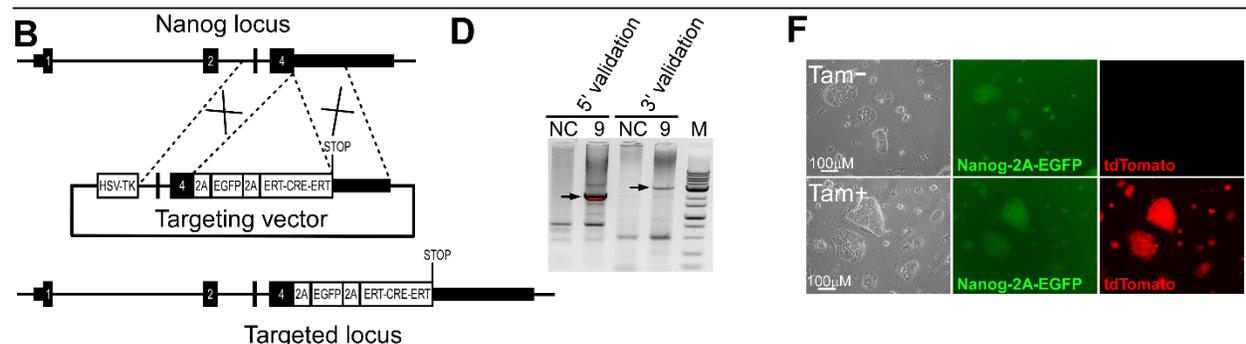
665 **Figure 4. NGFP1^{N+/-} double heterozygous mutant MEFs and Nanog KO MEFs show strong**
666 **reprogramming inhibition. (A)** Sm-mRNA-FISH directed towards *Sall4* transcripts in NGFP2^{N+/-} and
667 NGFP2^{N+/-;S+/-} single iPSCs. **(B)** Schematic representation of the KI/KO targeting strategy for replacing one
668 allele of *Sall4* with a tdTomato reporter in NGFP1^{N+/-} and SGFP1^{S2+/-;S4+/-} respectively. **(C)** PCR analysis for
669 NGFP1^{N+/-} targeted iPSC clones demonstrating correctly targeted clones for one allele of *Sall4*. Correctly
670 targeted clones were verified using primers amplifying regions at the 5' and 3' end of the targeted *locus*.
671 **(D)** Western blot analysis demonstrating a reduction of ~50% of the protein levels of *Sall4* compared to
672 parental NGFP1^{N+/-} control. **(E, F)** NGFP1^{N+/-} parental iPSCs were transfected with CRISPR/Cas9 and gRNA
673 against *Nanog* to produce *Nanog* KO NGFP1^{N+/-} line. Western blot analysis (E) and immunostaining (F)
674 demonstrating a complete KO of *Nanog* compared to parental NGFP1^{N+/-} control. **(G)** Flow cytometry
675 analysis of Nanog-GFP-positive cells in NGFP1^{N+/-} control cells, NGFP1^{N+/-;S+/-}, NGFP1^{N+/-} and following
676 overexpression of *Nanog* in rescue experiments after 13 days of reprogramming following by 3 days of
677 dox withdrawal. **(H)** Comparative percentage of Nanog-GFP-positive cells for NGFP1^{N+/-}, NGFP1^{N+/-;S+/-},
678 NGFP1^{N+/-}, NGFP1^{N+/-} (+*Nanog* OE) and NGFP1^{N+/-;S4+/-} (+*Nanog* OE) following reprogramming with OSKM
679 (13 days of dox + 3 days dox withdrawal). **(I)** PCR validation for SGFP1^{S2+/-;S4+/-} clones. **(J)** Western blot
680 analysis detecting *Sall4* in SGFP1^{S2+/-} and SGFP1^{S2+/-;S4+/-} iPSCs. **(K)** Flow cytometry analysis of Sox2-GFP-
681 positive cells for SGFP1^{S2+/-} compared with SGFP1^{S2+/-;S4+/-} following reprogramming with OSKM (13 days
682 of dox + 3 days dox withdrawal). **(L)** Comparative percentage of SOX2-GFP-positive cells for SGFP1^{S2+/-}
683 compared with SGFP1^{S2+/-;S4+/-} following reprogramming with OSKM (13 days of dox + 3 days dox
684 withdrawal). **(M)** Graph summarizing the number of colonies counted at the end of the reprogramming for
685 SGFP1^{S2+/-} and SGFP1^{S2+/-;S4+/-}.

Figure 5

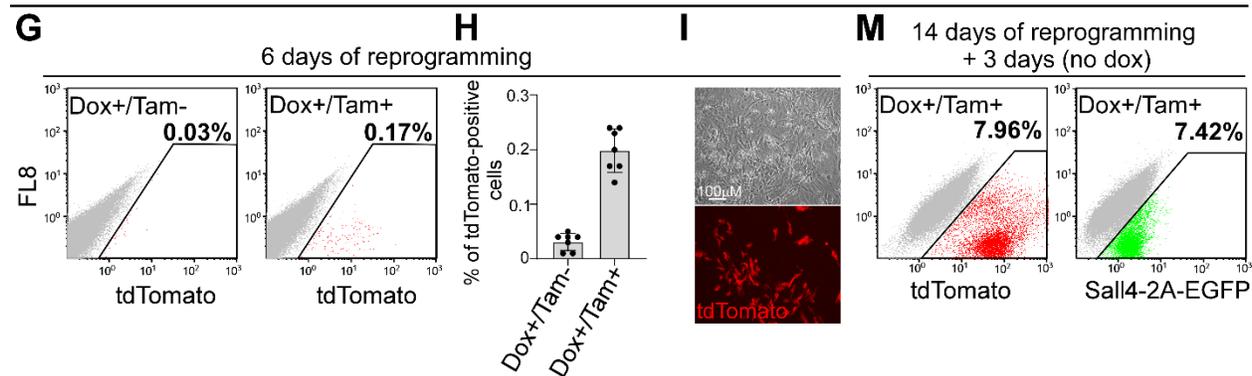
RL8 ESCs: *Sall4*^{2A-EGFP-2A-ERT-CRE-ERT};*Rosa26*^{M2rtTA;tdTomato-LSL}



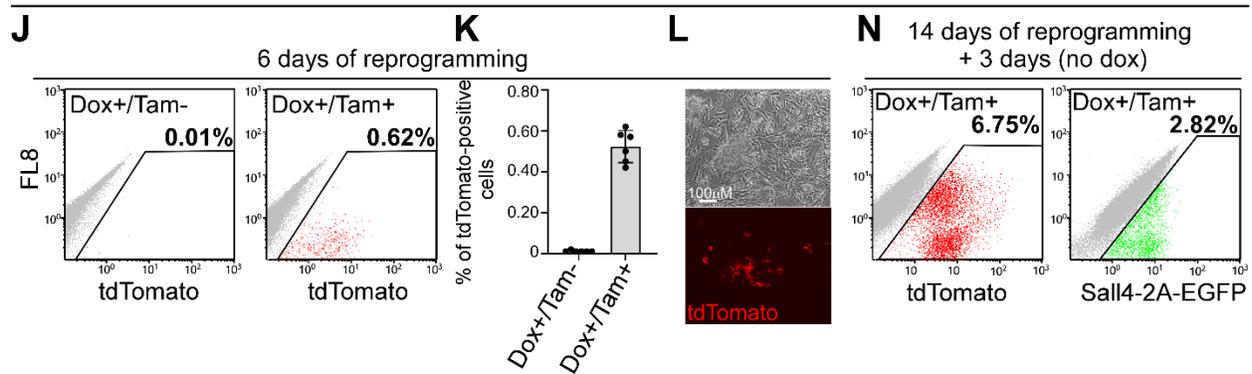
RL9 ESCs: *Nanog*^{2A-EGFP-2A-ERT-CRE-ERT};*Rosa26*^{M2rtTA;tdTomato-LSL}



Sall4 tracing system (RL8)

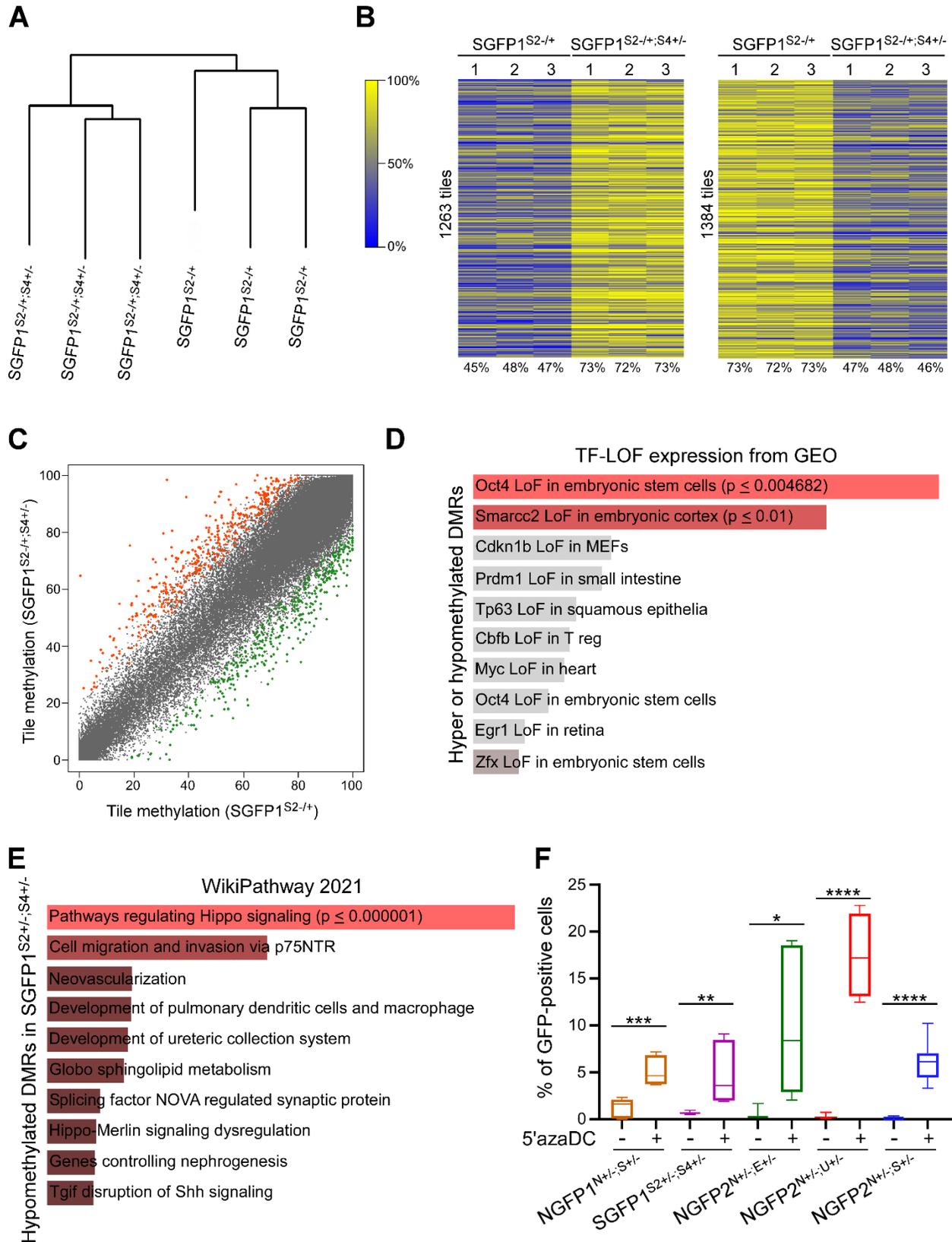


Nanog tracing system (RL9)



687 **Figure 5. Sall4 and Nanog tracing systems cannot explain the reprogramming blockage observed in the**
688 **double heterozygous mutant cells. (A, B)** Schematic representation of the targeting strategy to
689 introduce a 2A-EGFP-ERT-CRE-ERT cassette into the *Sall4* locus (A) or into the *Nanog* locus (B). **(C, D)** PCR
690 validation for targeted colonies demonstrating a correct targeting band size for the *Sall4* locus (C) and
691 for the *Nanog* locus (D) using both 5' and 3' regions of the incorporation point. Black arrows depict the
692 band size of correctly targeted allele. NC- negative control. **(E, F)** Representative bright field, RFP and
693 GFP channel images for the *Sall4* tracing system (RL8, E) and for the *Nanog* tracing system (RL9, F)
694 before and after tamoxifen addition. Scale bar, 100 μ m. **(G)** Flow cytometry analysis of tdTomato-
695 positive RL8 MEFs that were infected with OSKM and submitted to dox for 6 days with or without
696 tamoxifen (Tam). **(H)** Graph summarizing the percentages of tdTomato-positive cells of the *Sall4* tracing
697 system after 6 days of dox with or without Tamoxifen. **(I)** Bright field and RFP channel images of
698 tdTomato-positive cells from the *Sall4* tracing system after six days of dox and tamoxifen addition. **(J)**
699 Flow cytometry analysis of tdTomato-positive RL9 MEFs that were infected with OSKM and submitted to
700 dox for 6 days with or without tamoxifen (Tam). **(K)** Graph summarizing the percentages of tdTomato-
701 positive cells of the *Nanog* tracing system after 6 days of dox with or without Tamoxifen. **(L)** Bright field
702 and RFP channel images of tdTomato-positive cells from the *Nanog* tracing system after six days of dox
703 and tamoxifen addition. **(M, N)** Flow cytometry analysis of tdTomato and *Sall4*-GFP-positive cells (M) or
704 *Nanog*-GFP-positive cells (N) following 14 days of OSKM induction in the presence of dox and Tamoxifen
705 followed by 3 days of dox withdrawal.

Figure 6

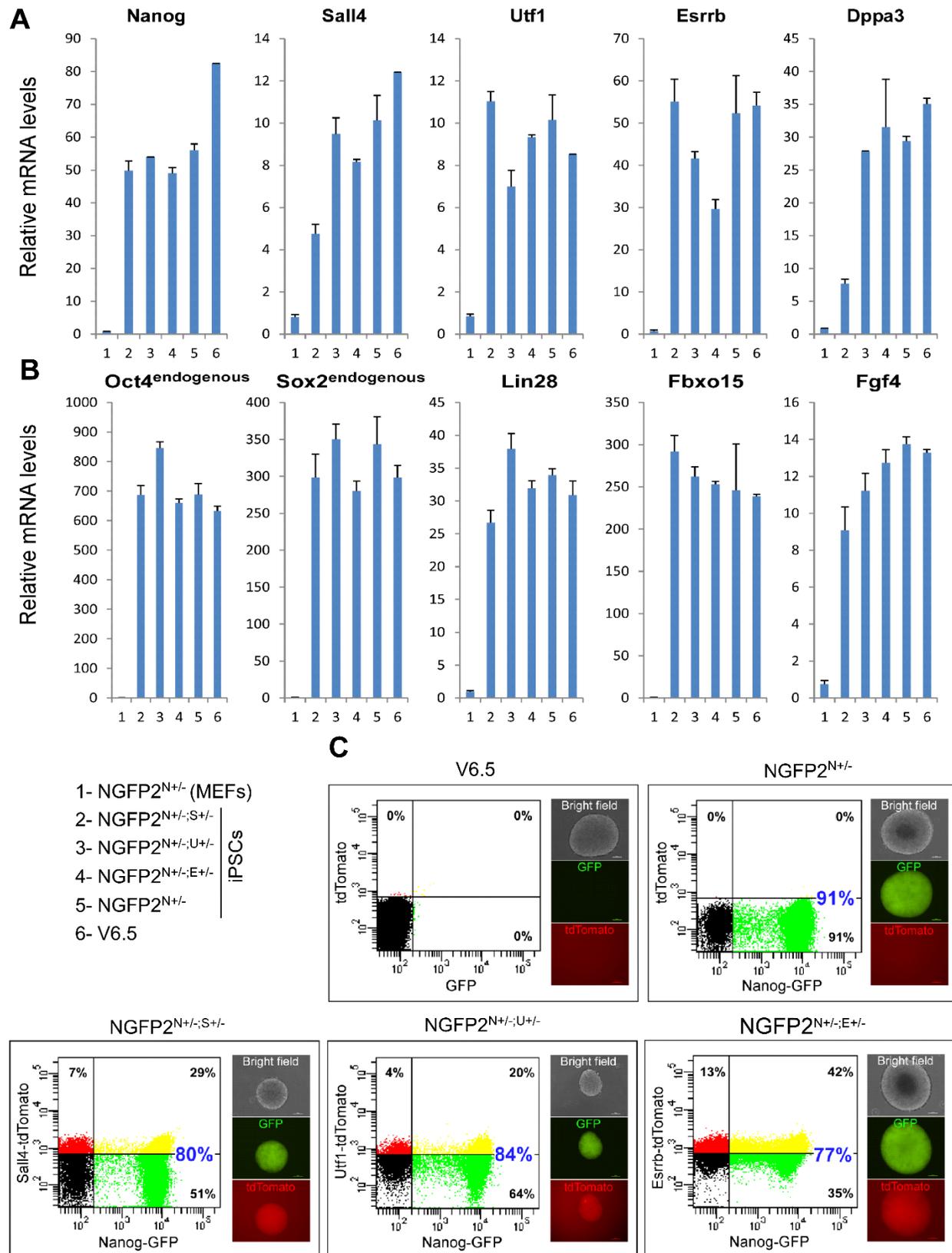


706

707

708 **Figure 6. DNA methylation abnormalities in the double heterozygous mutant fibroblasts hinder the**
709 **reprogramming process. (A)** Dendrogram for SGFP1^{S2+/-} MEFs and SGFP1^{S2+/-S4+/-} MEFs based on the level
710 of relative change observed at CpG sites with a threshold of 10 reads per site. **(B)** Heatmap of 20% of
711 differentially methylated regions (DMRs) which covers a 100bp genomic region/tile and filtered to
712 include at least 15 reads, which may be either hypermethylated or hypomethylated in SGFP1^{S2+/-S4+/-}
713 MEFs compared to its parental SGFP1^{S2+/-} control MEFs. p-value ≤ 0.001 . **(C)** Scatter plot analysis showing
714 all the differentially methylated regions between SGFP1^{S2+/-} MEFs and SGFP1^{S2+/-S4+/-} MEFs (average of 3
715 biological replicates). Blue dots represent regions that are significantly more methylated in the double
716 heterozygous mutant SGFP1^{S2+/-S4+/-} MEFs while red dots represent regions that are more methylated in
717 the control SGFP1^{S2+/-} MEFs. Red dots represent regions that are associated with genes that are related
718 to pluripotency and development and are significantly more methylated in SGFP1^{S2+/-S4+/-} MEFs while
719 green dots represent regions that are associated with genes that are related to pluripotency and
720 development and are significantly more methylated in SGFP1^{S2+/-} MEFs. Gray area represents no
721 significant differences between the samples. **(D, E)** EnrichR of GEO (D) and wiki pathways analysis (E) of
722 significantly over-represented genes that are either hyper or hypomethylated DMRs (D) or
723 hypomethylated DMRs significantly enriched in SGFP1^{S2+/-S4+/-} MEFs (E) **(F)** Bar plot graph displaying the
724 percentage of GFP-positive cells in the indicated samples after 13 days of reprogramming and 3 days of
725 transgene removal with and without prior treatment of 5'azaDC for two days. Error bars indicate
726 standard deviation between 6-7 biological replicates. *p-value ≤ 0.01 , **p-value ≤ 0.001 as calculated by
727 GraphPad Prism using 2-tailed Student's t-test.

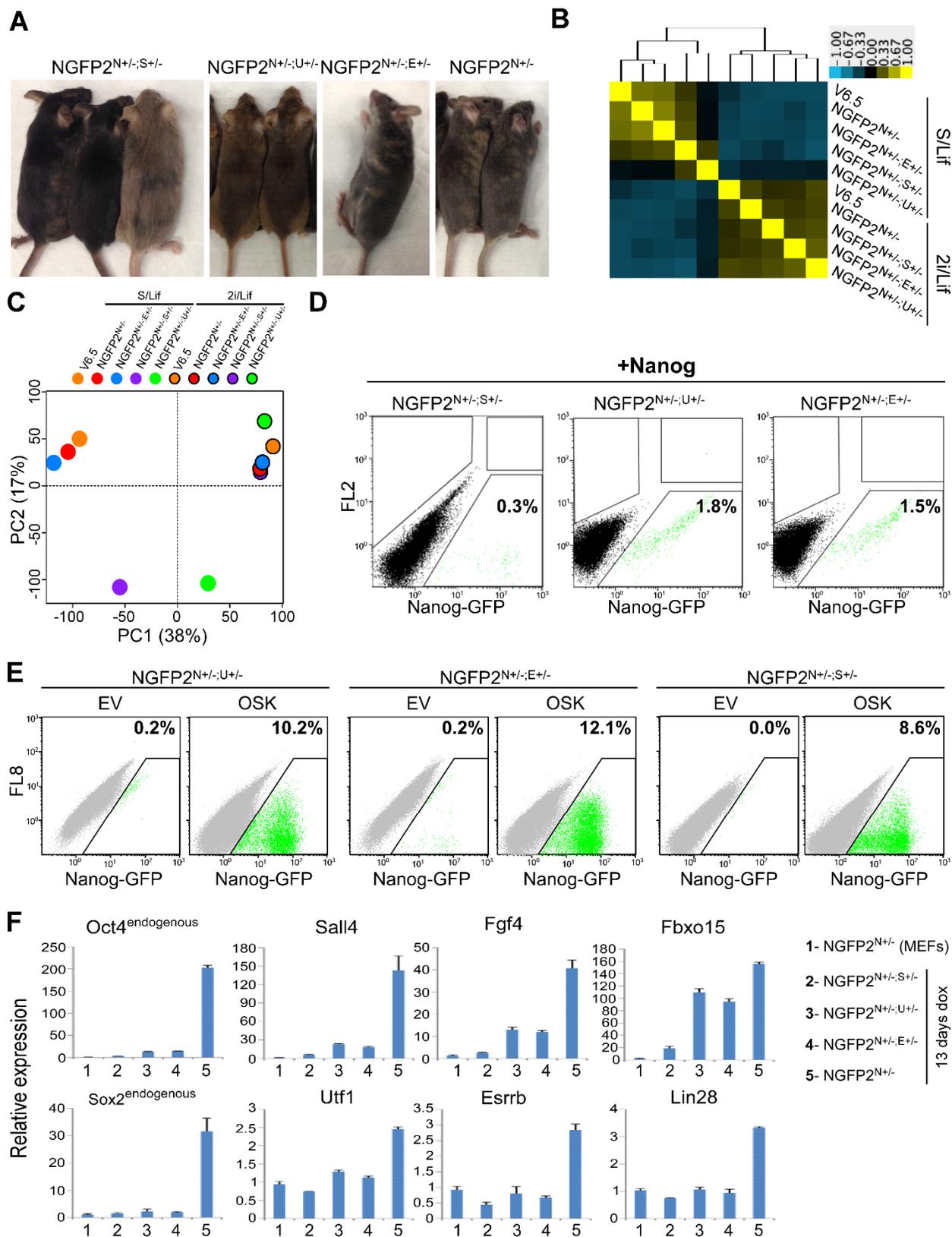
Figure S1



728

729 **Figure S1. Characterization of the double heterozygous mutant NGFP2^{N+/-} lines. (A)** qPCR of the
730 indicated genes normalized to the housekeeping control gene *Gapdh* in the various NGFP2^{N+/-} double
731 heterozygous mutant lines, NGFP2^{N+/-} parental line, and ESC (V6.5) and MEF controls. Error bars
732 presented as a mean \pm SD of 2 duplicate runs from a typical experiment out of 3 independent
733 experiments. **(B)** Flow cytometry analysis for GFP (*Nanog*) and tdTomato (*Utf1*, *Esrrb* or *Sal14*) in the
734 various double heterozygous mutant lines that grew under 2i/Lif conditions. Note that although
735 tdTomato lacks polyA, a red signal is still detectable

Figure S2

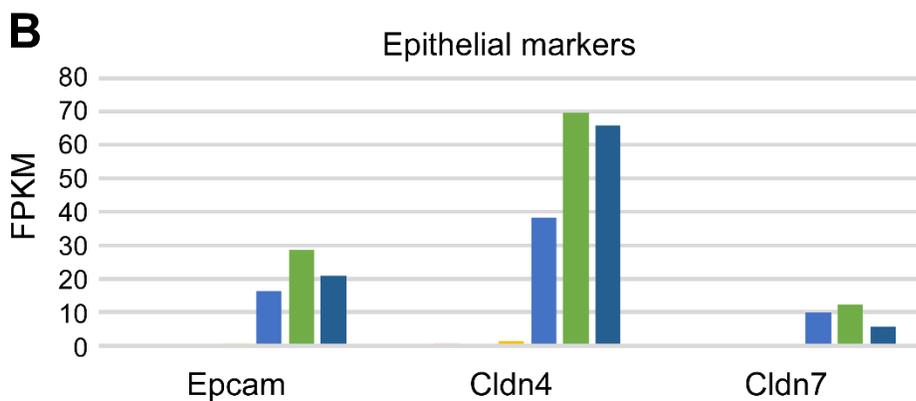
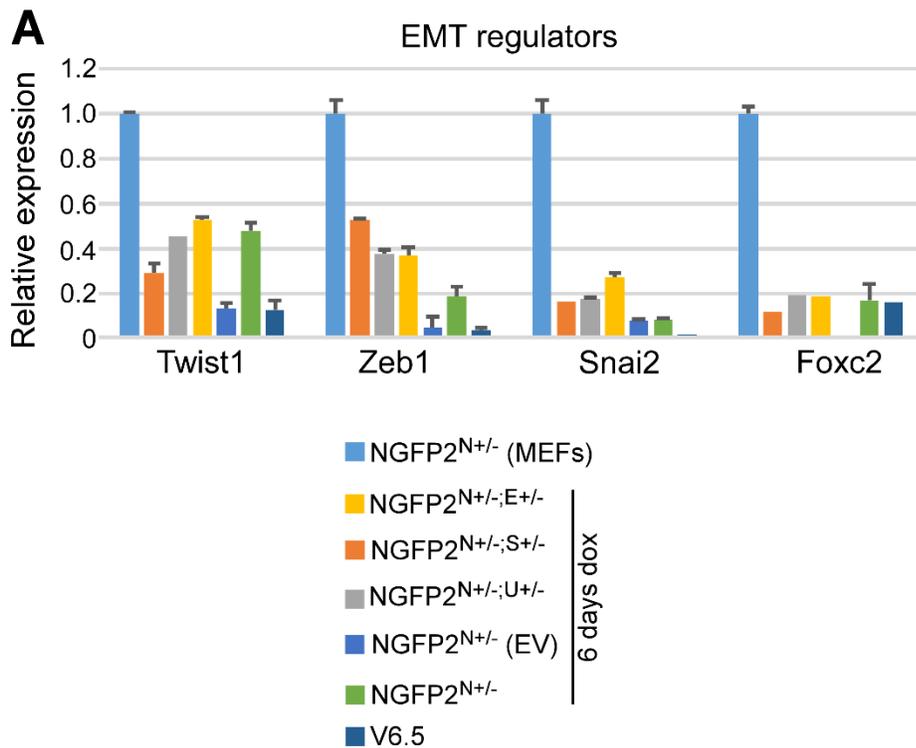


736

737

738 **Figure S2. The developmental potential and transcriptional profile of NGFP2^{N+/-} double heterozygous**
739 **mutant lines and rescue reprogramming experiments (A)** Representative images of adult chimeric mice
740 produced by the various NGFP2^{N+/-} double heterozygous mutant iPSC lines and control following
741 blastocyst injection and transplantation into foster mothers. **(B)** Correlation heatmap and dendrogram
742 of global gene expression profiles for two RNA-seq replicates for the indicated NGFP2^{N+/-} iPSC lines and
743 ESC (V6.5) control grown under S/Lif or 2i/Lif conditions. Replicate pairs were assigned a shared
744 numerical value. **(C)** Principle component analysis for the indicated samples using 500 most differentially
745 expressed genes among all samples. PC1, 38%; PC2, 17%. Each line is marked by a specific color. The
746 group names correspond to the names in (B). Cells that were grown in 2i/Lif are surrounded with black
747 circle. **(D, E)** Flow cytometry analysis of Nanog-GFP-positive cells for the various NGFP2^{N+/-} double
748 heterozygous mutant lines following overexpression of *Nanog* (D) or OSK (E) and reprogramming for 13
749 days following by 3 days of dox removal. OSK indicates *Oct4*, *Sox2* and *Klf4* and EV indicates empty
750 vector **(F)** NGFP2^{N+/-} double heterozygous mutant lines and control were reprogrammed for 13 days.
751 qPCR analysis showing the expression levels of the indicated genes, in the depicted samples, after
752 *Gapdh* normalization. Error bars presented as a mean \pm SD of 2 duplicate runs from a typical experiment
753 out of 3 independent experiments.

Figure S3

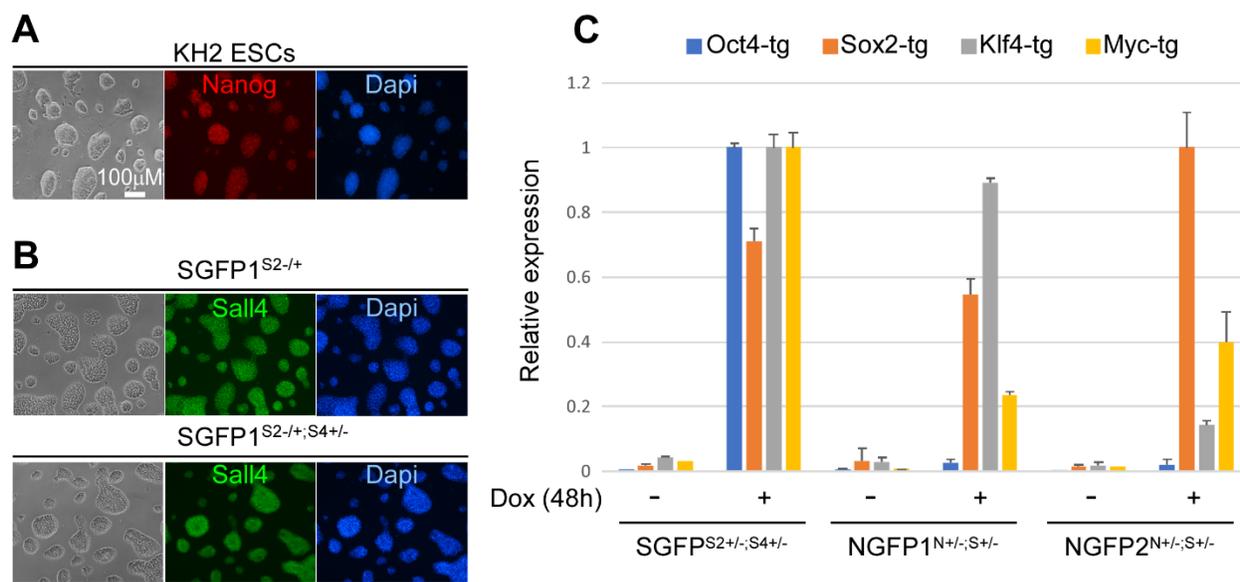


754

755

756 **Figure S3. NGFP2^{N+/-} double heterozygous mutant lines fail to activate the epithelial program during**
757 **reprogramming. (A)** qPCR of the indicated EMT genes normalized to housekeeping control gene *Gapdh*
758 in the various NGFP2^{N+/-} double heterozygous mutant lines following 6 days of dox and in ESCs (V6.5)
759 and NGFP2^{N+/-} MEF control. Error bars presented as a mean \pm SD of 2 duplicate runs from a typical
760 experiment out of 3 independent experiments. **(B)** Graph summarizing the expression level (FPKM-
761 Fragments Per Kilobase Million) of the indicated epithelial genes in the various NGFP2^{N+/-} double
762 heterozygous mutant lines after 6 days of dox and in ESCs (V6.5) and NGFP2^{N+/-} MEF control. Expression
763 level of the depicted genes was obtained from the RNA-seq data described in Figure 3.

Figure S4



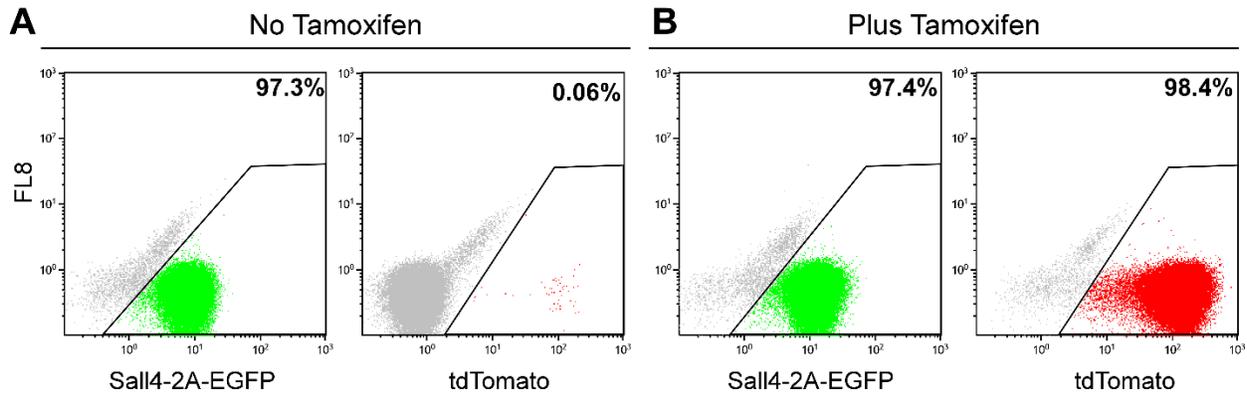
764

765

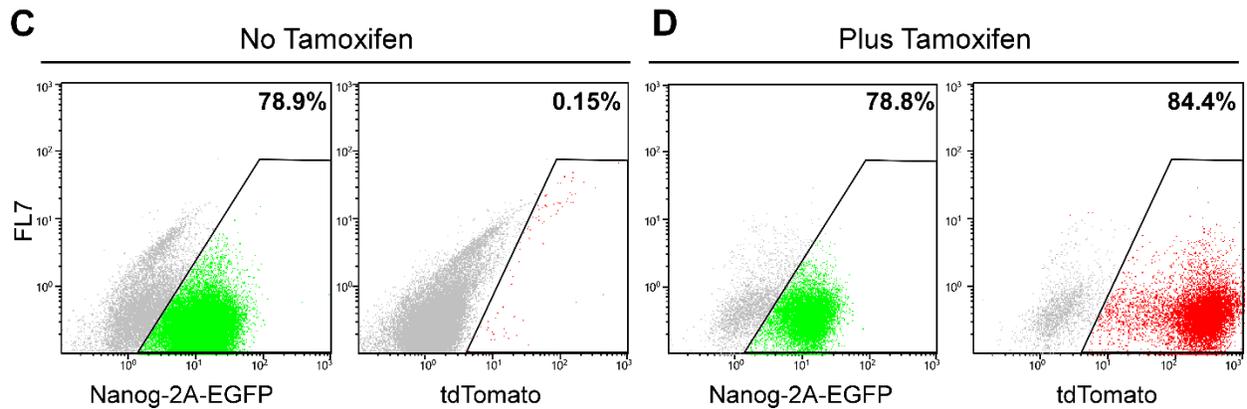
766 **Figure S4. Nanog and Sall4 protein level in targeted iPSC lines and controls (A)** Bright field images and
767 immunostaining images for *Nanog* (red) and Dapi (blue) in KH2 ESCs. Scale bar, 100 μ M. **(B)** Bright field
768 images and immunostaining images for Sall4 (green) and Dapi (blue) in SGFP1^{S2+/-} and SGFP1^{S2+/-;S4+/-} iPSC
769 lines. Scale bar, 100 μ M. **(C)** qPCR of the indicated OSKM transgenes normalized to housekeeping control
770 gene *Gapdh* in the various double heterozygous mutant MEF lines following 2 days of culture with or
771 without dox. Error bars presented as a mean \pm SD of 2 duplicate runs from a typical experiment out of 3
772 independent experiments.

Figure S5

RL8 ESCs: *Sall4*^{2A-EGFP-2A-ERT-CRE-ERT};*Rosa26*^{M2rtTA};tdTomato-LSL



RL9 ESCs: *Nanog*^{2A-EGFP-2A-ERT-CRE-ERT};*Rosa26*^{M2rtTA};tdTomato-LSL



773

774

775 **Figure S5. Sall4 and Nanog tracing system characterization. (A)** Flow cytometry analysis for Sall4-2A-
776 EGFP and tdTomato in the targeted ESC clone RL8 before and after tamoxifen addition (48 hours). **(B)**
777 Flow cytometry analysis for Nanog-2A-EGFP and tdTomato in the targeted ESC clone RL9 before and
778 after tamoxifen addition (48 hours).

779 **Table 1. primer list**

Gene	Application	Primer Sequence (5' --> 3')
Gapdh	qPCR analysis of mRNA expression normalization	F: CCTCAACGACCACTTTGTCAAG R: TCTTCCTCTGTGCTCTTGCTG
Thy1	qPCR analysis of mRNA expression	F: CCAGAACGTCACAGTGCTCA R: AGGTGTTCTGAGCCAGCAG
Col5a2	qPCR analysis of mRNA expression	F: TAGAGGAAGAAAGGGACAAAAGG R: GTTACAACAGGCACTAATCCTGGTT
Postn	qPCR analysis of mRNA expression	F: ACAACAATCTGGGGCTTTTT R: AATCTGGTTCCCATGGATGA
Des	qPCR analysis of mRNA expression	F: TGGAGCGTGACAACCTGATA R:AAGGCAGCCAAGTTGTTCTC
Cdh1	qPCR analysis of mRNA expression	F: CTCGACACCCGATTCAAAGT R: GCGTAGACCAAGAAATGGA
Dsp	qPCR analysis of mRNA expression	F: ACCGTCAACGACCAGAACTC R: TTTGCAGCATTTCTTGATG
Nanog	qPCR analysis of mRNA expression	F: AAACCAGTGGTTGAAGACTAGCAA R: GGTGCTGAGCCCTTCTGAATC
Oct4 endogenous	qPCR analysis of mRNA expression	F: TCAGTGATGCTGTTGATCAGG R: GCTATCTACTGTGTGCCAGTC
Sox2 endogenous	qPCR analysis of mRNA expression	F: CCGTTTTCTGTGGTCTTGTTT R: TCAACCTGCATGGACATTTT
Lin28	qPCR analysis of mRNA expression	F: GAAGAACATGCAGAAGCGAAGA R: CCGCAGTTGTAGCACCTGTCT
Fbxo15	qPCR analysis of mRNA expression	F: CGAGAATGGTGGACTAGCTTTTG R: GGCCATGGGAATGAATATTTG
Fgf4	qPCR analysis of mRNA expression	F: GCAGACACGAGGGACAGTCT R: ACTCCGAAGATGCTCACCAC
Sall4	qPCR analysis of mRNA expression	F: GCAAGTCACCAGGGCTCTT R: CCTCCTTAGCTGACAGCAATC
Utf1	qPCR analysis of mRNA expression	F: GTCCCTCTCCGCGTTAGC R: GGCAGGTTTCGTCATTTTCC
Esrrb	qPCR analysis of mRNA expression	F: CACCTGCTAAAAAGCCATTGACT R: CAACCCCTAGTAGATTTCGAGACGAT
Dppa3	qPCR analysis of mRNA expression	F: TCGGATTGAGCAGAGACAAAAA R: TCCCGTTCAAACCTCATTTCCTT
Twist1	qPCR analysis of mRNA expression	F: ACGCTGCCCTCGGACAA R: CCTGGCCGCCAGTTTG
Zeb1	qPCR analysis of mRNA expression	F: CCAGGTGTAAGCGCAGAAAG R: TCATCGGAATCTGAATTTGCT
Snai2	qPCR analysis of mRNA expression	F: ATCCTCACCTCGGGAGCATA R: TGCCGACGATGTCCATACAG
Foxc2	qPCR analysis of mRNA expression	F: AGAACAGCATCCGCCACAAC R: GCACTTTACGAAGCACTCATT
Oct4-transgene	qPCR analysis of transgenic	F : CGCCTGGAGACGCCATCCACGCT

	mRNA expression	R: GTTGGTTCCACCTTCTCCAA
Sox2-transgene	qPCR analysis of transgenic mRNA expression	F: GCCCAGTAGACTGCACATGG R: AGAATACCAGTCAATCTTTCA
Klf4-transgene	qPCR analysis of transgenic mRNA expression	F: CGCCTGGAGACGCCATCCACGCT R: ACGCAGTGTCTTCTCCCTTC
Myc-transgene	qPCR analysis of transgenic mRNA expression	F: TGTCCATTCAAGCAGACGAG R: AGAATACCAGTCAATCTTTCA
Nanog gRNA	gRNA for generating Nanog KO iPSCs	F: CACCGAGAACTATTCTTGCTTACA R: AAAGTGAAGCAAGAATAGTTCTC
Nanog KO	KO validation PCR	F: CGGCTCACTTCTTCTGACT R: TATTGCTCCGTCTGTGTCC
Nanog tracing 5 arm	PCR for generating arm for targeting vector	F: TAACAGCTGAAGTACCTCAGCCTCCAGCA R: TAACAGCTGTATTTACCTGGTGGAGTCACA
Nanog tracing 3 arm	PCR for generating arm for targeting vector	F: GGTACCCAGCCCCTGGTTTATTTTT R: CCGCGGACCCACACAGCCTCTCAAGT
Nanog gRNA	gRNA tracing	F: CACCGGATTTGAACTCCTGACCTT R: AAACAAGGTCAGGAGTTCAAATCC
Nanog validation 5 arm tracing	PCR analysis of integration into genomic DNA	F: CCACCCCGTGAAGTACTGACT R: CGTCACCGCATGTTAGAAGA
Nanog validation 3 arm tracing	PCR analysis of integration into genomic DNA	F: GGTACCCAGCCCCTGGTTTATTTTT R: CCCTGTGAGTGGTCAGGAGT
Sall4 tracing 5 arm	PCR for generating arm for targeting vector	F: GTTAACGCAAGGGAGAGCCAGTATT R: GTTAACGCTGACAGCAATCTTATT
Sall4 tracing 3 arm	PCR for generating arm for targeting vector	F: GGTACCCTGATATGCAAGTGATGT R: CCGCGGATACACACAAGCCCGCCTC
Sall4 gRNA	gRNA tracing	F: CACCGGAGGAGAGGAGTCTTCTGC R: AAACGCAGAAGACTCCTCTCCTCC
Sall4 validation 5 arm tracing	PCR analysis of integration into genomic DNA	F: TAATCCAGCCTTGCTCGTCT R: CGTCACCGCATGTTAGAAGA
Sall4 validation 3 arm tracing	PCR analysis of integration into genomic DNA	F: ACAGCTGTGAGGTACCCTGA R: GTGTGTGTGTGTCCTCCTC
Nanog-cDNA	Primers used for cloning of cDNA for lentiviral gene overexpression	F: CGCCATCACACTGACATGA R: TGGAAGAAGGAAGGAACCTG
Sall4-cDNA	Primers used for cloning of cDNA for lentiviral gene overexpression	F: GCAAGTCACCAGGGCTCTT R: CCTCTTAGCTGACAGCAAT
Esrrb-cDNA	Primers used for cloning of cDNA for lentiviral gene overexpression	F: GCTGGAACACCTGAGGGTAA R: GGTCTCCACTTGGATCGTGT
Utf1-cDNA	Primers used for cloning of cDNA for lentiviral gene overexpression	F: CTACCTGGCTCAGGGATGCT R: GACTGGGAGTCGTTTCTGGA
Sall4 gRNA	gRNA for generating Sall4 KI/KO in NGFP1 and SGFP1	F: CACCGCCAGCTCTCCGCGGATGGT R: AAACACCATCCGCGGAGAGCTGGC
Sall4 5arm validation PCR	PCR analysis of integration	F: CATAACAAAGCCCCAGGTT

	into genomic DNA	R: GCGCATGAACTCTTTGATGA
Sall4 3arm validation PCR	PCR analysis of integration into genomic DNA	F: CGGGATCCGAAGTTCCTATT R: AGCTTGCAAAGGGAAAGACA
Sall4 targeting 5arm	PCR for generating arm for targeting vector	F: GTTAACGTGGTGCAGGCCTGTTATCT R: AAGCTTCTCCTCCCAGTTGATGTGCT
Sall4 targeting 3arm	PCR for generating arm for targeting vector	F: CCGCGGTGGTCCACCTGGAACAAAA R: CCGCGGAGAAGGGAGCTATGGCACA

780

781 REFERENCES

- 782 Adachi, K., Kopp, W., Wu, G., Heising, S., Greber, B., Stehling, M., Arauzo-Bravo, M.J., Boerno, S.T.,
783 Timmermann, B., Vingron, M., *et al.* (2018). Esrrb Unlocks Silenced Enhancers for Reprogramming to
784 Naive Pluripotency. *Cell stem cell* 23, 900-904.
- 785 Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E.
786 (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation
787 profiles. *Genome Biol* 13, R87.
- 788 Arnold, K., Sarkar, A., Yram, M.A., Polo, J.M., Bronson, R., Sengupta, S., Seandel, M., Geijsen, N., and
789 Hochedlinger, K. (2011). Sox2(+) adult stem and progenitor cells are important for tissue regeneration
790 and survival of mice. *Cell stem cell* 9, 317-329.
- 791 Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell
792 lineages in early mouse development depend on SOX2 function. *Genes Dev* 17, 126-140.
- 793 Benchetrit, H., Herman, S., van Wietmarschen, N., Wu, T., Makedonski, K., Maoz, N., Yom Tov, N., Stave,
794 D., Lasry, R., Zayat, V., *et al.* (2015). Extensive Nuclear Reprogramming Underlies Lineage Conversion
795 into Functional Trophoblast Stem-like Cells. *Cell stem cell* 17, 543-556.
- 796 Benchetrit, H., Jaber, M., Zayat, V., Sebban, S., Pushett, A., Makedonski, K., Zakheim, Z., Radwan, A.,
797 Maoz, N., Lasry, R., *et al.* (2019). Direct Induction of the Three Pre-implantation Blastocyst Cell Types
798 from Fibroblasts. *Cell stem cell* 24, 983-994 e987.
- 799 Boiani, M., Eckardt, S., Scholer, H.R., and McLaughlin, K.J. (2002). Oct4 distribution and level in mouse
800 clones: consequences for pluripotency. *Genes Dev* 16, 1209-1219.
- 801 Boyle, P., Clement, K., Gu, H., Smith, Z.D., Ziller, M., Fostel, J.L., Holmes, L., Meldrim, J., Kelley, F., Gnirke,
802 A., *et al.* (2012). Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA
803 methylation profiling. *Genome Biol* 13, R92.
- 804 Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van
805 Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming
806 reveal an early stochastic and a late hierarchic phase. *Cell* 150, 1209-1222.
- 807 Buganim, Y., Faddah, D.A., and Jaenisch, R. (2013). Mechanisms and models of somatic cell
808 reprogramming. *Nature reviews Genetics* 14, 427-439.
- 809 Buganim, Y., Markoulaki, S., van Wietmarschen, N., Hoke, H., Wu, T., Ganz, K., Akhtar-Zaidi, B., He, Y.,
810 Abraham, B.J., Porubsky, D., *et al.* (2014). The developmental potential of iPSCs is greatly influenced by
811 reprogramming factor selection. *Cell stem cell* 15, 295-309.
- 812 Carey, B.W., Markoulaki, S., Hanna, J.H., Faddah, D.A., Buganim, Y., Kim, J., Ganz, K., Steine, E.J., Cassady,
813 J.P., Creighton, M.P., *et al.* (2011). Reprogramming factor stoichiometry influences the epigenetic state
814 and biological properties of induced pluripotent stem cells. *Cell stem cell* 9, 588-598.
- 815 Carter, A.C., Davis-Dusenbery, B.N., Koszka, K., Ichida, J.K., and Eggan, K. (2014). Nanog-independent
816 reprogramming to iPSCs with canonical factors. *Stem cell reports* 2, 119-126.
- 817 De Carvalho, D.D., You, J.S., and Jones, P.A. (2010). DNA methylation and cellular reprogramming.
818 *Trends in cell biology* 20, 609-617.
- 819 Eastman, A.E., Chen, X., Hu, X., Hartman, A.A., Pearlman Morales, A.M., Yang, C., Lu, J., Kueh, H.Y., and
820 Guo, S. (2020). Resolving Cell Cycle Speed in One Snapshot with a Live-Cell Fluorescent Reporter. *Cell*
821 *Rep* 31, 107804.
- 822 Elling, U., Woods, M., Forment, J.V., Fu, B., Yang, F., Ng, B.L., Vicente, J.R., Adams, D.J., Doe, B., Jackson,
823 S.P., *et al.* (2019). Derivation and maintenance of mouse haploid embryonic stem cells. *Nat Protoc* 14,
824 1991-2014.
- 825 Feng, B., Jiang, J., Kraus, P., Ng, J.H., Heng, J.C., Chan, Y.S., Yaw, L.P., Zhang, W., Loh, Y.H., Han, J., *et al.*
826 (2009). Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor
827 Esrrb. *Nat Cell Biol* 11, 197-203.

828 Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A.,
829 Tomlinson, S.R., and Chambers, I. (2012). *Esrrb* is a direct Nanog target gene that can substitute for
830 Nanog function in pluripotent cells. *Cell stem cell* *11*, 477-490.

831 Galili, T. (2015). *dendextend*: an R package for visualizing, adjusting and comparing trees of hierarchical
832 clustering. *Bioinformatics* *31*, 3718-3720.

833 Guo, L., Lin, L., Wang, X., Gao, M., Cao, S., Mai, Y., Wu, F., Kuang, J., Liu, H., Yang, J., *et al.* (2019).
834 Resolving Cell Fate Decisions during Somatic Cell Reprogramming by Single-Cell RNA-Seq. *Mol Cell* *73*,
835 815-829 e817.

836 Haenebalcke, L., Goossens, S., Dierickx, P., Bartunkova, S., D'Hont, J., Haigh, K., Hochepped, T., Wirth, D.,
837 Nagy, A., and Haigh, J.J. (2013). The ROSA26-iPSC mouse: a conditional, inducible, and exchangeable
838 resource for studying cellular (De)differentiation. *Cell Rep* *3*, 335-341.

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

842 Heslop, J.A., Pournasr, B., Liu, J.T., and Duncan, S.A. (2021). GATA6 defines endoderm fate by controlling
843 chromatin accessibility during differentiation of human-induced pluripotent stem cells. *Cell Rep* *35*,
844 109145.

845 Leeb, M., and Wutz, A. (2011). Derivation of haploid embryonic stem cells from mouse embryos. *Nature*
846 *479*, 131-134.

847 Martinez-Val, A., Lynch, C.J., Calvo, I., Ximenez-Embun, P., Garcia, F., Zarzuela, E., Serrano, M., and
848 Munoz, J. (2021). Dissection of two routes to naive pluripotency using different kinase inhibitors. *Nat*
849 *Commun* *12*, 1863.

850 Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A.,
851 Matoba, R., Sharov, A.A., *et al.* (2007). Pluripotency governed by *Sox2* via regulation of *Oct3/4*
852 expression in mouse embryonic stem cells. *Nat Cell Biol* *9*, 625-635.

853 Meissner, A., Wernig, M., and Jaenisch, R. (2007). Direct reprogramming of genetically unmodified
854 fibroblasts into pluripotent stem cells. *Nat Biotechnol* *25*, 1177-1181.

855 Miyanari, Y., and Torres-Padilla, M.E. (2012). Control of ground-state pluripotency by allelic regulation of
856 *Nanog*. *Nature* *483*, 470-473.

857 Morshedi, A., Soroush Noghabi, M., and Droge, P. (2013). Use of UTF1 genetic control elements as iPSC
858 reporter. *Stem Cell Rev Rep* *9*, 523-530.

859 Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and
860 Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU
861 transcription factor *Oct4*. *Cell* *95*, 379-391.

862 Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou,
863 E., Alaei, S., Cloutier, J., *et al.* (2012). A molecular roadmap of reprogramming somatic cells into iPS cells.
864 *Cell* *151*, 1617-1632.

865 Schwarz, B.A., Bar-Nur, O., Silva, J.C., and Hochedlinger, K. (2014). *Nanog* is dispensable for the
866 generation of induced pluripotent stem cells. *Current biology : CB* *24*, 347-350.

867 Sevilla, A., Papatsenko, D., Mazloom, A.R., Xu, H., Vasileva, A., Unwin, R.D., LeRoy, G., Chen, E.Y.,
868 Garrett-Bakelman, F.E., Lee, D.F., *et al.* (2021). An *Esrrb* and *Nanog* Cell Fate Regulatory Module
869 Controlled by Feed Forward Loop Interactions. *Front Cell Dev Biol* *9*, 630067.

870 Shanak, S., and Helms, V. (2020). DNA methylation and the core pluripotency network. *Dev Biol* *464*,
871 145-160.

872 Sim, Y.J., Kim, M.S., Nayfeh, A., Yun, Y.J., Kim, S.J., Park, K.T., Kim, C.H., and Kim, K.S. (2017). *2i* Maintains
873 a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms. *Stem cell reports* *8*, 1312-
874 1328.

875 Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency
876 reprogramming factors' initial engagement with the genome. *Cell* *151*, 994-1004.
877 Tan, M.H., Au, K.F., Leong, D.E., Foygel, K., Wong, W.H., and Yao, M.W. (2013). An Oct4-Sall4-Nanog
878 network controls developmental progression in the pre-implantation mouse embryo. *Mol Syst Biol* *9*,
879 632.
880 Theunissen, T.W., and Jaenisch, R. (2014). Molecular control of induced pluripotency. *Cell stem cell* *14*,
881 720-734.
882 Tsubooka, N., Ichisaka, T., Okita, K., Takahashi, K., Nakagawa, M., and Yamanaka, S. (2009). Roles of Sall4
883 in the generation of pluripotent stem cells from blastocysts and fibroblasts. *Genes Cells* *14*, 683-694.
884 Wakayama, T., Perry, A.C., Zuccotti, M., Johnson, K.R., and Yanagimachi, R. (1998). Full-term
885 development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* *394*, 369-374.
886 Wernig, M., Lengner, C.J., Hanna, J., Lodato, M.A., Steine, E., Foreman, R., Staerk, J., Markoulaki, S., and
887 Jaenisch, R. (2008). A drug-inducible transgenic system for direct reprogramming of multiple somatic cell
888 types. *Nat Biotechnol* *26*, 916-924.
889 Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A.,
890 Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M., *et al.* (2021). Gene Set Knowledge Discovery with
891 Enrichr. *Curr Protoc* *1*, e90.
892